Parameter Sensitivity and Uncertainty Analysis in Simplified Conceptual Urban Drainage Models

Cintia B. S. Dotto

Thesis submitted to Monash University in total fulfilment

of the requirements for the degree of

Doctor of Philosophy

May, 2013

Monash *Water for Liveability*
Department of Civil Engineering
MONASH University

This thesis is by publication, in
accordance with Section 17.2 of the
Research Graduate School Committee
Regulations (Doctorate).

Abstract

Stormwater models are powerful tools to aid the planning, design and performance of different stormwater management strategies. Although these models provide a great platform for decision making, they all have an intrinsic level of uncertainty. Little is understood about the sources and magnitude of this uncertainty, which could be due to the errors in measured data (input and calibration data) and/or due to the model itself. To better understand these sources and their impacts on the model predictions, robust model calibration and sensitivity analysis should be performed. The methodologies used for such an exercise should not only be able to provide an assessment of the uncertainties in the model's parameter values and an evaluation of the confidence level of the model's predictions, but also be able to identify and propagate the different sources of uncertainties.

The main aim of this research project is to assess uncertainties in conceptual urban stormwater flow and pollution generation models, with different levels of complexity, by evaluating the impact of different sources of uncertainties on the model predictions and parameter sensitivity. The research focuses on three main steps: (i) identifying suitable global sensitivity analysis method(s) to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models; (ii) exploring parameter calibration, model sensitivity and the resulting predictive uncertainties in models with different level of complexities; and, (iii) investigating the impact of measured input and calibration data uncertainty on the performance, sensitivity and predictive uncertainty of stormwater models.

Four methods were applied for calibration, sensitivity and uncertainty analysis of a simple stormwater (quantity and quality) model: one is a formal Bayesian approach, and three are methods based on Monte Carlo simulations coupled with different sampling and acceptance criteria. While the application of the four methods generated similar posterior parameter distributions and predictive uncertainty, results indicated that the selection of the most appropriate method is a trade-off between the need for a strong theory-based description of uncertainty (but limited by the requirements on prior knowledge), simplicity (but limited by the subjectivity) and computational efficiency (also affected by subjectivity). The results also suggested that modellers should select the method which is most suitable for the system they are modelling, their skill/knowledge level, the available information, and the purpose of their study. Further analysis of the application of the Bayesian approach verified the potential of the method to assess urban drainage models (with different level of complexities) in urban catchments of different sizes and land-use types. The tested Bayesian approach was selected to be used in the remaining activities of this research.

The likelihood function in the applied Bayesian approach assumes that the model errors (residuals) are normally distributed. This study demonstrated that this assumption is often not met in stormwater modelling (i.e. model residuals are not normally distributed), and therefore, the data was transformed (Box-Cox) to ensure the normality of the model residuals. The main finding was that the parameter sensitivity varied significantly between the scenarios in which the normality assumption of the residuals was verified or not. The main reason for this being the fact that the data transformation method to meet the assumption altered the intrinsic content of the measured data, which then influenced the emphasis on various parts of the hydrograph.

The Bayesian approach was used to assess two conceptual catchment rainfall runoff models (MUSIC, which simulates runoff from both impervious and pervious areas as a series of reservoirs; and, KAREN that simulates runoff from impervious surfaces using the time-area method) and few

simple stormwater quality models (empirical regressions and build-up/wash-off based models). Results from parameter calibration and sensitivity analysis of the rainfall runoff models demonstrated that the effective impervious fraction is the main parameter governing the prediction of runoff in urbanised catchments. Other key parameters are those related to the time of concentration. Indeed, the analysis indicated that the pervious area parameters play a secondary role when modelling highly urbanised catchments, which implies that the tested models could be simplified. The uncertainty analysis showed that the total predictive uncertainty bands (i.e. the total uncertainty derived from the specific modelling application) was considerably larger than the uncertainty bands contributed from parameter uncertainty alone, indicating that there are other prominent sources of uncertainty for these models. The water quality models were shown to be 'ill-posed' and unable to reproduce the pollutant processes in the catchment.

The impact of both input and calibration data errors on the parameter sensitivity and predictive uncertainty was evaluated by means of propagating these errors through the selected urban stormwater model (rainfall runoff model KAREN coupled with a build-up/wash-off water quality model). It was found that random errors in measured data had minor impact on the model performance and sensitivity. Systematic errors in input and calibration data impacted the parameter distributions (e.g. changed their shapes and location of peaks). In most of the systematic error scenarios (especially those where uncertainty in input and calibration data was represented using 'best-case' assumptions), the errors in measured data were fully compensated by the parameters. For example, when rainfall was systematically under or overestimated, the effective impervious area parameter varied systematically to compensate for the changes in the input data. Parameters were unable to compensate in some of the scenarios where the systematic uncertainty in the input and calibration data were represented using extreme worst-case scenarios. As such, in these few worst case scenarios, the model's performance was reduced considerably. Systematic errors in the calibration data error did not significantly impact the parameter probability distributions of the water quality model, mainly because the model cannot even reproduce TSS concentrations when the 'true' data is used. This finding suggested that the current main limitation in water quality modelling is related to poor model structure, and not to errors in measured data.

This research provides a comprehensive study of the propagation of different sources of uncertainties through stormwater models. It identifies how the different uncertainty sources impact on parameter sensitivity and the predictive uncertainty. In addition, the analysis of model parameters and their interactions provides practical recommendations for refining and further developing stormwater rainfall runoff and pollution generation models.

Declaration

This is to certify that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other institution and affirms that to the best of the candidate's knowledge the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature____

Date_________________________

Acknowledgements

I would like to express a big thanks to my supervisors Prof. Ana Deletic and Dr. David T. McCarthy for their unmeasurable support and encouragement during this long and not so smooth journey. Thanks for enduring with me. Dear, Ana, I am really grateful for the enthusiasm and support during all these years. Thanks Dave for your dedication and patient guidance.

I extend my gratitude to Tim Fletcher for his support and patience mainly during the first years of my PhD. I would also like to thank the staff of the Civil Engineering Department, in particular to Irene Sougas, Chris Powell, Jane Moody, Godwin Vaz and Alan Taylor. I am also grateful to Jenny, who made university life easier since the very beginning. Furthermore, I appreciate Ross Allen's understanding during the last year in which I was juggling between work and PhD. I would also like to acknowledge eWater CRC and CAPES for the scholarship provided.

I am indebted to the support from my colleagues in the department, especially to Ferny, Kathy, Ale, Peter Bach, Riane, Scott, Anja, Ranmalee, Dusan and Dale, who each, in their own way, helped me to go through this journey. I'm grateful to my officemates for the moral support and company along the way, and also to my dear Chris for her friendship and fitness inspiration.

I should mention the active collaboration of Manfred Kleidorfer to my research. In addition, I acknowledge the collaboration with the International Working Group on Data and Models, specifically the joint work with Luca Vezzaro, Malte Henrichs, Wolfgang Rauch, Giorgio Mannina and Gabriele Freni.

To my friends Estela, Leti, Lu, Mateus, Oli, Gi, Cla, a sincere thank-you for encouraging and being there for me even from the other side of the world. Also, thanks to my friends in Australia for bringing colour and fun into my journey. Specially, Burcu, Murat, Debs, Mi, Tan, Sile, Vin and Kate thanks for becoming part of my family during this journey. Also, thanks Anthony for your encouraging words and for the great camping escapes. And, to Nurses, my heartfelt thanks for your proof-readings and for the encouragement during rough times. But above all, thanks for your friendship!

Finally, huge thanks to my great family for all their love, support and understanding. Especially to my sister, dad and grandma, thanks for being there for me.

List of Publications

This thesis is done through publications. The following publications have resulted from the studies undertaken for this degree, of which the seven journal papers lead by the candidate authored formed the core of the thesis:

Refereed Journal Papers

Lead Authorship Papers:

1. **Dotto, C.B.S.**, Deletic, A. and Fletcher, T.D. (2009). Analysis of parameter uncertainty of a flow and quality stormwater model. *Water Science and Technology*, 60(3), 717-725. *(Impact Factor: 1.09, Number of citations: 14).*

2. **Dotto, C.B.S.**, Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D.T., Freni, G., Rauch, W. and Deletic, A. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research,* 46(8): 2545-2558. *(Impact Factor: 4.865, Number of citations: 5).*

3. **Dotto, C.B.S.**, Deletic, A., McCarthy, D.T. and Fletcher, T.D (2011). Calibration and sensitivity analysis of stormwater models. *Australian Journal of Water Resources,* 15(1), 85-93.

4. **Dotto, C.B.S.**, Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T. and Rauch, W. (2010). Stormwater quality models: performance and sensitivity analysis. *Water Science and Technology,* 62(4), 837–843. *(Impact Factor: 1.09, Number of citations: 15).*

5. **Dotto, C.B.S.**, Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T. and Fletcher, T.D. (2011). Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. *Environmental Modelling and Software*, 26(10), 1225-1239. *(Impact Factor: 3.166, Number of citations: 9).*

6. **Dotto, C.B.S.**, Deletic, A. and McCarthy, D.T. (in press). Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals? *Water Science and Technology. (Impact Factor: 1.09).*

7. **Dotto, C.B.S.**, Kleidorfer, M., Deletic, A., Rauch W. and D. McCarthy, D.T. (submitted). Impacts of measured data uncertainty on urban stormwater models. *Journal of Hydrology. (Impact Factor: 2.656).*

Co-authored Paper:

8. 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. (2012). Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth*, Parts A/B/C 42–44(0): 3-10. *(Impact Factor: 1.110, Number of citations: 6).*

Conference Papers

9. **Dotto, C.B.S.**, Deletic, A. and McCarthy, D.T. (2012). Urban drainage uncertainty analysis: should we break our back for normally distributed residuals? *9th International Conference on Urban Drainage Modelling*, Belgrade, Serbia, Sept 3-7, 2012.

10. **Dotto, C.B.S.**, Kleidorfer, M., McCarthy, D.T., Deletic, A., Rauch, W. and Fletcher, T.D. (2010). Towards Global Assessment of Modelling Errors. *6th International Conference on Sewer Processes and Networks (SPN6), Surfers Paradise, Australia, Nov 7-10, 2010.*

11. Deletic, A., **Dotto, C.B.S.**, Fletcher, T.D., McCarthy, D.T., Bertrand-Krajewski, J.-L., Rauch, W., Kleidorfer, M., Freni, G., Mannina, G. and Tait, S. (2009). Defining uncertainties in modelling of urban drainage systems. *8th International Conference on Urban Drainage Modelling jointly with the 2nd International Conference on Rainwater Harvesting and Management, Tokyo, Japan, Sept 7-11, 2009.*

12. **Dotto, C.B.S.**, Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T. and Rauch, W. (2009). Stormwater quality models: do they work? *8th International Conference on Urban Drainage Modelling jointly with the 2nd International Conference on Rainwater Harvesting and Management, Tokyo, Japan, Sept 7-11, 2009.*

13. **Dotto, C.B.S.**, Deletic, A., Fletcher, T.D. and McCarthy, D.T. (2009). Parameter sensitivity analysis of stormwater models. *6th International Conference on Water Sensitive Urban Design and Hydropolis #3, Perth, Australia, May 5-8, 2009.*

14. **Dotto, C.B.S.**, Deletic, A. and Fletcher, T.D. (2008). Analysis of uncertainty in flow and water quality from a stormwater model. *11th International Conference on Urban Drainage (ICUD). Edinburgh, Scotland, Aug 31-Sept 5, 2008.*

Preface

This thesis presents the results of the research into parameter sensitivity and uncertainties in urban drainage models in the form of seven journal papers (all with the PhD candidate as the lead author), of which five have been published and two have been submitted for review. These papers are accompanied by introduction, literature review, and discussion and conclusion chapters. The introduction and literature review focus on the current state and the main knowledge gaps in the following areas: (i) urban stormwater modelling approaches; (ii) sources of uncertainties in urban drainage models; (iii) uncertainties in urban drainage data; and, (iv) methods used to assess model uncertainties. The first and second papers, *Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling* (Water Research - 2012) and *Analysis of parameter uncertainty of a flow and quality stormwater model* (Water Science and Technology - 2009) investigate and test the most suitable methods for parameter calibration, model sensitivity and uncertainty analyses that could be used in urban drainage modelling. The third, fourth and fifth papers, *Calibration and sensitivity analysis of stormwater models* (Australian Journal of Water Resources - 2011), *Stormwater quality models: performance and sensitivity analysis* (Water Science and Technology - 2010) and *Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data* (Environmental Modelling and Software - 2011) investigate parameter sensitivity and model uncertainty of a number of urban drainage models using a Bayesian approach. The sixth paper, *Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?* (currently in press in Water Science and Technology) explores possible shortcomings of the Bayesian approach on model sensitivity and uncertainty. The seventh paper, *Impacts of measured data uncertainty on urban stormwater models* (submitted to Journal of Hydrology), presents results of an approach for propagating input and calibration data errors in stormwater models by taking into account the errors in the data sets and investigating their impact on parameter sensitivity and model predictive uncertainty. Finally, concluding remarks are given and possible future work is discussed.

In addition to the seven papers included in the thesis, the candidate co-authored a paper (Deletic et al, 2012) that is included as appendix. This paper is result of a major effort of International Working Group on Data and Models (that works under Joint Committee on Urban Drainage of IWA and IAHR) on development of a framework for assessing uncertainties in urban drainage models. The candidate also produced 6 conference papers (not included in the thesis), that she presented at major international conferences across the world.

Declaration of Publications and Authorship

In accordance with Monash University Doctorate Regulation 17/ Doctor of Philosophy and Master of Philosophy (MPhil) regulations the following declarations are made:

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes seven original papers; five published and two submitted for publication in peer-reviewed journals. The core theme of the thesis is the uncertainty assessment in stormwater models. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Department of Civil Engineering under the joint supervision of Prof. Ana Deletic and Dr. David T. McCarthy.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers, from the same and also different institutions, and acknowledges input into team-based research.

Chapter 4 includes two journal papers. The first co-authored with Ana Deletic and Tim Fletcher, and the second journal paper co-authored with members of the International Working Group on Data and Models (under the IWA/IAHR): Giorgio Mannina, Manfred Kleidorfer, Luca Vezzaro, Malte Henrichs, David T. McCarthy, Gabriele Freni, Wolfgang Rauch and Ana Deletic. **Chapter 5** contains three journal papers; the first co-authored with Ana Deletic, David T. McCarthy and Tim Fletcher and the two remaining co-authored with Ana Deletic, David T. McCarthy, Tim Fletcher, Manfred Kleidorfer and Wolfgang Rauch. **Chapter 6** holds one journal paper co-authored with Ana Deletic and David T. McCarthy. And finally, **Chapter 7** has a journal paper co-authored with Ana Deletic, David T. McCarthy, Manfred Kleidorfer and Wolfgang Rauch. **Appendix A** encloses a journal paper co-authored by the candidate, which relates to the methods used in the research conducted within this thesis. This paper was also a result of an international collaboration project (including 12 authors from different international universities). In each of 7 papers, the candidate's contribution to the work as the first author involved planning and initiation of the study, model runs and analyses of the results, discussion with the co-authors, and writing up the papers. More specifically, my contribution to the work developed in the papers involved the following:

Chapter	Publication Title	Publication status	Nature and extend of candidate's contribution
4	Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling	Published *Water Research*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 35%
	Analysis of parameter uncertainty of a flow and quality stormwater model	Published *Water Science and Technology*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 70%
5	Calibration and sensitivity analysis of stormwater models	Published *Australian Journal of Water Resources*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 70%
	Stormwater quality models: performance and sensitivity analysis	Published *Water Science and Technology*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 60%
	Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data	Published *Environmental Modelling and Software*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 60%
6	Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?	In press *Water Science and Technology*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 75%

Chapter	Publication Title	Publication status	Nature and extend of candidate's contribution
7	Impacts of measured data uncertainty on urban stormwater models	Submitted *Journal of Hydrology*	Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up. 60%

I hereby declare the statement of candidate's contribution to be true and correct.

Cintia B. S. Dotto, 24 May 2013

Table of Contents

Chapter 1

Data and models

1.1 Introduction

Non-point sources of pollution such as runoff from urban stormwater and agricultural areas are currently a significant concern all over the world. Among the diffuse pollution sources, stormwater runoff is one of the major sources of water pollution across urban areas. In the late 90s, runoff from urban areas was the fourth largest contributor of water pollution in rivers and streams in the United States (USEPA, 1998). However, due to investments made in management of point sources of pollution (such as industrial and domestic wastewater), as well as reduction in agricultural runoff pollution, stormwater became the leading source of pollution; e.g. it is recognised to be the number one source of pollution in many areas of the USA (USEPA, 2012). In Australia, the situation is not different, and stormwater is currently the main cause of coastal pollution (Department of Environment and Heritage NSW, 2012). In Melbourne, around 400 GL per year of stormwater flows down drains and into the Yarra River and surrounding creeks, finishing in the Port Phillip Bay (Melbourne Water, 2012). As such, stormwater management is at the forefront of many policies in Australia (as well as other developed countries), with increasing government funding becoming available (Department of Environment and Heritage NSW, 2012; USEPA, 2012). In Australia, the focus is not only on treating stormwater, but also realising its potential as an alternate water source. Indeed, many major Australian cities have just recovered from a severe and extended drought, from which stormwater emerged as viable source of water supply. Therefore, management of stormwater for both pollution protection of receiving waters and as a water resource is becoming regular practice in our cities (Wong et al., 2011). This is known as Water Sensitive Urban Design (WSUD) stormwater management in Australia (City of Melbourne, 2012), or implementation of Sustainable Urban Drainage Systems (SUDS) in the UK, or Low Impact Development (LID) strategies in the USA.

Stormwater models are powerful tools to aid the planning, design, and performance of different stormwater management strategies. Indeed, the Model for Urban Stormwater Improvement Conceptualisation (MUSIC model developed in 2001 for the conceptual assessment of stormwater management - eWater CRC, 2012) has enabled Australia to lead the world WSUD implementation; MUSIC now underpins the decision making process in urban water management, policies and regulation. Similar models exist around the world that are used in a similar way (e.g. SWMM (USEPA, 2007) in USA).

Although these models provide a great platform for decision making, they all have an intrinsic level of uncertainty, regardless of their formulations (e.g. whether they are physically based or purely statistical) (Bertrand-Krajewski et al., 2002). Understanding this uncertainty is important; in fact, incorrect estimates of stormwater flows and pollution concentrations would easily lead to an inadequate design of stormwater management systems. Little is understood about the sources and magnitude of this uncertainty, which could be due to the errors in measured data (input and calibration data) and/or due to the model structure and parameters. As a result, improving models and the confidence in their results requires more robust methodologies for model calibration, sensitivity and uncertainty analysis. These methodologies should not only be able to provide an assessment of the uncertainties in the model's parameter values and an evaluation of the confidence level of the model's predictions, but also be able to identify and propagate the different sources of uncertainties.

1.2 Uncertainties in stormwater models

As stated above, uncertainties are present in all models, yet it is often not addressed because uncertainty analysis is considered to be a difficult and time consuming activity. In most cases, it is avoided among practitioners because it actually reveals that the results are, in fact, highly unreliable (Larssen et al., 2007), or in other cases, the opposite, it is avoided because the uncertainty in the results seems very complex to deal with. In this context, if the different sources of errors compromise the level of accuracy of any model's output, then assessing uncertainties in stormwater models due to different sources of errors is crucial for advancing urban drainage modelling practice.

Typically, four key sources of uncertainties are identified: (1) uncertainty due to calibrated parameter values, (2) errors due to incomplete or biased model structure and (3) random and systematic errors in the measured input data and (4) errors in calibration data (Butts et al., 2004). In terms of stormwater modelling and related fields (e.g. environmental modelling), these sources of uncertainties are interlinked (Beck, 1987; Walker et al., 2003; Kleidorfer et al., 2009), suggesting that assessing the impacts of a single source is not enough and that simultaneous propagation of key sources of uncertainty is required.

As with most models, the calibration of urban drainage models rarely results in one unique parameter set, and instead many equally plausible parameter sets are obtained, which reduces the confidence in modelled results during the prediction period (Kuczera and Parent, 1998). Global sensitivity analysis methods have the advantage of performing uncertainty analysis while providing information about the most likely parameter sets to calibrate the model. However, there is no indication of the most suitable method to assess stormwater models. Therefore, the comparison of different methods to perform parameter calibration, model sensitivity and uncertainty analysis will identify the most suitable for stormwater models.

While a range of models have been applied worldwide to predict flows and pollution generation from stormwater, the assessment of the uncertainty associated with model structure has not been sufficiently explored. Furthermore, there is no standard method to evaluate structural uncertainty. Different approaches to evaluate structural uncertainties must be explored, even if from a heuristic perspective. Exploring parameter calibration, model sensitivity and the resulting predictive uncertainties in models with different levels of complexity will provide information about the models' limitations (including the existence of model structure and conceptual errors).

Measured data such as rainfall, flow rates and pollutant concentrations are required for the application of urban drainage models. However, these measured data are plagued with uncertainties (i.e. due to a range of random and systematic errors). As such, there is a need to understand the impacts of these uncertainties on the performance, sensitivity and predictive uncertainty of stormwater models. This will contribute to further understanding their consequences in the modelling exercise.

1.3 Overall aim

This research improves our understanding of the uncertainties in urban rainfall runoff and pollution generation models in order to define their reliability. The specific aim is to assess uncertainties in stormwater flow and pollution generation models (with different levels of complexity) and the impact of different sources of uncertainties on the models' results and parameter sensitivity. The findings of this research will be useful to inform stormwater management practices, such as: risk assessment, urban planning, design of stormwater facilities, optimisation of data monitoring campaigns (i.e. which kind and how much of data should be prioritised) and issues related to model under and over-parameterisation to the development of more accurate models.

The following are the main objectives of this study:

1. identify suitable method(s) to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models;
2. explore parameter calibration, model sensitivity and the resulting predictive uncertainties in models with different levels of complexity; and,
3. explore the impact of measured input and calibration data uncertainty on the performance, sensitivity and predictive uncertainty of stormwater models.

1.4 Scope of the thesis

The dataset used in this study has been previously collected as part of other research projects. The dataset contains long term and high resolution data on rainfall, flows and TSS and TN concentrations collected at the outlet of five urban catchments around Melbourne, Australia. The work was focused on separate storm drainage systems; i.e. the systems that collect and transport stormwater only (in Australia we do not have combined systems where stormwater and sewage are mixed).

Among the models representing the different processes happening in the catchment, only stormwater flow and pollution generation models are considered in this research. Treatment efficiency models are not included and should be part of future research projects in the field.

While a number of models and approaches to simulate discharges from urban catchments are available, it is not possible to cover all of them. They range from simple empirical equations (usually over-simplified and not able to illustrate the physical process) to complex models (that have a large number of calibration parameters and require a large number of input data). There is a group of conceptual models that falls in between; they do not represent the actual physical processes occurring in the catchment, but include equations describing the concept of the processes occurring in the system. Being widely used in practice (Butler and Davies, 2000; Wagener et al., 2004), conceptual rainfall runoff and pollution generation models ranging from simple to moderate complexity are the focus of this research.

A number of global sensitivity analysis method(s) is applied to a simple rainfall runoff model coupled with a simple pollution generation model. The most suitable methods are identified, however only one is selected to be used through the remaining of the research.

1.5 Outline of the thesis

Chapter 2 provides a review of the published literature, identifies the current research gaps, and presents the objectives and main hypotheses underlined in the present thesis. The review on urban drainage modelling is organised in terms of three major topics: modelling urban stormwater, sources of uncertainties in urban drainage models and assessing uncertainty in urban drainage models.

Chapter 3 provides an overview of the dataset and stormwater models employed in this study. The dataset was collected by former Monash University PhD students, thus only the key aspects of the monitoring are summarised. The selection of the models was done so that both water quantity and quality were included.

Chapter 4 explores different methods for parameter calibration, model sensitivity and uncertainty analyses of urban drainage models. Not only the results with respect to model parameter sensitivity and predictive uncertainty are presented, but the interaction between the complexity of the method used, computational time required and the knowledge/skill level of the modeller is also considered.

Chapter 5 further investigates parameter calibration, model sensitivity and uncertainty analysis in models with different levels of complexity by means of a Bayesian approach. The models' sensitivity to the different parameters is presented and the models' predictive uncertainty, originating from parameter uncertainty, is also reported.

Chapter 6 investigates the main impacts of verifying (or not) the assumed structure of model errors on model parameter sensitivity and associated predictive uncertainty of stormwater models, and it also explores alternative strategy to mitigate such impacts.

Chapter 7 applies a novel method for propagating input and calibration data errors in stormwater models using a Bayesian approach. The results are then presented by means of evaluating the impact of the input and calibration data errors on the sensitivity and predictive uncertainty of stormwater quantity and quality models.

Chapter 8 provides a summary of the key findings, a discussion of the strengths and weaknesses of the thesis and a summary of the areas requiring further investigation.

Appendix A encloses a journal paper co-authored by the candidate, which relates to the methods used in the research conducted within this thesis.

Appendix B includes a glossary with the definition of the terms related to stormwater modelling uncertainty that were widely used throughout this thesis.

1.6 References

Beck, M. B. (1987). Water Quality Modeling: A Review of the Analysis of Uncertainty. *Water Resources Research* **23**(8): 1393-1442.

Bertrand-Krajewski, J. L., Barraud, S. and Bardin, J. P. (2002). Uncertainties, performance indicators and decision aid applied to stormwater facilities. *Urban Water* **4**(2): 163-179.

Butler, D. and Davies, J. W. (2000). *Urban Drainage*, London: E&FN Spon.

Butts, M. B., Payne, J. T., Kristensen, M. and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* **298**(1-4): 242-266.

City of Melbourne (2012). WSUD Guidelines. Applying the Model WSUD Guidelines. An Initiative of the Inner Melbourne Action Plan (http://www.melbourne.vic.gov.au/Sustainability/SavingWater/Pages/Watersensitivedesign.aspx). Accessed in the year 2012.

Department of Environment and Heritage NSW. (2012). Department of Environment and Heritage NSW (http://www.environment.nsw.gov.au/beach/waterpollution.htm). Accessed in the year 2012.

eWater CRC (2012). Model for urban stormwater improvement conceptualisation (MUSIC) (Version 5.0), Canberra: eWater Cooperative Research Centre.

Kleidorfer, M., Deletic, A., Fletcher, T. D. and Rauch, W. (2009). Impact of input data uncertainties on stormwater model parameters. *Water Science and Technology* **60**(6): 1545 - 1554.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* **211**(1-4): 69-85.

Larssen, T., Høgåsen, T. and Cosby, B. J. (2007). Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Ecological Modelling* **207**(1): 22-33.

Melbourne Water (2012). Stormwater pollution (http://www.melbournewater.com.au/content/drainage_and_stormwater/stormwater/stormwater.asp?bhcp=1). Accessed in the year 2012.

USEPA (1998). The Quality of Our Nation's Water: A Summary of the National Water quality Inventory 1998. Report to Congress. US Environmental Protection Agency (USEPA). US Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (2007). Storm Water Management Model (SWMM) - User Manual version 5.0. US Environmental Protection Agency (USEPA). *US Environmental Protection Agency - EPA/600/R-05/039*. Washington DC.

USEPA (2012). US Environmental Protection Agency (http://water.epa.gov/polwaste). Accessed in the year 2012.

Wagener, T., Wheater, H. S. and Gupta, H. V. (2004). *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*. London, UK, Imperial College Press.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluis, J. P., van Asselt, M. B. A., Janssen, P. and Krayer von Krauss, M. P. (2003). Defining uncertainty – A conceptual basis for uncertaintry management in model-based decision support. *Integ. Assessment* **4**(1): 5-17.

Wong, T. H. F., Allen, R., Beringer, J., Brown, R. R., Chaudhri, V., Deletic, A., Fletcher, T. D., Gernjak, W., Hodyl, L., Jakob, C., Reeder, M., Tapper, N. and Walsh, C. (2011). *Blueprint2011 – Stormwater Management in a Water Sensitive City*. Melbourne, Australia, Centre for Water Sensitive Cities.

Chapter 2

Literature review, specific objectives and underlying hypotheses

2.1 Introduction

A large number of flow and pollution generation stormwater models are currently used worldwide. Regardless of whether they are physically based or purely statistical, they all have a high level of uncertainty (Bertrand-Krajewski et al., 2002). However, little is understood about the sources and magnitude of this uncertainty, which could be due to the errors in the measured data (input and calibration data) and/or due to the model structure and parameters. Improving the models and their effectiveness requires more robust methodologies for model calibration, sensitivity and uncertainty analysis. Such methodologies should not only be able to provide an assessment of the uncertainties in the model's parameter values and an evaluation of the confidence level of the model's predictions, but also be able to identify and propagate the different sources of errors.

Within this context, this chapter aims to present a review of literature on uncertainty analysis in urban drainage models and also on the related topics. Firstly, an introduction about stormwater models (their different principles and levels of complexity) is provided and the different sources of uncertainty in such models are then presented. Subsequently, the methods currently used to evaluate model uncertainty are summarised. Finally, the knowledge gaps in the topic are identified and the main research question is introduced.

2.2 Modelling urban stormwater

Stormwater models are essential in urban water management; they enable the quantification of urban discharges and the design of stormwater treatment and harvesting technologies. Moreover, they underpin the decision making process regarding water resource policies and regulations. The standard components of stormwater models are: (a) a rainfall runoff module to generate the runoff from the precipitation excess; (b) a water quality module to estimate the pollutant generation; (c) a transport modelling approach to route flows and pollutants through the system (channels/pipes); and, (d) a treatment module to design and analyse the performance of stormwater treatment and harvesting strategies. The outputs of each module are commonly used as inputs for the next one. For example, modelled flows might be used to estimate pollutant loads, which can be used to design treatment technologies. Rainfall runoff models are currently well developed and widely adopted in practice (Elliott and Trowsdale, 2007); they range from simple empirically-based to complex physical-based models. Contrary to the rainfall runoff models, reliable stormwater pollution generation models are almost non-existent (Elliott and Trowsdale, 2007). The available approaches for modelling water quality range from simple regression equations to conceptual models based on the concepts of build-up and wash-off processes (examples in McAlister et al., 2006).

Stormwater models simulating the catchment's runoff and pollution generation are reviewed in this section. Firstly, a general modelling approach scheme is presented, from which the main modelling components and tasks are identified. Then a description of the main modelling principles is presented. This is followed by an introduction to the common general stormwater modelling protocol (i.e. main tasks involved in the modelling exercise). The review of these general points is important to provide background for the next sub-sections that review the approaches currently used to model flows and pollution generated from urban stormwater.

2.2.1 General modelling concepts

A general modelling approach, presented in Figure 2.1, was adopted to describe the main components and tasks in the modelling exercise. The following components can be identified: model structure, model parameters, measured input and calibration data, calibration algorithms with objective functions and model outputs. In addition, model calibration and application are represented in the framework. This sub-section focuses on describing the main principles related to the model structure and the main required protocol in most of stormwater modelling routines. The remaining components in Figure 2.1 will be described later in this literature review.

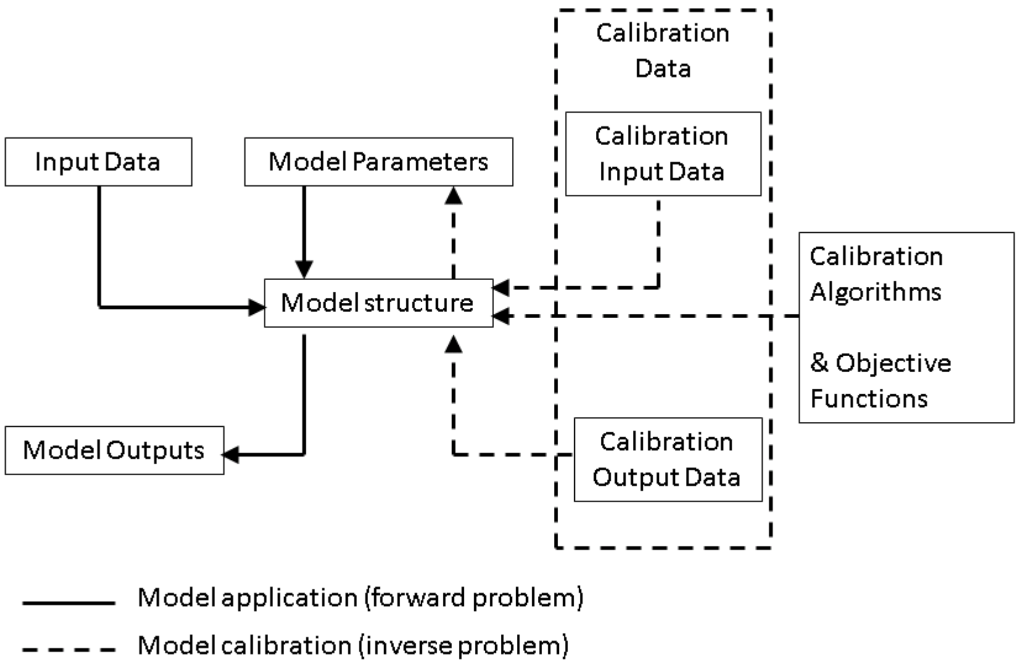


Figure 2.1 General modelling framework (after Deletic et al., 2012).

The different principles regarding the model structure are well known and established between the modelling community, and therefore they are only summarised in this section.

Deterministic models describe a physical process in the catchment in terms of mathematical equations that transform a certain set of input data into outputs. If however, one or more factors within the model (e.g. process description and/or model parameters) have a random nature, the model is classified as **stochastic** (Abbott and Refsgaard, 1996). Contrary to deterministic models, stochastic models will not generate the same outputs even if the same set of input data and/or parameters are used, as one or more variables are randomly sampled from a distribution (Parker et al., 2010). Stochastic models have the advantage of accounting for irreducible uncertainty by attempting to reproduce the natural variability of the processes, which is useful when the process being modelled is not fully understood. On the other hand, they might have limited application for some studies because the random variables are restricted to certain probability distributions (Zoppou, 2001). In addition, stochastic models do not allow a complete control of the user over the results, which imposes limitations for further model calibration, validation and sensitivity analysis. Meanwhile, deterministic models are accurate when the process being modelled is well described by the model. Deterministic models have become the standard approach in many areas (Butler and Davies, 2000).

Empirical models do not represent the physical process of a system and their parameters are not directly related to the physical processes (i.e. parameters have to be inferred by model calibration) (Wagener et al., 2004). **Conceptual** models, on the other hand, include equations that describe at least the concept of the processes occurring in the system. Their parameters might (a) have a physical background and therefore be estimated from measured data, and/or (b) not have any physical background and thus have to be calibrated (Wagener et al., 2004). The most complex group of models are **process based** (Wagener et al., 2004), where mathematical equations represent the actual physical processes occurring in the catchment and their parameters are also physically based (i.e. parameters can be easily estimated from the measured data). This group of models usually generates more accurate outputs (closer to reality). However, their formulations are often complex, their numerical solution might not be explicit (which can lead to numerical instabilities) and they also demand long computational times. Conceptual models are the most used in practice (Butler and Davies, 2000; Wagener et al., 2004) as they are a compromise; they represent the physical processes by simplified concepts and require less input data, a lower level of expertise from the user and a lower level of understanding of the fundamental processes when compared to the process based models.

Lumped models represent the study area as one homogeneous block (Beven, 2001). Only one set of parameters is required to generate the response for the whole area and therefore the spatial variability of the area is ignored. On the contrary, the spatial variability is better represented by **distributed** models, which disaggregate the study area in sub-areas with similar characteristics (Beven, 2001). In general, distributed models can represent the different processes to be modelled, while lumped models assume that all processes are punctual (in space and time). They use mean values for the various processes, and therefore are indicated to be applied with large temporal scales in which a detailed description of the processes is not important (Zoppou, 2001). However, it is often the case that point-based measured data is collected to represent the whole area. In this case, the benefits of subdividing the area are not significant.

In general, models can be applied for a range of temporal scales and timesteps. In the water related fields, **continuous** models simulate the system's response over a period of time (e.g. weeks, months, years) and therefore, account for the overall water balance in the system; continuous models can represent the event antecedent conditions. Models based on **event** simulation can only estimate the system's response from discrete events and cannot account for any between event characteristics (McAlister et al., 2006).

In most cases, the accuracy of stormwater models increases as the modelling **timestep** decreases, and the choice of an inadequate timestep can compromise the model results (Einfalt et al., 2002; McCarthy, 2008; eWater CRC, 2009). For example, if the model is used with a timestep larger than the transient time of the process in the catchment, results will not reflect reality (eWater CRC, 2009).

The majority of models require **calibration** prior to their application, especially those which have an empirical or conceptual structure (refer to Figure 2.1 for a schematic of model calibration in the general modelling framework). It is very unlikely that non-calibrated conceptual models will be able to reproduce reality (Wagener et al., 2004). In practice, urban stormwater models can be used without calibration (Rauch et al., 2002b). This is mainly because of the lack of accurate measurements required to calibrate and evaluate the performance of such models (Bertrand-Krajewski et al., 1993; Gaume et al., 1998). In the cases where sufficient measured data is available and models can be calibrated, the following concepts are used: measured input data is used as the input to the model, which is used to generate the outcomes (output data). The model is then calibrated through calibration algorithms and objective functions that are used to compare the modelled outputs to the measured calibration data. The choice of calibration datasets and objective function is critical as they tune the parameters to characterise different parts of the hydrograph or pollutograph (e.g. low or peak flows) (Diskin and Simon, 1977; Yapo et al., 1996; Madsen et al., 2002; Guinot et al., 2011). In addition, as with most models, the calibration of urban drainage models rarely results in one unique parameter set, and instead many equally plausible parameter sets are obtained. This effect is called equifinality (Beven and Freer, 2001) and is caused by several factors: (i) the parameter space presents several local minima regions; (ii) often the model is not equally sensitive to all the calibration parameters, in fact some models are over-parameterised and present a large number of insensitive parameters; and, (iii) parameters can present a high degree of correlation (usually non-linear interactions as per Wagener et al., 2004). The equifinality (mainly related to the influential parameters) effect drastically reduces the confidence in modelled results (Kuczera and Parent, 1998).

Model **validation** should be performed to verify if the model is able to reproduce the simulated process outside the calibration data (Mourad et al., 2005a; Mourad et al., 2005c). For validation, data management is a major task as the choice of how to split the data for calibration and validation may influence our understanding of the model's predictive ability (McCarthy, 1976; Klemes, 1986; Vaze and Chiew, 2003; Wagener et al., 2004). Finally, the model **application** is the process of using the model with the calibrated parameter sets to predict outside the calibration and validation scenarios (Figure 2.1).

The following two sub-sections use the concepts and processes reviewed in this sub-section to present a review of the different rainfall runoff and water quality models currently adopted.

2.2.2 Review of rainfall runoff models

Empirical models

The Soil Conservation Service curve number, mostly used to predict peak flows (SCS, 1956), and the polynomial equations and time-series methods used to generate the catchment's runoff (Chiew et al., 1993) are examples of empirical models. They have the advantage of being easy to apply and hence no hydrologic knowledge is required. While they produce reasonable results when applied to the simulation of urban developments with monthly and annual volumes, these models are too simple to represent the within-event characteristics of the processes being modelled. In other words, they are not able to reproduce the processes within events and consequently are not suitable for small timesteps (Chiew et al., 1993).

Conceptual models

The runoff is generated by assuming the catchment as a number of interlinked storages with mathematical functions describing the movement of water entering and leaving them (Boyd et al., 1994). These models produce reasonable results even when applied in small timesteps (e.g. sub-daily to few minutes) and range from simple to very complex urban drainage models.

The simplest conceptual models focus on the simulation of the impervious area runoff only, which is modelled as a single reservoir (Schueler, 1994). They are lumped models ideal for the estimation of total volumes, but they might not be very informative when the modeller is interested in a detailed study of the catchment's hydrological processes (e.g. baseflow). Such models are easy to use, computationally fast and usually involve a small number of parameters. The Rational Method, which generates runoff as a function of the rainfall and imperviousness of the catchment (Schueler, 1994) forms the basis for many continuous conceptual models. KAREN (Rauch and Kinzel, 2007) is an example of a conceptual simple model based on the Rational Method (Schueler, 1994). It is a continuous model designed to estimate flows from urban areas and presents only four parameters to be calibrated. The model was recently calibrated for two Australian urban catchments and resulted in reasonable estimates of urban flows, with the Nash-Sutcliffe efficiency coefficients (Nash and Sutcliffe, 1970) higher than 0.6 (Kleidorfer et al., 2009).

There is a range of models that are distributed catchment models and include runoff generation from impervious and pervious surfaces and simulate simplified channel/pipe flow. P8-UCM (P8-

Urban Catchment Model) by Palmstrom and Walker (1990) and MUSIC (Model for Urban Stormwater Improvement Conceptualisation) by eWater CRC (2012) are examples of this category. These models are based on continuous simulation and represent a simplified description of the rainfall runoff processes by a series of reservoirs. P8 UCM works with hourly timesteps and MUSIC allows a minimum of 6 minutes timestep. MUSIC is of particular interest because it was developed for Australian conditions, can operate at a range of temporal and spatial scales and has been widely used in Australia mainly for the conceptual design of drainage systems (Ladson, 2008; Mitchell et al., 2008), in particular treatment technologies (Walsh et al., 2005). While a number of studies contributed to the enhancement of MUSIC's stormwater treatment module (Scholes et al., 2008; Hatt et al., 2009), the rainfall runoff model is usually applied without calibration. Moreover, **validation of MUSIC's rainfall runoff model has not been explored.**

Some studies looked into the comparison of different conceptual models. For example, Chiew et al. (1993) compared six rainfall/runoff models, which ranged from a simple polynomial equation to a more complex conceptual model (MODHYDROLOG - Chiew and McMahon, 1994). They compared the models in terms of daily, monthly and annual volumes. Their results are in agreement with the fact that simple models can only be applied for the large timesteps (monthly or annual). The conceptual models were able to predict total daily flows and the rather complex MODHYDROLOG was the only one able to depict the low flows in the catchments. **The performance of these different conceptual model structures in sub-daily timesteps has not been explored.**

There is another group of more complex models, which contain representations of surface runoff, subsurface flow, evapotranspiration, and channel flow. However, they can be far more complicated due to their complex nature. MOUSE (DHI, 2004), Infoworks (Wallingford Software, 2009), CANOE (INSA/SOGREAH, 1999) and SWMM (USEPA, 2007) are examples of such models. They model both complex surface runoff and channel/pipe flow, including transition from unpressurised to pressurised pipe flow. These models can be applied to a wide range of temporal and spatial distributions and are suitable for a vast range of applications. However, they are probably too complex to be used by the general public or non-modelling professionals (Elliott and Trowsdale, 2007). They have the option of different concepts to estimate the catchment's hydraulic responses. These conceptual complex models range from linear reservoir routing routines to solving the full Saint-Venant equations for dynamic wave routing. As any other type of model, these are unlikely to perform well if not calibrated, but unlikely the simpler ones, these present a large number of parameters. Moreover, the required input data (topographic, geological, climatic, etc) is not very often available, mainly for practical industrial application.

Elliott and Trowsdale (2007) reviewed the qualitative strengths, weakness and potential uses of ten widely used conceptual urban stormwater models. They demonstrated that the popular MOUSE

(DHI, 2002c; 2004), MUSIC (eWater CRC, 2012), SWMM (USEPA, 2007) and P8-UCM (Palmstrom and Walker, 1990) are the only spatially distributed models with a link-node drainage network. Moreover, only MOUSE (DHI, 2002c; 2004), SWMM (USEPA, 2007) and MUSIC (eWater CRC, 2012) showed the capability of predicting flow rates from small catchments due to their distributed sub-hourly temporal scale. On the other hand, routing flow through the channel/pipe is available in the majority of models. In terms of potential use, MOUSE (DHI, 2002c; 2004) and SWMM (USEPA, 2007) were the most suitable for a large range of applications with the disadvantage of being complex, requiring a large amount of data and parameters to be measured or calibrated, and hence unlikely to be used by the practitioners.

The effects of long term simulations with the complex CANOE (INSA/SOGREAH, 1999) and the simple KAREN (Rauch and Kinzel, 2007) on the design of CSO structures and storage tanks was studied by Gamerith (2006). Interestingly, both models, with different levels of complexity, generated similar results. The author argued that this result was due to the simplicity of the sewer system and the large size of the sewer sections, which prevented flooding. Such result is in agreement with previous literature on hydrological models, which suggests that conceptual models are suitable for predicting steamflow, volumes or loads at the catchment scale (Refsgaard and Knudsen, 1996; Rauch et al., 2002b), while more complex cases, for example where spatial representation is important, would require more highly parameterised complex models (Refsgaard and Abbott, 1996).

Summary

While the empirical equations are usually over-simplified and cannot describe the physical process, the complex models require a large amount of input data and have a large number of (a) physical parameesters to be measured in the field or (b) conceptual parameters to be calibrated. Moreover, these data and parameter sets will only be valid for a specific area and thus, the models are not transferable for other catchments. In this context, conceptual models are preferred and are currently the most used in the field (Butler and Davies, 2000; Wagener et al., 2004)**. However, the assessment of these models' structures and their associated predictive uncertainty under the same conditions has not been explored.**

2.2.3 Review of pollution generation models

Stormwater quality models are important tools to control pollution generation, evaluate pollutant loads and investigate and rank alternative approaches for stormwater quality management (Ahyerre et al., 1998; Marshall et al., 2005). Different approaches are available to attempt reproducing the response of urban catchments in terms of pollutants generation (Huber, 1985; Huber, 1986):

- event mean concentration (EMC) models;
- continuous stochastic models;

- empirical regression models;
- process based build-up/wash-off models; and,
- physical models.

The current most used softwares such as P8-UCM (Palmstrom and Walker, 1990), XP-AQUALM (XP-SOFTWARE, 1999), SWMM (USEPA, 2007) and MUSIC (eWater CRC, 2009) usually include one or more of these approaches. The different approaches are reviewed in this section.

Event Mean Concentration (EMC) Models

These are the simplest modelling approaches for estimating pollution generation in urban drainage. Such group of model uses monitored data to associate the catchment's physical or hydrological characteristics with measured concentrations and loads (Charbeneau and Barrett, 1998) and assume that the constituent concentration is well represented by an event mean concentration (i.e. constant throughout an event). It can relate the parameters of an event (e.g. average rainfall intensity, rainfall total, etc) and climatic parameters (e.g. antecedent dry weather period, average previous day temperature, etc) to an EMC for a certain pollutant (e.g. Duncan, 1995; 1999; McCarthy, 2008; Dembélé et al., 2010). Duncan (2005), for example, reported a power relationship between the event load and the rainfall intensity. McCarthy (2008) found a strong relationship between *E. coli* EMCs and antecedent temperatures (and evaporation). EMC based models statistically describe the long-term pollution generation process in the system and thus, are able to evaluate the long-term impact of pollutants in the receiving bodies, i.e. annual loads can be estimated (Charbeneau and Barrett, 1998; Francey, 2010). The disadvantages of such models are that they do not represent any pollutant processes in the catchment and also they are not easily transferable to other areas without extensive data collection/calibration. This is due to the great variance of EMC values between sites, even when they have similar characteristics.

Stochastic and semi-stochastic models

These models are commonly used to represent the stochastic nature of pollutant inputs into the system, or our lack of understanding of the process (Butler and Davies, 2000; Rossi et al., 2005). For example, in MUSIC (eWater CRC, 2009) the pollutant distribution is defined by specifying the mean and standard deviation of a log-normal distribution, from which concentrations are stochastically generated for each timestep. XP-Storm (XP-SOFTWARE, 1977) and SWMM (USEPA, 2007) use a similar approach, in which a random EMC value is sampled from the measured data distribution of the pollutant at the beginning of each event; this value is then used as a constant throughout the specific event. A similar approach is available in XP-AQUALM (XP-SOFTWARE, 1999) for estimating the daily loads in the catchment. In addition to the group of semi-stochastic models there is a new approach proposed by Bach et al. (2010), in which water quality concentrations are predicted using the first flush theory.

Regression models

Simple or multiple regression models are another group of models to quantify discrete pollutant concentration in the catchment (Letcher et al., 2002). The wash-off rating curves proposed by Duncan (1995) are examples of this group. They can be used as event based or continuous models. These models generally relate the pollutant concentration or loads to some storm characteristics such as rainfall intensity, total rainfall and flow (Vaze and Chiew, 2003; Mourad et al., 2005a). For example, the kinetic energy of falling raindrops in the detachment of surface pollutants is considered in the power relationship between the event loads and the rainfall intensities (Duncan, 1995). Variations of this model include, for instance, the replacement of the rainfall intensities by the catchment runoff, in which the transport of the pollutants is represented by the shear stress generated by flow (Letcher et al., 2002). Vaze and Chiew (2003) and Francey (2010) tested a number of simple regression equations to estimate event pollutant loads from impervious surfaces. Results on total loads were reported to be promising when the models were tested in terms of TSS and its associated pollutants, TN and TP. Several popular stormwater models have the option of regression equations in their algorithms: XP-AQUALM (XP-SOFTWARE, 1999), SWMM (USEPA, 2007) and P8-UCM (Palmstrom and Walker, 1990). The drawback of such approach is that no build-up consideration is made and hence, the accumulation of pollutants on the surface is neither considered nor characterised. In addition, the unexplained variability in concentrations and loads are so large that the predictive power of deterministic empirical models becomes questionable (Vogel et al., 2005; Shaw et al., 2010).

Process-based build-up/wash-off models

The first build-up/wash-off model was proposed by Sartor and Boyd (1972) and its variations are currently used in some of the popular stormwater models, such as SWMM (USEPA, 2007) and MOUSE (DHI, 2004). In addition, these models have been subject of a number of studies (e.g. Deletic et al., 2000; Kanso et al., 2006; Hossain et al., 2010; Shaw et al., 2010). For instance, Kanso et al. (2006) tested variations of the original model in terms of TSS concentrations for two urban catchments in France. It was concluded that the model was unable to represent the complexity of the system at the scale of urban sub-catchments. In contrast, Gaume et al. (1998) used a build-up/wash-off model to simulate TSS concentrations from stormwater in an urban catchment and reported a good agreement between measured and modelled values. Vaze and Chiew (2003) compared a number of build-up/wash-off models to estimate event pollutant loads from impervious surfaces. Between 14 and 20 events, with rainfall, flow and concentration data from three urban catchments in Australia were used in the study. Their results indicated that, once calibrated, both approaches estimated event pollutant loads satisfactorily. The predictive power of these models in terms of concentration has been less explored.

Physical models

There are also attempts to develop physical models (e.g. Shaw et al., 2006), which have the advantage of better characterising the pollutant transport process within the catchments. These are not quite established and therefore their application is very limited. For example, a model based on the advection dispersion equation was developed by Mannina and Viviani (2010). The results suggested the model is promising in terms of modelling approach. However, the model with eleven parameters was calibrated with only five events, which makes it difficult to assess the quality of the model and highlights the importance of adequate datasets for model calibration. This group of models can only be tested and validated if a comprehensive measured input data set is available, which is usually not the case for water quality data.

Model application –temporal scale

In general, models can be applied for a range of temporal scales and timesteps. In the water related fields, **continuous** models simulate the system's response over a period of time (e.g. weeks, months, years) and therefore, account for the overall water balance in the system; continuous models can represent the event antecedent conditions. Models based on **event** simulation can only estimate the system's response from discrete events and cannot account for any between event characteristics (McAlister et al., 2006).

In most cases, the accuracy of stormwater models increases with as the modelling **timestep** decreases, and the choice of an inadequate timestep can compromise the model results (Einfalt et al., 2002; McCarthy, 2008; eWater CRC, 2009). For example, if the model is used with a timestep larger than the transient time of the process in the catchment, results will not reflect reality (eWater CRC, 2009).

Summary

While EMC based and stochastic models are adopted as an option in a number of stormwater models, they do not provide information about the physical processes of pollutants in the catchment. Simple regressions that relate pollutant concentration to some hydrological variable to model wash-off and process based build-up/wash-off models are preferred as they are not overly complex, yet attempt to reproduce the main processes of pollutants in the system. Due to the high variability of stormwater quality processes and pollutants both between and within events, continuous simulation is recommended over event based (Hossain et al., 2010; Shaw et al., 2010).

2.2.4 Summary

There are a number of approaches to represent the rainfall runoff and pollutant generation processes in urban stormwater. However, the physical process occurring in the systems are plagued with so many uncertainties that the ability of models to represent reality is very limited (Mourad et al., 2005a).

The general rainfall runoff mechanisms in the catchment are understood and there is a plethora of rainfall runoff models available to predict general figures as mean annual flows and total volumes, and also a number of models are formulated to quantify and represent the processes within and between events (Elliott and Trowsdale, 2007). The conceptualization errors in this models are related either to an over simplification of physical processes or to the complexity of some formulations used in physical models (Elliott and Trowsdale, 2007).

The context of water quality is much worse (Ahyerre et al., 1998; Rauch et al., 2002a). The mechanisms governing the processes and dynamics of accumulation and wash-off of pollutants are not well understood, especially in the sources and processes of pollution generation in drainage systems (Kanso et al., 2005; Rossi et al., 2005; Park and Roesner, 2012). The sum of all these factors leads to large model conceptualization errors, which results in models with low accuracy and a high level of uncertainty (Ahyerre et al., 1998; Bertrand-Krajewski and Bardin, 2002).

In this context, the uncertainties related to the misfit between measured and model data due to the model structure are immense and should be addressed (Butts et al., 2004; Refsgaard et al., 2006; Doherty and Welter, 2010). However, uncertainties were often ignored in the urban drainage field and only recently they have been addressed (Vezzaro and Mikkelsen, 2011; Vezzaro and Mikkelsen, 2012). As such, i**t is clear that a complete, but yet simple, exercise of assessing model uncertainty through rigorous parameter calibration, model sensitivity and estimation of the uncertainty associated with the model's predictions is required (Beven and Binley, 1992; Kuczera and Parent, 1998; Wagener et al., 2004).** The following section reviews the different sources of uncertainties associated with modelling urban drainage and related fields.

2.3 Sources of uncertainties in urban drainage models

2.3.1 Introduction

During the last decades, a number of studies have been done on the uncertainty associated to groundwater, wastewater, environmental and hydrological modelling (O'Donnell and Canedo, 1980; Canale and Seo, 1996; Krzysztofowicz and Kelly, 2000; Refsgaard, 2000; McIntyre et al., 2005; Belia et al., 2009). But the uncertainty associated with urban drainage modelling was only recently approached (Kanso et al., 2006; Vezzaro and Mikkelsen, 2012). To advance the uncertainty analysis in stormwater model, it is important to understand the different sources of uncertainties. This sub-section introduces the main sources of uncertainties in urban drainage models and related fields.

The sources of uncertainties in environmental and hydrological models were extensively mapped in the literature (Beck, 1987; Melching, 1995; Refsgaard, 2000; Walker et al., 2003; Wagener et al., 2004; Gourley and Vieux, 2006). While these sources have been named or grouped differently in some of the mentioned studies, the content was the same. The general modelling framework in Figure 2.1 can be used to help map those sources of uncertainties in urban drainage models:

i. **model structure uncertainty** is the limitation of the model structure in representing real physical processes, usually because of a lack of knowledge about the process being modelled, the model assumptions and boundary conditions. In addition, models usually reproduce different aspects of the system with different parameters, which often leads to high level of parameter interaction.

ii. **measured data uncertainty** is related to the errors in data measurements, including collection, handling and post-processing of model input and calibration data. It can also relate to some variables that are "estimated" rather than measured data (e.g. catchment area);

iii. **model calibration uncertainty** is about the methods used for calibration, also including the selection of the objective functions and data selection; and,

iv. **model calibration parameter uncertainty** relates to the calibrated model parameters;

The following sub-sections describe each of them in more detail.

2.3.2 Model structure

The causes of this uncertainty are numerous and include: conceptualisation errors, such scale-issues or omitting key processes; equations, which could be ill posed and thus inadequately represent the process; and, numerical methods and boundary conditions, which can be ill defined leading to inaccurate solutions (Refsgaard et al., 2006).

Renard et al. (2008) presented a Bayesian based framework that is promising in quantifying uncertainties arising from structural errors (among other sources). However, its application is not straightforward and very computationally demanding (Renard et al., 2008). As an attempt to deal with structural errors, Refsgaard et al. (2006) and Wagener et al. (2003) proposed frameworks to assess uncertainty due to model structure errors. The proposed schemes differ about method and criteria, but both suggest that comparing model structures is the way of assessing this source of uncertainty. Nevertheless, the causes of this source of uncertainty are very complex, and there is no generic approach to evaluate model structure uncertainty (Refsgaard et al., 2006; Doherty and Welter, 2010).

It is also true that model validation is very important to assess the model's efficiency in simulating specific physical processes outside the calibration period (McCarthy, 2008), and therefore could also be used as an indication of the model structure error. The assessment of the model structure is sometimes performed by accepting or rejecting a model structure depending on the number of observations covered in the predictive uncertainty bands (Freni et al., 2009). However, this method is too subjective, as it depends on how wide confidence bands are (e.g. 65%, 95% or 99% confidence interval) (Refsgaard and Henriksen, 2004). In addition, is rejecting the model the best way to go? There might be cases that results from less accurate models are still useful (Refsgaard and Henriksen, 2004).

In summary, it is suggested that comparing results from different model structures when applied to the same case study could be a starting point to evaluate model structure uncertainty in stormwater models.

2.3.3 Measured data

This source refers to the uncertainty in any measured or estimated data used as model input or for model calibration. Independent of the model type, all models require some input measured dataset which will inherently contain a certain degree of error. Furthermore, to calibrate models, another measured dataset is required, which also includes some error. Finally, models often require extra information as input to the model, which also contains a degree of uncertainty: e.g. catchment characteristics (e.g. area and slope) and/or extra climate data (e.g. evapotranspiration and antecedent dry weather periods). Uncertainties in the measured data are generally caused by (i) systematic and/or (ii) random errors. The following paragraphs outline the uncertainties inherent in the major data sources used in rainfall runoff and water quality modelling of stormwater systems.

The most common input to rainfall runoff and water quality models is rainfall intensity, and it is usually required in time-series format (Achleitner et al., 2007; McCarthy, 2008; eWater CRC, 2012). Tipping bucket rainfall gauges are the standard and most used device for measuring rainfall data (Sevruk, 2002). The main sources of uncertainties in the data measured with these gauges are related to both catching and counting errors (Molini et al., 2005b). While splashing losses were found to be only up to 2% and evaporation losses were up to 4%, the wind losses were found to be inversely proportional to the rain intensity and were up to 30% for rainfall intensities around 0.25 mm/h (Sevruk, 1982). Battery, logger and computer clock failures are significant source of errors in rainfall measurements. Time drifts are inherent to any battery controlling logging devices and values around 0.07 min/day were reported by McCarthy (2008). The spatial variability of rainfall is another issue. It is common that the point rainfall measured with the tipping bucket is different from the average rainfall calculated if several gauges were installed along the catchment. To address spatial rainfall distribution radar rainfall data can be used. But this also requires a calibration on rain gauge measurements to reduce the radar uncertainties (Einfalt et al., 2004).

Measured flow data is often used for the calibration of rainfall runoff models (e.g. Sherman, 1932; Marshall et al., 2004; Refsgaard et al., 2006; Huard and Mailhot, 2008), and is sometimes used as input into water quality models (e.g. Kanso et al., 2003; Vaze and Chiew, 2003). The errors in flow data are usually related to the measurement equipment and installation methods. Flow measurement uncertainties for the velocity-area method are caused by the uncertainties in the estimation of the channel's cross section and velocity estimates. The Law of Propagation of

Uncertainty (LPU) (Taylor and Kuyatt, 1994), which is used to propagate and combine individual sources of uncertainties, can be used to estimate this uncertainty associated with the measured flows (Bertrand-Krajewski and Muste, 2007). Values of ±20% were reported in the literature for flows with the velocity-area method (Ahyerre et al., 1998; Harmel et al., 2006a). Uncertainties in flow measurements due to systematic errors were not explored. They are related to the height measurement and an inaccurate velocity calibration or incorrect probe set-up (McCarthy, 2008).

To adequately calibrate stormwater quality models, water quality samples need to be taken and analysed for the parameter of interest (or in-situ probes are utilised). However, monitoring stormwater quality is plagued by a wide range of errors, including those related to sampling, storage and analytical /laboratorial analysis (Harmel et al., 2006a). While the errors relating to sampling methods (i.e. the process of actually extracting the sample from the water source) are significant for TSS measurements (e.g. up to 33%), they are often less significant for dissolved pollutants (Harmel et al., 2006a). Some pollutants are also impacted by storage uncertainties; for example, uncertainties of up to 50% were found for TN concentrations, even for those samples which were stored appropriately (iced <6hrs) (Kotlash and Chessman, 1998). Uncertainty related to the laboratorial analysis was less explored, but values from -9.8 % to 5.1 % have been reported for TSS (Harmel et al., 2006a) while 10.4 % has been reported for TN (Donohue and Irvine, 2008).

In general, uncertainties in measured input and calibration data can be characterised and assessed according to international standards as ISO (1993; 1995; 2007; 2008; 2009a; b). In these standards, uncertainty is defined as the variable associated with a measurement result which characterises the dispersion of the values that could be reasonably attributed to the measured variable, LPU (Taylor and Kuyatt, 1994) is an example.

Errors in measured data could strongly impact the model outputs (Andreassian et al., 2001; Haydon and Deletic, 2009). Mainly systematic errors would propagate the error over an over through the model (Ahmad et al., 2010). For example, if the rainfall data is constantly over- or under-estimated and/or the data logger is suffering from time drifts, the modelled flows or pollutant concentrations would systematically suffer/respond for these errors. Errors in discrete pollutant concentrations can lead to an inadequate design of stormwater treatment technologies (Vaze and Chiew, 2003).

2.3.4 Model calibration

Introduction

The confidence of the model outputs relates to the model uncertainty remaining after the model has been calibrated. Therefore, there is a need for robust and reliable automatic calibration procedures (Beven and Freer, 2001; Moore and Doherty, 2005). However, even when using complex algorithms, which are capable of calibrating highly non-linear functions, there is never certainty that the best solution (or global optimum) will always be found (Beven and Freer, 2001; Wagener et al., 2004).

Auto-calibration methods

Two approaches are extensively used for the estimation of parameters. These are the frequentist and Bayesian approaches. In the frequentist approach, unknown parameters are assumed as fixed and a calibration dataset is used to estimate their values. One disadvantage of such methods is that they usually depend on the initial parameter values assumed at the beginning of the calibration process, which might compromise the optimisation by not searching all the possibilities in the parameter space. Old optimization tools such as downhill simplex method (Nelder and Mead, 1965), the pattern search method (Hooke and Jeeves, 1961) and the rotating directions method (Rosenbrock, 1960) can easily fail because their search for the best estimates can considerably float according to the choice of the initial values and therefore get stuck in some random local minimum. Some gradient based methods that start calibration from different points in the parameter space, which are selected in a manner that minimizes the chance of finding the same local minimum twice, can overcome the problem of the objective function surface in parameter space being pitted with local minima. PEST (Doherty, 2004) is one example of this type of tool. Many studies reported successful applications of the software for calibrating conceptual hydrological models (Doherty and Johnston, 2003; Kunstmann et al., 2006; Skahill and Doherty, 2006).

The Metropolis algorithm (Metropolis et al., 1953), a general Monte Carlo Markov Chain (MCMC) sampling method, has also been widely used for model calibration and sensitivity analysis of models in related fields (e.g. conceptual hydrological models - Kuczera and Parent (1998) and Feyen et al. (2007) and water distribution hydraulic models - Kapelan et al. (2007)). Contrary to frequentist approaches, the Metropolis algorithm identifies not only a best parameter set, but a probability distribution of parameters according to measured data; it estimates the true posterior probability distribution of parameters, which may differ significantly from the multinormal distributions used in classical parameter uncertainty estimation methods. This is a major advantage of this method, as it can overcome the identifiability problem (Kuczera and Parent, 1998). In addition, it is possible to account for model uncertainty while evaluating model performance.

Objective function

The choice of appropriate objective functions is a fundamental consideration when estimating model parameters. Different objective functions influence the calibrated parameter distributions and the uncertainty of model predictions. All objective functions sacrifice the fit of a certain portion of the dataset to achieve a good performance in another portion (Diskin and Simon, 1977; Sorooshian et al., 1983; Servat and Dezetter, 1991; Wagener et al., 2004). It is a common view that the selection of the most appropriate objective function is not an easy task and should reflect the modelling aims (Madsen et al., 2002; Krause et al., 2005). In addition, Croke (2009) recommended that objective functions should consider the uncertainties in the measured and modelled data. As such, he suggested that accounting for the non-homoscedastic and serial correlation of the model residuals (i.e. misfit between the measured and modelled data) can reduce the uncertainty in the estimated parameters and improve the ability of evaluating the model performance. In the same context, Doherty and Welter (2010) added that the choice of objective function, rather than the exact nature of the statistical characterisation of model-to-measurement misfit, becomes an issue of critical importance in the model calibration process. In particular, the choice of an objective function that tunes a model to make predictions of a certain type can lower the uncertainty associated with predictions of that type. The potential and benefits of multi-objective methods to calibrate models have been investigated and a review of these approaches can be found in Efstratiadis and Koutsoyiannis, (2010). Multiple objective functions have some disadvantages as increasing the number of objective functions can turn the problem of model calibration into a decision-making process (Khu and Madsen, 2005).

Least square based objective functions place emphasis on medium/large values, which are often the goals for stormwater management practices (i.e. high volumes - e.g. Chiew and McMahon, 1999). In addition, its statistical background is rather simple, which is the reason why they are still the most adopted functions in the field (e.g. Feyen et al., 2007; Freni et al., 2009).

Calibration data availability

While the uncertainty in measured calibration data was covered in the previous section, the relationship between model uncertainty and the data availability for calibration and validation is addressed here. Model predictions depend on calibration and calibration depends on data. For example, Gaume et al. (1998) used a build-up/wash-off model to simulate TSS concentrations from stormwater in an urban catchment and reported a good agreement between measured and modelled values. The limitation was that the authors used eight events, all over summer, to calibrate the model (with four calibration parameters). It could be argued that if their events were more evenly distributed along the year (or more events evenly distributed through the year) the model would behave differently. Similarly Rodríguez et al. (2010), obtained satisfactory TSS estimations when applying different build-up/wash-off formulations for a small urban catchment in Bogota, Colombia. Nevertheless, only two events were taken into account for calibration.

In addition, the influence of the calibration data availability is reflected in the uncertainty of a model's prediction outside the calibration period (Vaze and Chiew, 2003; Mourad et al., 2005b), and also on model's parameter probability distributions (Larssen et al., 2007). For example, Mourad et al. (2005a; 2005c) found that the commonly used build-up/wash-off and multiple regression models were sensitive to the amount of calibration data (i.e. number of events) and that more data should be allocated for model calibration in comparison to the amount of data for model validation. An alternative to decrease this dependence is to use datasets that include different contexts, such as data from different seasons, to better represent the natural variability of the natural process (Bertrand-Krajewski, 2007; Renard et al., 2008). However, the objective functions do not endow the model with the ability to make all types of predictions equally, and therefore it is unlikely that the model will be able to capture the differences.

More recently, Sun and Bertrand-Krajewski (2012) proposed few methods to select representative calibration datasets in order to optimise the performance of selected regression models. They concluded that the method that uses multiple-dimension information of the measured inputs is more effective than the methods that consider only one-dimension information of the outputs. In the context of flows, Yapo et al. (1996) concluded that sensitivity of the calibration to the data length (e.g. data availability) and period selection (e.g. wet or dry periods) was determined by the different objective functions. Boughton (2007) found that the estimates for long term runoff are more data-dependent than model-dependent when short periods of data are used for calibration. It is important to remember that such finding might be study or data dependent and not a general finding, as the model structure does represent a significant source of errors in models.

In summary, some studies have examined how to divide the available data into calibration and validation sets (McCarthy, 1976; Klemes, 1986; Vaze and Chiew, 2003; Wagener et al., 2004). No specific protocol was found yet, but the overall suggestion is that datasets should include data covering a range of intrinsic content (e.g. different climate seasons and hydrologic behaviours).

Variables to be calibrated

Most water quality models have been calibrated against fluxes and total loads from urban catchments (Sriananthakumar and Codner, 1992; Charbeneau and Barrett, 1998; Dembélé, 2010; Francey, 2010). However, fluxes are driven by flow rates, which generate a degree of spurious correlation in load based models, and therefore mask the real predictive capability of the models (Vogel et al., 2005; Shivers and Moglen, 2008; McCarthy et al., 2011). Moreover, false conclusions about the intrinsic characteristics of pollutants and their sources could be easily derived. For these reasons, estimation of concentrations rather than fluxes would be indicated. Although few studies investigated pollutant concentrations modelling, they usually utilised a very limited number of events.

Summary

The choice among different auto-calibration methods, objective functions, calibration datasets and variables to be calibrated, impacts the model results and uncertainties, and therefore influences the model calibration parameters.

2.3.5 Model calibration parameters

Understanding the uncertainties associated with the stormwater model parameters is crucial for advancing urban drainage modelling practice. For example, the high level of uncertainty in the calibrated parameters has been recognised as one of the main problems in the establishment of water quality models (Kanso et al., 2003; Kanso et al., 2005). The parameter uncertainty may result from (1) a poor fit between model outcomes and measured data (Yapo et al., 1996), (2) a high level of parameter correlation (Vanrolleghem and Keesman, 1996; Lindenschmidt, 2006); and/or, (3) the insensitivity or practical identifiability problem (Vanrolleghem et al., 1995).

The model sensitivity to its parameters is usually determined by either local (Deletic, 2001; Haydon and Deletic, 2007) or global sensitivity methods (Beven and Binley, 1992; Kuczera and Parent, 1998). While local sensitivity methods can identify how the model results change with the different parameter values, they usually do not provide information about the global influence of the different sets of model parameters (Saltelli, 2005). Global sensitive methodologies are preferred as they allow all parameters to vary simultaneously over a wide range of possible parameter values (Neumann et al., 2009; Varella et al., 2010; Vezzaro and Mikkelsen, 2012) and provides information not only about the different parameter sets, but also about parameter interaction. In addition, they can provide insights about the model structure, because most of the methods valuate the model sensitivity while quantifying the uncertainty associated with the parameters (Kuczera and Parent, 1998). Model sensitivity analysis is one of the main interests of this research and is further reviewed in detail in Section 2.4.2.

2.3.6 Summary

Measured data used for input and calibration of stormwater models are not free of errors. Therefore, it is likely that the model's responses will reflect those uncertainties and consequently, the model's calibration parameters will also be impacted. In addition, the combination of different calibration methods and objective functions determines which parts of the dataset the model is tuned to reproduce and these modelling setup choices impact the values of the calibrated parameters. As such it seems clear that the sources of uncertainties in models are interlinked and eventually they all impact the model parameters. It is hypothesized that some of these sources can add-up or compensate for each other, suggesting that assessing the impact of a single source is not enough and that simultaneous propagation of key sources of uncertainties is required.

2.4 Assessing uncertainty in urban drainage models

2.4.1 Introduction

Model sensitivity and uncertainty analysis are imperative prior to model application. While the uncertainty analysis quantifies the uncertainty in the model results, sensitivity analysis complement the uncertainty analysis by providing information about the importance and relevance of model parameters in determining the change in the results (model outputs). This section introduces and compares the most common adopted methods for model sensitivity and uncertainty analysis in urban drainage models and related fields.

2.4.2 Model sensitivity

Model sensitivity analysis reveals how sensitive the model outputs are to each parameter or input. The results can be used just to screen for the most important parameters (Weijers and Vanrolleghem, 1997; Reichl et al., 2006; Haydon and Deletic, 2007; Kleidorfer et al., 2009) or, as in most cases, model sensitivity results can be used to estimate confidence intervals around the model's results (Feyen et al., 2007; Yang et al., 2008; Li et al., 2010). This sub-section reviews the most popular sensitivity analysis methods in the area and related fields.

Global uncertainty analysis

Adequate sensitivity analyses have to include an investigation over the full range of plausible parameter values and their interactions (Saltelli, 2005). In summary, global sensitivity analysis is the assessment of how the variation in the model outputs can be assigned to the uncertainty in the model parameters (Vezzaro and Mikkelsen, 2012). Methods including the variance based measures and Monte Carlo approaches (Ratto et al., 2001; Dorini et al., 2011) have been used in different fields (Saltelli, 2005). During the last decades, a number of studies investigated the uncertainty associated with groundwater, environmental and hydrological modelling (Beck, 1987; Beven and Binley, 1992; Canale and Seo, 1996; Krzysztofowicz and Kelly, 2000; Refsgaard, 2000; Reichert and Vanrolleghem, 2001; Refsgaard et al., 2007). McIntyre et al. (2005) analysed uncertainty in a semi-distributed catchment nutrient model, focusing on the spatial significance of parameters and model outputs, and associated uncertainties. Results suggested that even the most influential parameters suffered from high uncertainty due to (i) spatial inconsistencies in the estimated optimum values; (ii) the sampling error associated with the calibration method; and, (ii) parameter equifinality. A number of studies compiled and qualitatively compared different methods used in integrated environmental modelling (Matott et al., 2009). Makowski et al. (2002), Willems (2008) and Yang et al. (2008) compared the application of different uncertainty analysis techniques in different fields and concluded that modellers should choose the method which is most suitable for the system they are modelling (e.g. complexity of the model's structure including the number of parameters), their skill and knowledge level and the purpose of their study.

Uncertainty associated with urban drainage modelling was only recently investigated (Kanso et al., 2006; Kleidorfer et al., 2009; Lindblom et al., 2011; Vezzaro et al., in press). Some studies have assessed the impact of uncertainties in model parameters (e.g. Kanso et al., 2003; Thorndahl et al., 2008). The key methods and concepts already used in water resources modelling were adopted for urban drainage models and many methodologies (some packed in software tools) are now available to evaluate the model sensitivity, while calibrating and quantifying the uncertainty associated with the parameters. They range from formal Bayesian approaches (Bayes, 1763) as the Markov Chain

Monte Carlo (MCMC) approaches (e.g. MICA by Doherty (2003), or DREAM by Vrugt et al. (2008)) to less formal likelihood methods as the Generalized Likelihood Uncertainty Estimation (GLUE) by Beven and Binley, (1992).

Bayesian inference based on MCMC methods express the uncertainties associated with parameters and model outputs in terms of probability. Samples are generated from the Markov Chains, which will converge to the posterior distribution of the parameters. One of the most used MCMC methods is the Metropolis-Hasting algorithm (Hastings, 1970), which uses an adaptive proposal distribution to sample parameters and is thus better at finding the high posterior density region. Its effectiveness is now well established (e.g. Bates and Campbell, 2001). GLUE has largely been applied to uncertainty assessment in general hydrological models (e.g. Montanari, 2005). The principle of GLUE is to generate parameter samples from a uniform distribution in order to provide a scan of the parameters' space. The method requires a large number of Monte Carlo simulations, while the criteria for accepting a parameter set (the choice of a certain threshold, usually based on a measure of the model performance, that defines which of the sampled parameter sets will be considered for further analysis) is subjective and is defined by the user. In addition, the results obtained with GLUE are very sensitive to this acceptance threshold, which places some limitation in the application of such methodology (Mantovan and Todini, 2006; Freni et al., 2008).

Many likelihood functions used in some Bayesian approaches assume that the model errors (or residuals between the measured and modelled values) are normally distributed. However, this assumption is often not checked; this is the case for both scientific literature (Maksimovic et al., 1991; Kanso et al., 2006; Varella et al., 2010) and modelling practitioners, who are often not fully acquainted with uncertainty procedures. In the cases where these assumptions are checked, it is commonly found that the error does not follow any specific distribution and the results are still presented 'as is'. In the literature, a transformation of measured and modelled data (e.g. log or Box-Cox transformation) is used by some modellers to ensure they meet the assumptions (Gallagher and Doherty, 2007; Yang et al., 2008). However, it is noted that all transformation methods will intrinsically change the implied information content of the observations (Beven et al., 2008). For example, in an urban drainage model, using a log or Box-Cox transformation (Box and Cox, 1964) to meet normality of residuals will place more emphasis on different parts of the hydrograph (i.e. lower flow rates), which in turn significantly influences the sensitivity to the model parameters (Yang et al., 2008). **The impacts of verifying, or not verifying these assumptions on the model sensitivity and associated parameter uncertainty have not been studied**.

According to Freni et al. (2009), the classical Bayesian method is more effective at discriminating models according to their uncertainty, but the GLUE approach performs similarly when it is based on the same founding assumptions as the Bayesian method. However, this conclusion is still debated (Beven, 2009; Vrugt et al., 2009). The different approaches have also been compared in related fields other than urban drainage; for example, Makowski et al. (2002) compared GLUE and Markov Chain Monte Carlo (MCMC) methods (in particular the Metropolis-Hasting sampling approach) using a simplified crop model with 22 parameters. Both methods presented similar results, but the authors recommend the use of the Metropolis-Hasting algorithm. This is because the Metropolis-Hasting method converges to the true posterior distribution even if the model includes a large number of parameters, while GLUE found this challenging because it required a large number of simulation runs. Nevertheless, MCMC procedures also have their own limitations and a misspecification of the error structure (or likelihood function) in the Bayesian approach can lead to an erroneous quantification of the model predictive uncertainty (Beven et al., 2008). Although these insights are valuable, it is difficult to relate the results between different comparison studies, as they employ different models and different datasets.

In summary, the reviewed global uncertainty analysis methods have the advantage of performing uncertainty analysis while providing information about the most likely parameter sets that calibrate the model. They do, however, have disadvantages (Bayesian - assumption or knowledge about the likelihood function and GLUE - subjective parameter acceptance criteria) that might limit their real potential of performing sensitive analysis. **Nevertheless, there is no information in the literature which suggests which is the most suitable method to assess parameter uncertainties in urban drainage models.**

2.4.3 Propagation of measured data uncertainty in stormwater models

The sensitivity analysis methods reviewed in the above section (Section 2.4.2) are used to identify the most influential parameters and also to assess the model predictive uncertainty due to parameter uncertainty. Nevertheless, the measured data and model structure are other sources of uncertainty that should be considered (as summarised in Sub-section 2.3.3).

Impacts of input data uncertainties on urban drainage modelling are largely unknown, although their importance in other related fields was already noted (e.g. hydrologic models: Krzysztofowicz and Kelly, 2000; Haydon and Deletic, 2009). Korving and Clemens (2005) evaluated the sensitivity of a distributed hydrologic model to parameter and radar rainfall uncertainty. Among other results, they found that as the drainage area increased, the uncertainty in flows modelled with the distributed model also increased. Some work has been done on the propagation of input data uncertainties through urban drainage models, mainly by methods based on Monte Carlo simulations (Rauch et al., 1998; Bertrand-Krajewski et al., 2003; Korving and Clemens, 2005). However, in these studies, the models were first calibrated assuming that measured inputs and outputs are without error, and the impact of input data uncertainties were then propagated through the models, while keeping the model parameters fixed. Kleidorfer et al. (2009) developed this further by assessing the impact of input data uncertainties on model parameters. The techniques used to measure urban discharges and associated water quality are of limited accuracy (Harmel et al., 2006b; Bertrand-Krajewski, 2007; McCarthy et al., 2008). However, the impact of this contribution to the model's overall uncertainty is not well understood. **The effect of input and calibration data uncertainty on the parameters and outputs of urban drainage models has not been explored**.

While most of the existing studies about input and calibration data uncertainty on the outputs have been restricted to hydrologic models of large natural catchments, the methods used are complex and as a result have limited practical application for urban stormwater models. For example, Renard et al. (2008) and Thyer et al (2009a; 2009b). applied the Bayesian Total Error Analysis methodology (BATEA proposed by Kuczera et al., 2006) to evaluate the uncertainties in hydrological models arising from model input, outputs and structural errors. The BATEA framework is based on hierarchical Bayesian models, in which each source of uncertainty is explicitly considered. The model input errors, structure errors and calibration errors are considered to be independent, and the error from one source should not be compensated by another one (Kuczera et al., 2006; Thyer et al., 2009b). For example, the uncertainty in the model inputs should not be compensated by re-calibrating the model and adjusting the parameters. Uncertainties are considered at their source and therefore different parameters should not compensate for each other. This methodology has significant advantages for estimation of predictive uncertainties when circumstances change. For example, changes in the availability of data (e.g. use of new measurement devices, installation of more rain gauges) can easily be implemented by adapting the specific error model without the requirement for a complete new calibration of the model (Renard et al., 2008). On the other side, the model can be changed, improved or extended without changing error models (Renard et al., 2008). This is attractive for urban drainage modelling because the systems are continuously changing with the changes in the urban infrastructure. In BATEA, error models are used to improve the model predictions. One of the disadvantages of this approach is the inclusion of a large number of extra parameters from the error models and there is a chance that the relationship between the number of parameters to be calibrated and the amount of measured data is such that it generates spurious results (Kleidorfer, 2010). In addition, validation of the estimated uncertainty is limited as BATEA is a probabilistic approach and the estimates are made in terms of predictive distributions. The approach seems very promising; however, its application is not straightforward, is computationally demanding and requires a significant level of expertise about model structures and probabilistic approaches, which might limit the range of users (Renard et al., 2008). **Perhaps a less complex approach is required to assess urban drainage models.**

2.5 Conclusions from the literature review

There are a vast variety of modelling approaches to simulate stormwater processes in the urban drainage system. While rainfall runoff models are well established, water quality models are still being researched, as the existing ones are unable to adequately represent the pollutant processes in stormwater systems. Furthermore, there are a wide range of uncertainties which can impact the modelling results. For further improvement of urban drainage models, it is imperative that these uncertainties are acknowledged and evaluated.

The key findings of this literature review are summarised as follows:

1. a number of rainfall runoff and pollution generation models have been used to predict flows and pollutants from stormwater. Independently of the modelling approach, conceptualisation errors will always be present. This is even more evident for pollution generation models because the related processes are not yet fully understood. It is assumed that such errors lead to a large amount of uncertainty in these models;
2. the uncertainties associated with urban drainage models have not been fully investigated. Therefore, there is a need to further explore this field of research. First, the sources of uncertainty in urban drainage models have to be understood, then their impact on the model's predictions can be evaluated;
3. model calibration parameters are likely to respond to all sources of uncertainty in the model; consequently, rigorous assessment of the uncertainty associated with model parameters is required, and robust methodologies should be used for this task;
4. global sensitivity analysis methods have the advantage of performing uncertainty analysis while providing information about the most likely parameter sets to calibrate the model. However, there is no indication of the most suitable method to assess stormwater models. Therefore, the comparison of different methods to perform parameter calibration, model sensitivity and uncertainty analysis will identify the most suitable one for stormwater models;
5. while a range of models have been applied worldwide to predict flows and pollution generation from stormwater, the assessment of the uncertainty associated with model structure has not been explored in any great detail. Furthermore, there is no standard method to evaluate this source of uncertainty. Different approaches are required to evaluate structural uncertainties, even if from a heuristic perspective. Testing models with different formulations (levels of complexity) through the application of a sound global sensitivity analysis method will provide information about the model structure (e.g. under or over parameterised, ability of representing (or not) different processes);

6. there are various assumptions made about global sensitivity analysis methods, and while these assumptions need to be verified, in practice it is common that such assumptions are not even checked. Research is required to determine the impact of verifying (or not) assumptions on the model sensitivity and predictive uncertainty. The comparison of scenarios in which the method assumption is verified and unverified will guide the approach to future applications; and,
7. input and calibration data are plagued with uncertainties. As such, there is a need to understand the impacts of these uncertainties on the performance, sensitivity and predictive uncertainty of stormwater models. This will contribute to further understanding their consequences in the modelling exercise.

2.6 Research aims and objectives

As stated in Chapter 1, the overall aim of this thesis is to further understand the impact of different sources of uncertainties on urban drainage models. The underlying hypothesis is that the sources of uncertainty are linked and that the model parameters respond to all the different uncertainty sources.

2.6.1 Specific aims and hypotheses

The literature review found that significant knowledge and data gaps exist in order to better understand uncertainties in urban drainage models. The overall aim presented above will be accomplished by completing a number of more specific objectives and hypotheses as follows:

(I) identify suitable method(s) to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models.

- it is hypothesised that different uncertainty analysis methods lead to different results with respect to model parameter sensitivity and predictive uncertainty because they rely on different formulations (e.g. formal probabilistic or not) and assumptions (e.g. assumption about the model errors structure, such as the assumption that the residuals are independent and normally distributed).
- it is hypothesised that there is a complex interaction between the complexity of the method used, computational time required and the knowledge/skill level of the modeller;
- it is hypothesised that verifying the underlying assumption of the sensitivity and uncertainty analysis method will result in the most comprehensive understanding of the model's uncertainty;

(II) explore parameter calibration, model sensitivity and the resulting predictive uncertainties in models with different levels of complexity;

- it is hypothesised that a well-posed and well-calibrated model (which has influential and identifiable parameters) will have a higher model efficiency. Providing inadequate calibration for a well-posed model may neglect important processes represented by the model;

- it is hypothesised that the results from a sound model sensitivity analysis will indicate if the model is well or 'ill-posed', as the identifiability of parameters, the confidence in the model results and the existence of model structure and conceptual errors will be determined.

- the assessment of the uncertainty originating from model parameters allows a comprehensive analysis of model structure and parameter interaction. Nevertheless, other sources of uncertainties (e.g. input measured data, model formulation and assumptions and selected objective function) should be investigated because they impact on the total uncertainties in the modelled results.

(III) explore the impact of measured input and calibration data uncertainty on the performance, sensitivity and predictive uncertainty of stormwater quantity and quality models;

- it is hypothesised that the model parameters can entirely compensate for the uncertainty in input and calibration data. As such, if the model parameters are considered initially as reflecting reality, then these uncertainties will reduce this representation; and,

- it is hypothesised that systematic errors in measured data will have more impact on the model sensitivity and uncertainty than random errors because they are time-dependent, and therefore they will be continuously propagated through the model.

2.6.2 Methodology used to complete the aims

In total, there are eight main chapters in this thesis, each contributing to one, or more, of the above listed aims.

Figure 2.2 provides an overview of how the chapters of the thesis are organised to achieve the major aim described at the beginning of Section 2.6. Data collected from different urban catchments in Melbourne, Australia, were used to complete many of the above aims and hypotheses. Assessing model sensitivity and uncertainty analysis forms the major part of the overall thesis.

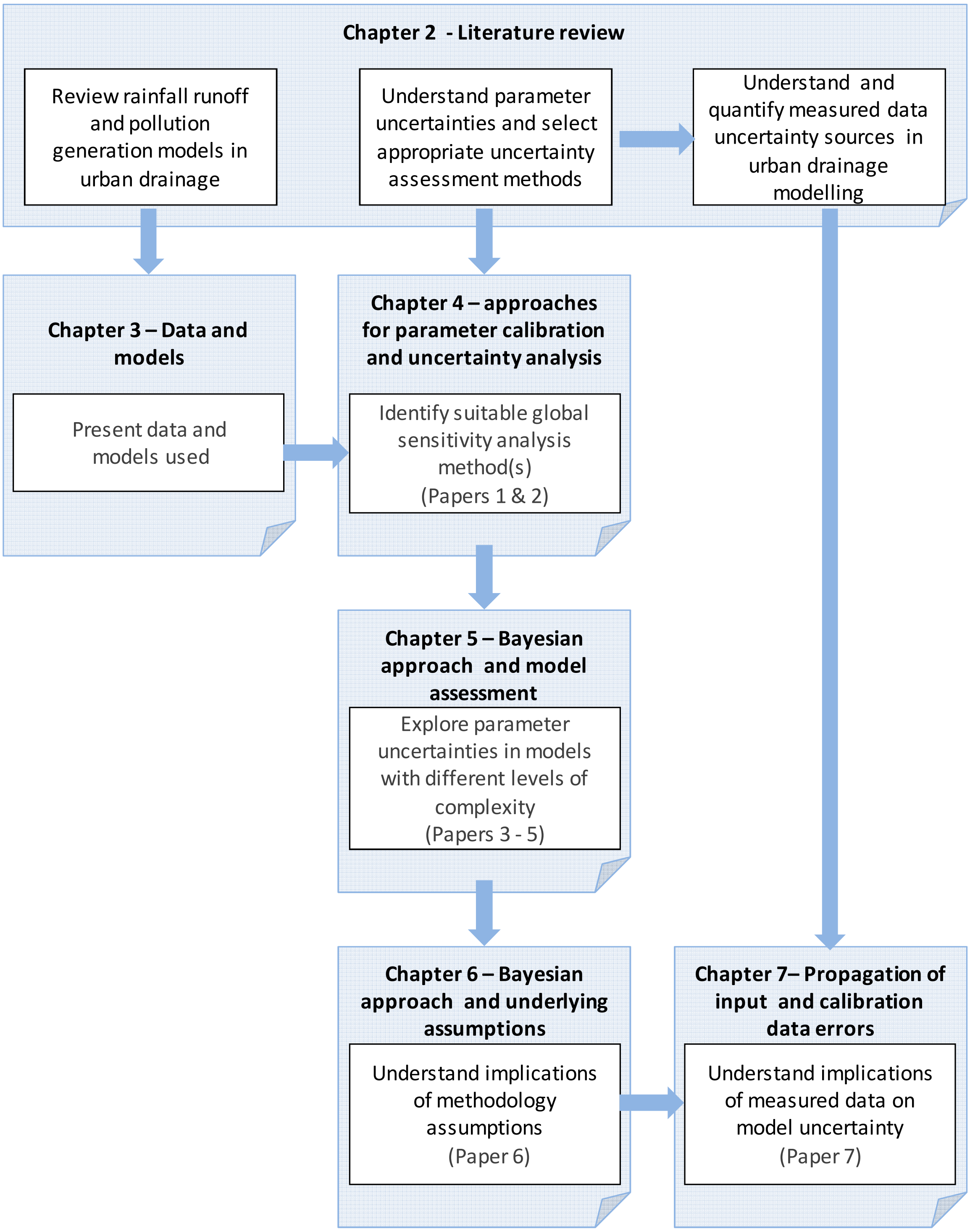


Figure 2.2 Flow chart describing the chapters of the thesis, the main roles they play in achieving the aim and how the papers fit in each of the chapters and objectives.

The information collated in the literature review will assist in the selection of different urban stormwater rainfall runoff and pollution generation models to be tested. Model calibration, sensitivity and uncertainty analyses will be conducted using an array of different methods. The most suitable one will be used to assess different models and sources of uncertainties.

2.6.3 Thesis by publication

This is a thesis by publication, and the results of this research are presented in the form of journal papers that are integrated within the chapters. Figure 2.2 shows how these journal papers are incorporated into the thesis and how they meet the key objectives.

The search for suitable global sensitivity analysis method(s) to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models is presented in Chapter 4 in the form of two journal papers: (1) *Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling*; and, (2) *Analysis of parameter uncertainty of a flow and quality stormwater model*.

Next, parameter calibration, model sensitivity and predictive uncertainties (originating from parameter uncertainties), in models with different formulations, are extensively investigated in Chapter 5 through three journal papers: (3) *Calibration and sensitivity analysis of urban drainage models: MUSIC rainfall/runoff module and a simple stormwater quality*; (4) *Stormwater quality models: performance and sensitivity analysis*; and, (5) *Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data*.

This is followed by an evaluation of the main assumption of the sensitivity analysis method in Chapter 6. This is presented in one journal paper: (6) *Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?* The last part of this research focuses on propagating measured data uncertainty through stormwater models and this is presented in Chapter 7 through the following journal paper: (7) *Impacts of measured data uncertainty on urban stormwater models*. Finally, the last chapter presents the main conclusions and summarises the main topics for further investigation (Chapter 8).

2.7 References

Abbott, M. B. and Refsgaard, J. C. (1996). *Distributed hydrological modelling (Eds.)*, Kluwer Academic Pub.

Achleitner, S., Moderl, M. and Rauch, W. (2007). CITY DRAIN © – An open source approach for simulation of integrated urban drainage systems. *Environmental Modelling and Software* **22**(8): 1184-1195.

Ahmad, M., Ghumman, A., Ahmad, S. and Hashmi, H. (2010). Estimation of a Unique Pair of Nash Model Parameters: An Optimization Approach. *Water Resources Management* **24**(12): 2971-2989.

Ahyerre, M., Chebbo, G., Tassin, B. and Gaume, E. (1998). Storm water quality modelling, an ambitious objective? *Water Science and Technology* **37**(1): 205-213.

Andreassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. and Lavabre, J. (2001). Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology* **250**(1-4): 206-223.

Bach, P. M., Daly, E., McCarthy, D. T. and Deletic, A. (2010). Investigating pollutant variability for the development of a semi-stochastic water quality model of urban stormwater runoff. *6th International Conference on Sewer Processes and Networks*. Surfers Paradise, Gold Coast, Australia.

Bates, B. C. and Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research* **37**(4): 937–947.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society London* **53**(A): 370–418.

Beck, M. B. (1987). Water Quality Modeling: A Review of the Analysis of Uncertainty. *Water Resources Research* **23**(8): 1393-1442.

Bedient, P. B. and Huber, W. C. (1992). *Hydrology and Flood Plain Analysis*. New York, Addison-Wesley.

Belia, E., Amerlinck, Y., Benedetti, L., Johnson, B., Sin, G., Vanrolleghem, P. A., Gernaey, K. V., Gillot, S., Neumann, M. B., Rieger, L., Shaw, A. and Villez, K. (2009). Wastewater treatment modelling: dealing with uncertainties. *Water Science and Technology* **60**(8): 1929-1941.

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

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

Bertrand-Krajewski, J. L., Bardin, J. P., Mourad, M. and Beranger, Y. (2003). Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. *Water Science and Technology* **47**(2): 95-102.

Bertrand-Krajewski, J. L., Barraud, S. and Bardin, J. P. (2002). Uncertainties, performance indicators and decision aid applied to stormwater facilities. *Urban Water* **4**(2): 163-179.

Bertrand-Krajewski, J. L., Briat, P. and Scrivener, O. (1993). Sewer sediment production and transport modelling: A literature review. *Journal of Hydraulic Research* **31**(4): 435-460.

Bertrand-Krajewski, J. L. and Muste, M. (2007). Chapter 6: Understanding and managing uncertainty. *in: Flectcher, T.D. and Deletić, A. (Eds.) Data Requirements for Integrated Urban Water Management*. 18 London, UK, Taylor and Francis, Urban Water series - UNESCO IHP**:** 65-90.

Beven, K. (2009). Comment on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" by Jasper A. Vrugt, Cajo J. F. terBraak, Hoshin V. Gupta and Bruce A. Robinson. *Stochastic Environmental Research and Risk Assessment* **23**(7): 1059-1060.

Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* **6**(3): 279-298.

Beven, K. and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **246**: 11-29.

Beven, K. J. (2001). *Rainfall-runoff modelling – the primer*. New York, John Wiley & Sons Ltd.

Beven, K. J., Smith, P. J. and Freer, J. E. (2008). So just why would a modeller choose to be incoherent? *Journal of Hydrology* **354**(1-4): 15-32.

Blasone, R.-S., Madsen, H. and Rosbjerg, D. (2008). Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* **353**(1-2): 18-32.

BOM (2012). Bureau of Meteorology (www.bom.gov.au). Accessed in the year 2012.

Boughton, W. C. (2007). Effect of data length on rainfall-runoff modelling. *Environmental Modelling and Software* **22**(3): 406-413.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*. **Series B 26**(2): 211-252.

Boyd, M. J., Bufill, M. C. and Knee, R. M. (1994). Predicting pervious and impervious storm runoff from urban drainage basins. *Hydrological Sciences* **39**(4): 321-332.

Brisbane City Council (2006). Estimation of MUSIC parameters for use in Brisbane city.

Butler, D. and Davies, J. W. (2000). *Urban Drainage*, London: E&FN Spon.

Butts, M. B., Payne, J. T., Kristensen, M. and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* **298**(1-4): 242-266.

Canale, R. P. and Seo, D.-I. (1996). Performance, reliability and uncertainty of total phosphorus models for lakes—II. Stochastic analyses. *Water Research* **30**(1): 95-102.

Carrera, J. and Neuman, S. P. (1986). Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 2. Uniqueness, Stability, and Solution Algorithms. *Water Resources Research* **22**(2): 211-227.

Charbeneau, R. J. and Barrett, M. E. (1998). Evaluation of Methods for Estimating Stormwater Pollutant Loads. *Water Environment Research* **70**(7): 295-1302.

Chiew, F. and McMahon, T. (1994). Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. *Journal of Hydrology* **153**(1-4): 383-416.

Chiew, F. H. S. and McMahon, T. A. (1999). Modelling runoff and diffuse pollution loads in urban areas. *Water Science and Technology* **39**(12): 214-248.

Chiew, F. H. S., Stewardson, M. J. and McMahon, T. A. (1993). Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology* **147**(1-4): 1-36.

City of Melbourne (2012). WSUD Guidelines. Applying the Model WSUD Guidelines. An Initiative of the Inner Melbourne Action Plan (http://www.melbourne.vic.gov.au/Sustainability/SavingWater/Pages/Watersensitivedesign.aspx). Accessed in the year 2012.

CRCCH (2005). MUSIC Version3 - User guide, Cooperative Research Centre for Catchment Hydrology.

Croke, B. F. W. (2009). *Representing uncertainty in objective functions: extension to include the influence of serial correlation*. 18th World IMACS / MODSIM Congress, Cairns, Australia.

Cunge, J. A. (1969). On the Subject of a Flood Propagation Computation Method (Muskingum Method). *Journal of Hydraulic Research* **7**(2): 205-230.

Deletic, A. (2001). Modelling of water and sediment transport over grassed areas. *Journal of Hydrology* **248**(1-4): 168-182.

Deletic, A., Ashley, R. and Rest, D. (2000). Modelling Input of Fine Granular Sediment into Drainage Systems via Gully-Plots. *Water Research* **34**(15): 3836-3844.

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. (2012).

Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth, Parts A/B/C* **42–44**(0): 3-10.

Deletic, A. and Maksimovic, C. T. (1998). Evaluation of water quality factors in storm runoff from paved areas. *Journal of Environmental engineering - ASCE* **124**: 869-879.

Dembélé, A. (2010). MES, DCO et polluants prioritaires des rejets urbains de temps de pluie: mesure et modélisation des flux événementiels (TSS, COD and priority pollutants in urban wet weather discharges: monitoring and modelling of event loads). *INSA*. Lyon, France. **PhD Thesis:** 286 (in French).

Dembélé, A., Bertrand-Krajewski, J.-L., Becouze, C. and Barillon, B. (2010). A new empirical model for stormwater TSS event mean concentrations (EMCs). *8th international conference on sustainable techniques and strategies in urban water management - NOVATECH 2010*. Lyon, France.

Department of Environment and Heritage NSW. (2012). Department of Environment and Heritage NSW (http://www.environment.nsw.gov.au/beach/waterpollution.htm). Accessed in the year 2012.

DHI (2002c). MOUSE pipe flow reference manual. Software, D. Horsolm, Denmark.

DHI (2004). MOUSE surface runoff models reference manual. *DHI Software*. Horsolm, Denmark.

Diskin, M. H. and Simon, E. (1977). A procedure for the selection of objective functions for hydrologic simulation models. *Journal of Hydrology* **34**(1-2): 129-149.

Doherty, J. (2003). Model Independent Markov Chain Monte Carlo Analysis - MICA, Watermark Numerical Computing and US EPA, US.

Doherty, J. (2004). PEST - Model-Independent Parameter Estimation Version 10, Watermark Numerical Computing and US EPA, US.

Doherty, J. and Johnston, J. M. (2003). Methodologies for calibration and predictive analysis of a watershed model *Journal of the American Water Resources Association*: 251-265.

Doherty, J. and Welter, D. (2010). A short exploration of structural noise. *Water Resources Research* **46**(W05525, doi:10.1029/2009WR008377).

Donohue, I. and Irvine, K. (2008). Quantifying variability within water samples: the need for adequate subsampling. *Water Research* **42**: 476-482.

Dorini, G., Kapelan, Z. and Azapagic, A. (2011). Managing uncertainty in multiple-criteria decision making related to sustainability assessment. *Clean Technologies and Environmental Policy* **13**(1): 133-139.

Duncan, H. P. (1995). A Review of Urban Storm Water Quality Processes. *Cooperative Research Centre for Catchment Hydrology*. Melbourne, Australia**:** 46 pp.

Duncan, H. P. (1999). Urban stormwater quality: A statistical overview. *Cooperative Research Centre for Catchment Hydrology*. Melbourne, Australia**:** 80 pp.

Efstratiadis, A. and Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal* **55**(1): 58-78.

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

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.-E., Quirmbach, M., Vaes, G. and Vieux, B. (2004). Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology* **299**(3-4): 186-202.

Einfalt, T., Arnbjerg-Nielsen, K. and Spies, S. (2002). An enquiry into rainfall data measurement and processing for model use in urban hydrology. *Water Science and Technology* **45** (2): 147-152.

Elliott, A. H. and Trowsdale, S. A. (2007). A review of models for low impact urban stormwater drainage. *Environmental Modelling and Software* **22**(3): 394-405.

eWater CRC (2009). Model for urban stormwater improvement conceptualisation (MUSIC) (Version 4.0), Canberra: eWater Cooperative Research Centre.

eWater CRC (2012). Model for urban stormwater improvement conceptualisation (MUSIC) (Version 5.0), Canberra: eWater Cooperative Research Centre.

Feyen, L., Vrugt, J. A., Nualláin, B. Ó., van der Knijff, J. and De Roo, A. (2007). Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology* **332**(3-4): 276-289.

Fletcher, T. D., Wong, T. H. F., Duncan, H. P., Deletić, A., Coleman, J. R. and Jenkins, G. A. (2004). Recent advances in decision-supprt for stormwater management: the MUSIC approach. *6th international conference on sustainable techniques and strategies in urban water management - NOVATECH 2004*. Lyon, France.

Francey, M. (2010). Characterising urban pollutant loads. *Civil Engineering*. Melbourne, Monash University. **PhD thesis:** 248.

Francey, M., Fletcher, T. D., Deletic, A. and Duncan, H. (2010). New insights into water quality of urban stormwater in South Eastern Australia. *Journal of Environmental Engineering* **136**(4): 381-390.

Freni, G., Mannina, G. and Viviani, G. (2008). Uncertainty in urban stormwater quality modelling: The effect of acceptability threshold in the GLUE methodology. *Water Research* **42**(8-9): 2061-2072.

Freni, G., Mannina, G. and Viviani, G. (2009). Urban runoff modelling uncertainty: Comparison among Bayesian and pseudo-Bayesian method. *Environmental Modelling and Software* **24**(9): 1100-1111.

Gallagher, M. and Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling and Software* **22**(7): 1000-1020.

Gamerith, V. (2006). Long term simulations of combined sewer facilities, INSA de Lyon, URGC - Hydrologie Urbaine & Graz University of Technology, Austria. **Diploma thesis:** 88.

Gaume, E., Villeneuve, J. P. and Desbordes, M. (1998). Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. *Journal of Hydrology* **210**: 38-50.

Gold Coast City Council (2006). MUSIC modelling guidelines.

Gourley, J. J. and Vieux, B. E. (2006). A method for identifying sources of model uncertainty in rainfall-runoff simulations. *Journal of Hydrology* **327**(1-2): 68-80.

Guinot, V., Cappelaere, B., Delenne, C. and Ruelland, D. (2011). Towards improved criteria for hydrological model calibration: theoretical analysis of distance- and weak form-based functions. *Journal of Hydrology* **401**: 1-13.

HACH, C. (2008). Sigma 950 Flow Meter - Instrument Manual. 2009, from http://www.hach.com/fmmimghach?/CODE%3A3314-2008-1117081%7C1.

Harmel, R., Cooper, R., Slade, R., Haney, R. and Arnold, J. (2006a). Cumulative uncertainty in measured stream flow and water quality data for small watersheds. *American Society of Agricultural and Biological Engineers* **49**(3): 689-701.

Harmel, R. D., King, K. W., Haggard, B. E., Wren, D. G. and Sheridan, J. M. (2006b). Practical guidance for discharge and water quality data collection on small watersheds. *American Society of Agricultural and Biological Engineers* **49**(4): 937-948.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97-109.

Hatt, B. E., Morison, P., Fletcher, T. D. and Deletic, A. (2009). Adoption guidelines for stormwater biofiltration systems, Facility for Advancing Water Biofiltration. University, M.

Haydon, S. and Deletic, A. (2007). Sensitivity testing of a coupled Escherichia coli - Hydrologic catchment model. *Journal of Hydrology* **338**(3-4): 161-173.

Haydon, S. and Deletic, A. (2009). Model output uncertainty of a coupled pathogen indicator-hydrologic catchment model due to input data uncertainty. *Environmental Modelling and Software* **24**(3): 322-328.

Hooke, A. R. and Jeeves, A. T. A. (1961). "Direct Search" Solution of Numerical and Statistical Problems. *Journal of the ACM* **8**(2): 212-229.

Hossain, I., Imteaz, M., Gato-Trinidad, S. and Shanableh, A. (2010). Development of a catchment water quality model for continuous simulations of pollutants build-up and wash-off. *International Journal of Civil and Environmental Engineering* **2**(4): 210-217.

Huard, D. and Mailhot, A. (2008). Calibration of hydrological model GR2M using Bayesian uncertainty analysis. *Water Resources Research* **44**(W02424, doi:10.1029/2007WR005949).

Huber, W. C. (1985). Deterministic modeling of urban runoff quality. In: Torno, H. C.,Marsalek, J.andDesbordes, M. *Urban Runoff Pollution*. 527, NATOASI Series, Series G: Ecological Sciences, 10:167-242, Springer-Verlag, New York, NY.

Huber, W. C. (1986). *Modeling urban runoff quality: state of the art*. Conference on Urban Runoff Quality, Impact and Quality Enhancement Technology, Engineering Foundation, ASCE, New York, NY.

INSA/SOGREAH (1999). CANOE: Manuel d'utilisation. Villeurbanne, France, ALISON, INSA LYON.

ISO (1993). Guide to the Expression of Uncertainty in Measurement, 1st edn, International Organization for Standardization. Switzerland.

ISO (1995). Guide to the Expression of Uncertainty in Measurement.

ISO (2007). ISO/TS 25377:2007 Hydrometric uncertainty guide (HUG). November 2007, Geneva, Switzerland**:** 62.

ISO (2008). ISO/IEC Guide 98-3/Suppl.1:2008(E) Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) Supplement 1: Propagation of distributions using a Monte Carlo method. December 2008, Geneva, Switzerland**:** 98.

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

ISO (2009b). ISO/IEC Guide 98-3/S1/AC1:2009(E) Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM:1995), Supplement 1: Propagation of distributions using a Monte Carlo method, Technical corrigendum 1. May 2009, Geneva, Switzerland**:** 2.

Kacker, R. and Jones, A. (2003). On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent. *Metrologia* **40**: 235-248.

Kanso, A., Chebbo, G. and Tassin, B. (2006). Application of MCMC-GSA model calibration method to urban runoff quality modeling. *Reliability Engineering and System Safety* **91**(10-11): 1398-1405.

Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B. and Chebbo, G. (2003). Bayesian approach for the calibration of models: application to an urban stormwater pollution model. *Water Science and Technology* **47**(4): 77-84.

Kanso, A., Tassin, B. and Chebbo, G. (2005). A benchmark methodology for managing uncertainties in urban runoff quality models. *Water Science and Technology* **51**(2): 163-170.

Kapelan, Z., Savic, D. and Walters, G. (2007). Calibration of water distribution hydraulic models using a Bayesian-type procedure. *Journal of Hydraulic Engineering* **133**(8): 927-936.

Khu, S. T. and Madsen, H. (2005). Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Research* **41**: W03004, doi:10.1029/2004WR003041.

Kleidorfer, M. (2010). Uncertain calibration of urban drainage models. A scientific approach to solve practical problems. *Civil Engineering*, University of Innsbruck. **PhD thesis:** 288.

Kleidorfer, M., Deletic, A., Fletcher, T. D. and Rauch, W. (2009). Impact of input data uncertainties on stormwater model parameters. *Water Science and Technology* **60**(6): 1545 - 1554.

Korving, H. and Clemens, F. (2005). Impact of dimension uncertainty and model calibration on sewer system assessment. *Water Science and Technology* **52**(5): 35-42.

Kotlash, A. R. and Chessman, B. C. (1998). Effects of water sample preservation and storage on nitrogen and phosphorus determinations: implications for the use of automated sampling equipment. *Water Research* **32**(12): 3731-3737.

Krause, P., Boyle, D. P. and Base, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5**: 89-97.

Krzysztofowicz, R. and Kelly, K. (2000). Hydrologic Uncertainty Processor For Probabilistic River Stage Forecasting. *Water Resources Research* **36**(11): 3265-3277.

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology* **331**(1-2): 161-177.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* **211**(1-4): 69-85.

Kunstmann, H., Krause, J. and Mayr, S. (2006). Inverse distributed hydrological modelling of Alpine catchments. *Hydrology and Earth System Sciences* **10**: 395-412.

Ladson, A. (2008). Chapter 10: Rainfall-runoff modelling. *Hydrology: an Australian introduction*. 352 New York, Oxford University Press.

Larssen, T., Høgåsen, T. and Cosby, B. J. (2007). Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Ecological Modelling* **207**(1): 22-33.

Lee, J. G. and Heaney, J. P. (2003). Estimation of Urban Imperviousness and its Impacts on Storm Water Systems. *Journal of Water Resources Planning and Management* **129**(5): 419-426.

Leecaster, M. K., Schiff, K. and Tiefenthaler, L. L. (2002). Assessment of efficient sampling designs for urban stormwater monitoring. *Water Research* **36**(6): 1556-1564.

Letcher, R. A., Jakeman, A. J., Calfas, M., Linforth, S., Baginska, B. and Lawrence, I. (2002). A comparison of catchment water quality models and direct estimation techniques. *Environmental Modelling and Software* **17**(1): 77-85.

Li, L., Xia, J., Xu, C.-Y. and Singh, V. P. (2010). Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *Journal of Hydrology* **390**(3-4): 210-221.

Lindblom, E., Ahlman, S. and Mikkelsen, P. S. (2011). Uncertainty-based calibration and prediction with a stormwater surface accumulation-washoff model based on coverage of sampled Zn, Cu, Pb and Cd field data. *Water Research* **45**(13): 3823-3835.

Lindenschmidt, K. E. (2006). The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling. *Ecological Modelling* **190**: 72–86.

Macleod, A. (2008). MUSIC calibration based on soils conditions. *SW and QLD Stormwater Industry Association*. Surfers Paradise, QLD, Australia.

Madsen, H., Wilson, G. and Ammentorp, H. C. (2002). Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology* **261**(1-4): 48-59.

Makowski, D., Wallach, D. and Tremblay, M. (2002). Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie* **22**(S1): 191-203.

Maksimovic, C., Buzek, L. and Petrovic, J. (1991). Corrections of rainfall data obtained by tipping bucket rain gauge. *Atmospheric Research* **27**(1-3): 45-53.

Mannina, G. and Viviani, G. (2010). A hydrodynamic water quality model for propagation of pollutants in rivers. *Water Science and Technology* **62**(2): 288-299.

Mantovan, P. and Todini, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology* **330**(1-2): 368-381.

Marshall, L., Nott, D. and Sharma, A. (2004). A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research* **40**.

Marshall, L., Nott, D. and Sharma, A. (2005). Hydrological model selection: A Bayesian alternative. *Water Resources Research* **41, W10422, doi:10.1029/2004WR003719**.

Matott, L. S., Babendreier, J. E. and Purucker, S. T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research* **45**(6): W06421.

McAlister, T., Mitchell, G., Fletcher, T. and Phillips, B. (2006). Modelling Urban Stormwater Management Systems. In: Wong, T. H. F. *Australian Runoff Quality*. 153, Institution of Engineers, Australia.

McCarthy, D. T. (2008). Modelling microorganisms in urban stormwater. *Civil Engineering*. Melbourne, Monash University, Australia. **PhD thesis:** 488.

McCarthy, D. T., Deletic, A., Mitchell, V. G. and Diaper, C. (2011). Development and testing of a model for Micro-Organism Prediction in Urban Stormwater (MOPUS). *Journal of Hydrology* **409**(1-2): 236-247.

McCarthy, D. T., Deletic, A., Mitchell, V. G., Fletcher, T. D. and Diaper, C. (2008). Uncertainties in stormwater E. coli levels. *Water Research* **42**(6-7): 1812-1824.

McIntyre, N., Jacksona, B., Wadeb, A. J., Butterfieldb, D. and Wheater, H. S. (2005). Sensitivity analysis of a catchment-scale nitrogen model. *Journal of Hydrlogy* **316**: 71-92.

Melbourne Water (2012). Stormwater pollution (http://www.melbournewater.com.au/content/drainage_and_stormwater/stormwater/stormwater.asp?bhcp=1). Accessed in the year 2012.

Melching, C. S. (1995). Reliability Estimation. *Singh, V.P. (ed.) Computer Models of Watershed Hydrology*. 469 USA, Water Resources Publications**:** 69-18.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H. (1953). Equation of state calculations by fast Computing machines. *The Journal of Chemical Physics* **21**(6): 1087-1092.

Michelson, D., Einfalt, T., Holleman, I., Gjertsen, U., Friedrich, K., Haase, G., Lindskog, M. and Jurczyk, A. (2005). Weather Radar Data Quality in Europe: Quality Control and Characterization. Review, COST Action 717 - Use of radar observations in hydrological and NWP models. Luxembourg.

Mitchell, V. G., McCarthy, D. T., Deletic, A. and Fletcher, T. D. (2008). Urban stormwater harvesting - sensitivity of a storage behaviour model. *Environmental Modelling and Software* **23**(6): 782-793.

Molini, A., Lanza, L. G. and La Barbera, P. (2005). Improving the accuracy of tipping-bucket rain records using disaggregation techniques. *Atmospheric Research* **77**(1-4): 203-217.

Montanari, A. (2005). Large sample behaviours of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research* **41**(8): W08406.

Moore, C. and Doherty, J. (2005). Role of the calibration process in reducing model predictive error. *Water Resour. Res* **41**: W05020, doi:10.1029/2004WR003501.

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

Mourad, M., Bertrand-Krajewski, J. L. and Chebbo, G. (2005b). Sensitivity to experimental data of pollutant site mean concentration in stormwater runoff. *Water Science and Technology* **51**(2): 155-162.

Mourad, M., Bertrand-Krajewski, J. L. and Chebbo, G. (2005c). Stormwater quality models: sensitivity to calibration data. *Water Science and Technology* **52**(5): 61-68.

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

Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal* **7**(4): 308-313.

Neumann, M. B., Gujer, W. and von Gunten, U. (2009). Global sensitivity analysis for model-based prediction of oxidative micropollutant transformation during drinking water treatment. *Water Research* **43**(4): 997-1004.

O'Donnell, T. and Canedo, P. (1980). The reliability of conceptual basin model calibration. *Hydrlogical forecasting (Proceedings of the Oxford Symposium, April 1980).*

Palmstrom, N. and Walker, W. W. J. (1990). P8 urban catchment model: User's guide. Program documentation, and evaluation of existing models, design concepts, and Hunt-Potowomut data inventory. NBP-90-50, T. N. B. P. R. N.

Park, D. and Roesner, L. A. (2012). Evaluation of pollutant loads from stormwater BMPs to receiving water using load frequency curves with uncertainty analysis. *Water Research*(0).

Parker, G. T., Droste, R. L. and Rennie, C. D. (2010). Stochastic calibration of riverine water quality models. *Water Environment Research* **82**(2): 99-108.

Pitt, R. (1998). Unique features of the Source Loading and Management Model (SLAMM). *James, W. (Ed.), Modeling the Management of Stormwater Impacts* **6** (Computational Hydraulic International): 13-35.

Ratto, M., Tarantola, S. and Saltelli, A. (2001). Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communications* **136**(212-224).

Rauch, W., Bertrand-Krajewski, J.-L., Krebs, P., Mark, O., Schilling, W., Schütze, M. and Vanrolleghem, P. A. (2002a). Mathematical modelling of integrated urban drainage systems. *Water Science and Technology* **45**(3): 81-94.

Rauch, W. and Kinzel, H. (2007). KAREN - User Manual, hydro-IT GmbH, Innsbruck, (in German).

Rauch, W., Krejci, V. and Gujer, W. (2002b). REBEKA – software tool for planning urban drainage on the basis of predicted impacts on receiving waters. *Urban Water* **4**(4): 355-361.

Rauch, W., Thurner, N. and Harremoes, P. (1998). Required accuracy of rainfall data for integrated urban drainage modeling. *Water Science and Technology* **37**(11): 81-89.

Refsgaard, J. C. (2000). Towards a formal approach to calibration and validation of models using spatial data. In: Grayson, R.andBloshchl, G. *Spatial patterns in catchment hydrology: observations and modelling*. 398, Cambridge University Press**:** 329-354.

Refsgaard, J. C. and Abbott, M. B. (1996). The role of distributed hydrological modelling in water resources management. In: Abbott, M. B.andRefsgaard, J. C. *Distributed hydrological modelling*. 526, Kluwer Academic Pub.

Refsgaard, J. C. and Henriksen, H. J. (2004). Modelling guidelines-terminology and guiding principles. *Advances in Water Resources* **27**: 71–82.

Refsgaard, J. C. and Knudsen, J. (1996). Operational Validation and Intercomparison of Different Types of Hydrological Models. *Water Resour. Res.* **32**(7): 2189-2202.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. and van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* **29**(11): 1586-1597.

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L. and Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling and Software* **22**(11): 1543-1556.

Reichert, P. and Vanrolleghem, P. (2001). Identifiability and uncertainty analysis of the River Water Quality Model No. 1 (RWQM1). *Water Science and Technology* **43**(7): 329-338.

Reichl, J. P. C., Chiew, F. H. S. and Western, A. W. (2006). *Model Averaging, Equifinality and Uncertainty Estimation in the Modelling of Ungauged Catchments*. iEMSs Third Biennial Meeting, Summit on Environmental Modelling and Software, International Environmental Modelling and Software Society (Burling).

Renard, B., Kuczera, G., Kavetski, D., Thyer, M. and Franks, S. (2008). *Bayesian Total Error Analysis for Hydrologic Models: Quantifying Uncertainties Arising from Input, Output and Structural Errors*. Water Down Under 2008 Modbury, SA: Engineers Australia ; Causal Productions

Rodríguez, J. P., Achleitner, S., Möderl, M., Rauch, W., Maksimović, Č., McIntyre, N., Díaz-Granados, M. A. and Rodríguez, M. S. (2010). Sediment and pollutant load modelling using an integrated urban drainage modelling toolbox: An application of City Drain. *Water Science and Technology* **61**(9): 2273-2282.

Rosenbrock, H. H. (1960). An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal* **3**(3): 175-184.

Rossi, L., Krejci, V., Rauch, W., Kreikenbaum, S., Fankhauser, R. and Gujer, W. (2005). Stochastic modeling of total suspended solids (TSS) in urban areas during rain events. *Water Research* **39**(17): 4188-4196.

Saltelli, A. (2005). Global Sensitivity Analysis: An Introduction. *in: Manson, K.M. ang Hemez, F.M. (Eds) Sensitivity Analysis of Model Output*. 492, Los Alamos National Laboratory.

Sartor, J. D. and Boyd, G. B. (1972). Water pollution aspects of street surface contaminants. *Rep. EPA-R2-72-081*. Washington, D. C., U.S. Environ. Prot. Agency.

Sartor, J. D., Boyd, G. B. and Agardy, F. J. (1974). Water pollution aspects of street surface contaminants. *Journal of the Water Pollution Control Federation* **46**(3): 458-467.

Scholes, L., Revitt, D. M. and Ellis, J. B. (2008). A systematic approach for the comparative assessment of stormwater pollutant removal potentials. *Journal of Environmental Management* **88**(3): 467-478.

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

SCS (1956). *National Engineering Handbook, Hydrology, Section 4, Soil Conservation Service*. Washington, D.C., US Department of Agriculture.

Servat, E. and Dezetter, A. (1991). Selection of calibration objective functions in the context of rainfall-runoff modelling in a sudanese savannah area. *Hydrological Sciences Journal* **36**(4): 307-330.

Sevruk, B. (1982). Methods of correction for systematic error in point precipitation measurement for operational use,. *WMO Operational Hydrology Report 21* **21**(589): 91.

Sevruk, B. (2002). WMO questionnaire on recording precipitation gauges: state-of-the-art. *Water Science and Technology* **45**(2): 139-145

Shaw, S. B., Stedinger, J. R. and Walter, M. T. (2010). Evaluating Urban Pollutant Buildup/Wash-Off Models Using a Madison, Wisconsin Catchment. *Journal of Environmental Engineering* **136**(2): 194-203.

Shaw, S. B., Walter, M. T. and Steenhuis, T. S. (2006). A physical model of particulate wash-off from rough impervious surfaces. *Journal of Hydrology* **327**(3-4): 618-626.

Sherman, L. K. (1932). Stream-flow from Rainfall by the Unit-Graph Method. *Engineering News-Record 108: 501-505*. **108:** 501-505.

Shivers, D. E. and Moglen, G. E. (2008). Spurious Correlation in the USEPA Rating Curve Method for Estimating Pollutant Loads. *Journal of Environmental Engineering* **134**(8): 610-618.

Skahill, B. E. and Doherty, J. (2006). Efficient accommodation of local minima in watershed model calibration. *Journal of Hydrology* **329**(1-2): 122-139.

Sorooshian, S., Gupta, V. K. and Fulton, J. L. (1983). Evaluation of Maximum Likelihood Parameter Estimation Techniques for Conceptual Rainfall-Runoff Models : Influence of Calibration Data Variability and Length on Model Credibility. *Water Resources Research* **19**(1): 251–259.

Sriananthakumar, K. and Codner, G. P. (1992). *Factors affecting pollutant washoff decay coefficients in urban catchments*. International Symposium on Urban Stormwater Management, Sydney, Australia.

Sun, S. and Bertrand-Krajewski, J.-L. (2012). On calibration data selection: The case of stormwater quality regression models. *Environmental Modelling and Software* **35**(0): 61-73.

Taylor, B. N. and Kuyatt, C. E. (1994). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. USA, NIST Technical Note 1297, National Institute of Standards and Technology.

Taylor, G. D. (2006). Improved effectiveness of nitrogen removal in constructed stormwater wetlands. Civil Engineering Department. Monash University, Australia.

Thorndahl, S., Schaarup-Jensen, K. and Jensen, J. B. (2008). Probabilistic modelling of combined sewer overflow using the First Order Reliability Method. *Water Science and Technology* **57**(9): 1337-1344.

Thyer, M., Renard, B., Kavetski, D. and Kuczera, G. (2009a). *Impact of runoff measurement error models on the quantification of predictive uncertainty in rainfall-runoff models*. 18th World IMACS / MODSIM Congress, Cairns, Australia.

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

USEPA (1998). The Quality of Our Nation's Water: A Summary of the National Water quality Inventory 1998. Report to Congress. US Environmental Protection Agency (USEPA). US Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (2007). Storm Water Management Model (SWMM) - User Manual version 5.0. US Environmental Protection Agency (USEPA). *US Environmental Protection Agency - EPA/600/R-05/039*. Washington DC.

USEPA (2012). US Environmental Protection Agency (http://water.epa.gov/polwaste). Accessed in the year 2012.

Vanrolleghem, P. A., Daele, M. V. and Dochain, D. (1995). Practical identifiability of a biokinetic model of activated sludge respiration. *Water Research* **29**(11): 2561-2570.

Vanrolleghem, P. A. and Keesman, K. J. (1996). Identification of biodegradation models under model and data uncertainty. *Water Science and Technology* **36**(5): 69-79.

Varella, H., Guérif, M. and Buis, S. (2010). Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environmental Modelling and Software* **25**(3): 310-319.

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

Vaze, J. and Chiew, F. H. S. (2003). Comparative evaluation of urban storm water quality models. *Water Resources Research* **39**(10): 1280.

Vezzaro, L., Eriksson, E., Ledin, A. and Mikkelsen, P. S. (in press). Quantification of uncertainty in modelled partitioning and removal of heavy metals (Cu, Zn) in a stormwater retention pond and a biofilter. *Water Research*(0).

Vezzaro, L. and Mikkelsen, P. S. (2011). Investigating the appropriate level of complexity for stormwater micropollutants modelling. *12th International Conference on Urban Drainage*. Porto Alegre, Brazil.

Vezzaro, L. and Mikkelsen, P. S. (2012). Application of global sensitivity analysis and uncertainty quantification in dynamic modelling of micropollutants in stormwater runoff. *Environmental Modelling and Software* **27–28**(0): 40-51.

Vogel, R. M., Rudolph, B. E. and Hooper, R. P. (2005). Probabilistic Behavior of Water-Quality Loads. *Journal of Environmental Engineering* **131**(7): 1081-1089.

Vrugt, J. A., Braak, C. J. F. t., Diks, C. G. H., Robinson, B. A., Hyman, J. M. and Higdon, D. (2009). 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): 271-288.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M. and Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.* **44**: W00B09.

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

Wagener, T., Wheater, H. S. and Gupta, H. V. (2004). *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*. London, UK, Imperial College Press.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluis, J. P., van Asselt, M. B. A., Janssen, P. and Krayer von Krauss, M. P. (2003). Defining uncertainty – A conceptual basis for uncertaintry management in model-based decision support. *Integ. Assessment* **4**(1): 5-17.

Wallingford Software (2009). InfoWorks CS Help, Wallingford Software, United Kingdom.

Walsh, C. J., Fletcher, T. D. and Ladson, A. R. (2005). Stream restoration in urban catchments through redesigning stormwater systems: looking to the catchment to save the stream. *Journal of the North American Benthological Society* **24**(3): 690-705.

Weijers, S. R. and Vanrolleghem, P. A. (1997). A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Science and Technology* **36**(5): 69-79.

Willems, P. (2008). Quantification and relative comparison of different types of uncertainties in sewer water quality modeling. *Water Research* **42**(13): 3539-3551.

Wong, T. H. F., Allen, R., Beringer, J., Brown, R. R., Chaudhri, V., Deletic, A., Fletcher, T. D., Gernjak, W., Hodyl, L., Jakob, C., Reeder, M., Tapper, N. and Walsh, C. (2011). *Blueprint2011 – Stormwater Management in a Water Sensitive City*. Melbourne, Australia, Centre for Water Sensitive Cities.

WP Software (1995). *XP-SWMM stormwater management model with XP graphical interface, User's manual, Version 2*. Tampa, Forida and Belconnen, USA.

XP-SOFTWARE (1977). XP-Storm - Technical Description. *XP Software*. Australia.

XP-SOFTWARE (1999). XP-AQUALM - Technical Description, XP Software, Australia.

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J. and Yang, H. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology* **358**(1-2): 1-23.

Yapo, P. O., Gupta, H. V. and Sorooshian, S. (1996). Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* **181**(1-4): 23-48.

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

Chapter 3

Data and models

3.1 Introduction

This chapter provides an overview of the dataset and stormwater models employed in this study. The data collection was conducted by two former Monash University PhD students, and the full details of the monitoring campaigns can be found in their thesis: McCarthy (2008) and Francey (2010). The key aspects of the monitoring are summarised in this section. The models selection was done so that both water quantity and quality were included. The models were of different complexities and reliabilities to investigate impacts of model structure on the results.

3.2 Overview of the monitoring sites

The study sites used in this thesis were part of a large monitoring program which focused on the measurement of rainfall and flow data and on the collection of typical stormwater pollutants, such as Total Suspended Solids (TSS) and Total Nitrogen (TN) (Francey, 2010; Francey et al., 2010). Four out of the five sites also included a comprehensive data set on *E. coli* (McCarthy, 2008). The sites were selected to represent different characteristics of urban catchments, which included the coverage of:

- different catchment sizes;
- different levels of development (i.e. from high density developments to low density developments), which reflects the various levels of imperviousness;
- different types of land-uses (i.e. industrial, commercial and residential land-uses); and,
- relatively established catchments to ensure that construction or other development works were kept to a minimum during the sampling period.

In addition, the security of the sampling point and the proximity of the sites to the research team were also taken into account when selecting the sites. According to these criteria, five study catchments in the inner suburbs of Melbourne, Australia were selected (McCarthy, 2008; Francey, 2010).

Table 3.1 shows a summary of the characteristics of the catchments. The total impervious fraction of the sites (*TIF*) ranges from 0.2 to to 0.8 and catchment areas ranges from just 10 to over 100 ha. The level of development of the catchments is also diverse. Figure 3.1 presents the aerial photographs of each catchment showing the relative positions of the centroid, rainfall gauge and flow gauges. All catchments are serviced by separate stormwater and wastewater systems, but some cross-connections between systems are expected. Narre Warren and Doncaster are the only sites that contain septic systems (tanks).

Table 3.1 Summary of the characteristics of the catchments (McCarthy, 2008; Francey, 2010).

Site	**Gilby Rd (GR)**	**Richmond (RICH)**	**Ruffeys Lake, Doncaster (RD)**	**Shepherds Bush (SB)**	**Narre Warren (NW)**
Primary Land Use	Commercial	High Density Residential	Medium Density Residential	Medium Density Residential	Rural Residential
Area (ha)	28.2	89.1	105.6	38	10.5
TIF*	0.8	0.74	0.51	0.45	0.2
Catchment Average slope (%)	1	3.5	5	4	4
Time of concentration (min)	23	31	14	14	16

* Total Impervious Fraction (*TIF*)

3.3 Overview of the monitoring programs

0.2 mm tipping bucket rainfall gauges were used in all the sites and the bucket size volume was regularly calibrated. The gauges were installed in areas where they were not obstructed by trees or buildings, as close as possible to the catchment's centroid, accessible by monitoring staff, and safe to leave unattended (Francey, 2010). Figure 3.1 shows the location of each of the rainfall gauges in relation to the catchments' centroids.

Rainfall loggers were installed next to each rainfall gauge and were programmed to count the number of tips which occur during each minute. Visual inspection of each of the loggers and rainfall gauges were made on each visit and if this inspection revealed any inconsistency then the loggers and rainfall gauges were re-calibrated and tested. The rainfall intensities were then calculated taking into account the time passed between the recorded tips and the number of tips registered (see Chapter 3, Section 3.1.1 of McCarthy, 2008 for full details).

Flows were measured with the American Sigma/HACH area-velocity 950 sensor (HACH, 2008) installed in the outlet pipes of each of the sites. Figure 3.1 shows the location of each of the flow meters on the aerial photographs of each site. The Sigma 950 uses a submerged area velocity sensor probe containing a pressure transducer to measure depth of flow in conjunction with ultrasonic transducers for velocity measurement. Subsequently the flow rates were calculated by the product between the wetted cross-sectional area and the velocity. Flow loggers were installed close to the flow meter and were programmed to measure the instantaneous average cross sectional velocity (m/s) and depth (m) of the stormwater and these readings were recorded at the end of every minute. Visual inspection of each of the loggers and flow gauges were made on most the fortnightly visits and if this inspection revealed any inconsistency (e.g. debris, etc.) then the loggers and flow gauges were re-calibrated and tested.

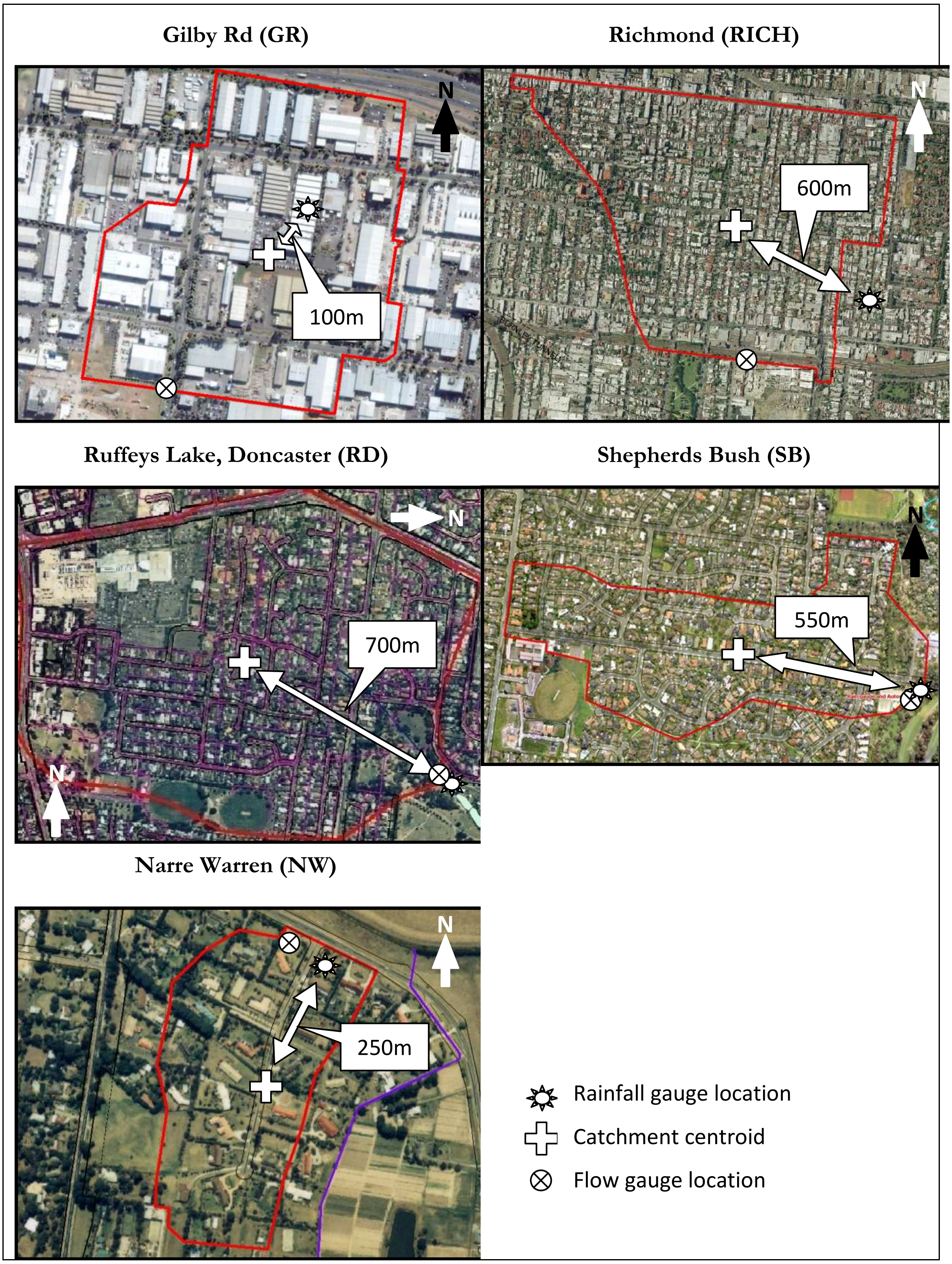


Figure 3.1 Aerial photographs of each catchment showing the relative positions of the centroid, rainfall gauge and flow gauges (after McCarthy, 2008).

The water quality samples were collected at the outlet of the catchments using a discrete sampling methodology. Non refrigerated autosamplers were used to withdraw samples from the stormwater using peristaltic pumps through reinforced suction tubes. The suction tubes were placed at a continuous gradient from the autosampler to the stormwater pipe, to avoid any ponding of water within the tube. A flow-weighed sampling approach was used, which means that subsequent samples were taken after predetermined volumes passed through the pipe. In general, pollutant concentrations vary considerably during the ascending peak of the hydrograph and regularise during the descending peak (Leecaster et al., 2002). As such the trigger volumes were set to allow a representative coverage of the event with the sample intervals becoming larger as the event progresses (e.g. see Figure 3.2).

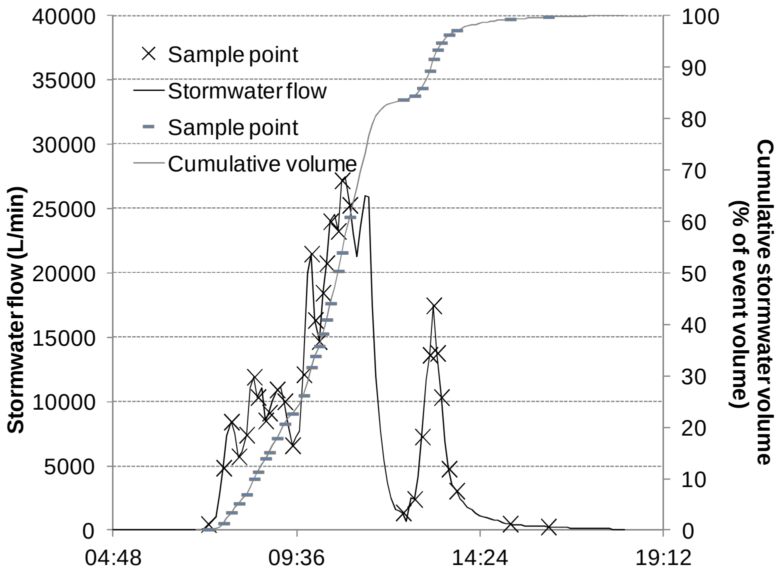


Figure 3.2. Example hydrograph and sampling points for Richmond catchment (RICH) taken on 14th of April 2005.

An event was considered to be representative if four or more samples were collected. The starting point of an event was considered to be at 3 hours before the first sample and the end point after 3 hours past from the last sample. In addition, the event was ended if rainfall or runoff did not occur for two hours or more.

3.4 Overview of the data used in this study

Most of the sites were monitored from January 2004 to December 2007. Rain and flow data, collected between 2004 and 2005, were used for model calibration (used in Chapters 4, 5, 6 and 7), while data collected from 2006 to 2007 was used for model validation (see Chapter 5). Table 3.2 reports on the characteristics of events used for model calibration, and also on the events used for validation (these are presented in square parentheses).The validation of the water quality models was not performed as preliminary tests and previous studies (e.g. Kanso et al., 2006) confirmed the models' very low performance even during calibration. As such, validation of these models would not make sense and therefore data is not presented.

Water quantity data: The mean annual rainfall in these catchments ranges from 370 to over 720 millimetres per year. From the rainfall figures in Table 3.2, it is possible to note that the data used for validation reflects the severe draught that Melbourne went through during 2006 and 2007. The mean rainfall among all the catchments that was 661 mm for the data period used for calibration (and reflects the normal average rainfall in Melbourne) went down to 452 mm during the period of data used for validation.

Figure 3.3 presents the flow duration curves for each catchment in mm/min (i.e. flow per catchment area) during the period used for calibration, 2004 and 2005 (on the left) and validation, 2006 and 2007 (on the right). Again the effects of the severe draught are identified in the flow duration curves.

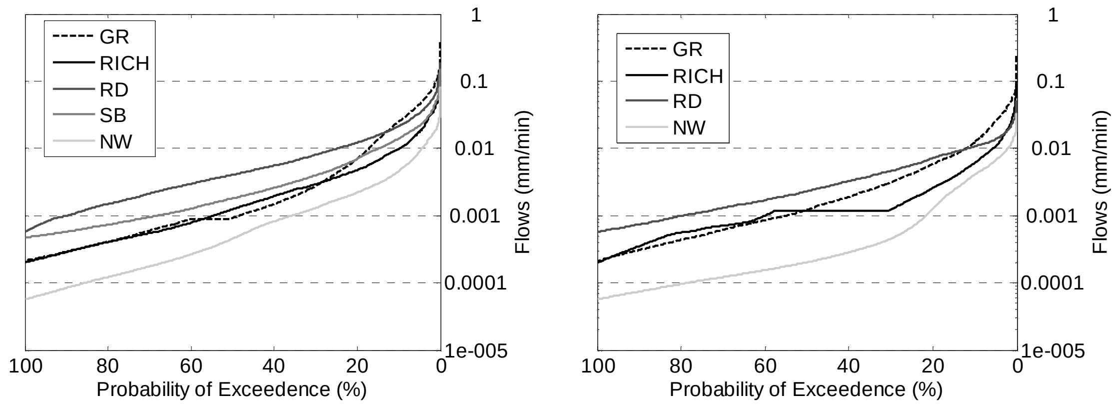


Figure 3.3 Flow duration curves for each catchment (mm/min) - calibration (left) and validation (right) period.

Although measures were taken to avoid measurement errors, uncertainty can only be minimised but not eliminated. Therefore, rainfall data used in this study was processed to cope with gaps and time drifts, which are inherent in any battery controlled logging device. Flow data was checked for any discrepancy (e.g. backflow effects indicated by negative velocities). As such, these 'corrected' or validated datasets are used for the subsequent chapters (see Chapter 3, Section 3.1.1 of McCarthy, 2008 for full details).

Table 3.2 Summary of measured data details (McCarthy, 2008; Francey, 2010). The characteristics of events used for model calibration are presented while the characteristics of events used for validation are given in brackets [ ].

Site	Gilby Rd (GR)	Richmond (RICH)	Ruffeys Lake, Doncaster (RD)	Shepherds Bush (SB)	Narre Warren (NW)
Distance from catch centroid to rain gauge (m)	100	600	700	550	250
Mean annual rainfall (mm/year)	723 [536]	650 [500]	650 [370]	580	700 [400]
Mean event maximum rainfall intensity (mm/hr)	10 [7.5]	10 [8.4]	9 [7]	6	10 [8]
Range of event maximum rainfall intensity (mm/hr)	2 - 86 [2 - 36]	2 – 60 [2 - 44]	2 - 44 [2 – 28]	2 - 60	2.5 - 86.3 [2-32]
Mean event maximum runoff rate (L/s)	408 [50]	547 [212]	723 [165]	214	44 [20]
Range of event maximum flow rates (L/s)	75 – 2241 [30 - 200]	67 -3867 [25 - 1430]	164 - 3069 [20 - 908]	29 – 1200	14 -90 [10 – 58]
N. of events - TSS	49	40	54	19	41
Maximum TSS concentration (mg/L)	867	1600	1422	1545	2398
TSS CV (%)**	151.46	164.05	183.12	153.54	182.87
Mean of TSS EMC's* (mg/L)**	71.6	125.1	77.0	94.8	91.9
N. of events - TN	47	39	-	17	18
Maximum TN concentration (mg/L)	9	26	-	15	19
TN CV (%)**	83.18	101.32		85.22	76.82
Mean of TN EMC's* (mg/L)**	1.17	2.29	-	1.74	3.51

* Total impervious fraction (*TIF*)
**Coefficient of variation (CV)
*** Event Mean Concentration (EMC)

A time series of areal potential evaporation (PET) was also required for this study. The values are based on the evaporation data from the Australian Bureau of Meteorology (BOM, 2012) and were used in a daily format (mm/day). However, the time series was obtained from a long-term average, in which the monthly values are constant through the year. Figure 3.4 shows the constant daily values for Melbourne; as expected a seasonal pattern was observed with higher PET during the summer.

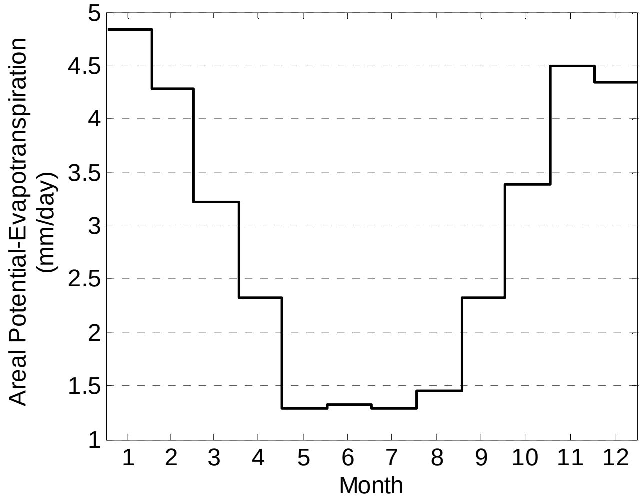


Figure 3.4 Melbourne daily PET (mm).

Water quality data: Approximately 20 to 50 pollutographs are available for each site, each event containing between 5 and 30 discrete samples. At each site, a range of pollutants were available. For this study we used pollutographs of Total Nitrogen (TN) and mostly pollutographs of Total Suspended Solids (TSS) for all catchments, except for RD where only TSS samples were available. These two pollutants were selected as TSS represents pollutants that are associated with particles, while TN is mainly dissolved in water (Taylor, 2006). The variability of TSS and TN concentrations between sites was quite large as shown by their coefficient of variation (CV) in Table 3.2 and Figure 3.5. These large coefficient values also indicate the log nature of these pollutants. The mean Event Mean Concentration (EMC) for TSS ranged from 72 to 125 mg/L between sites, and for TN this was between 1.17 and 3.51 mg/L.

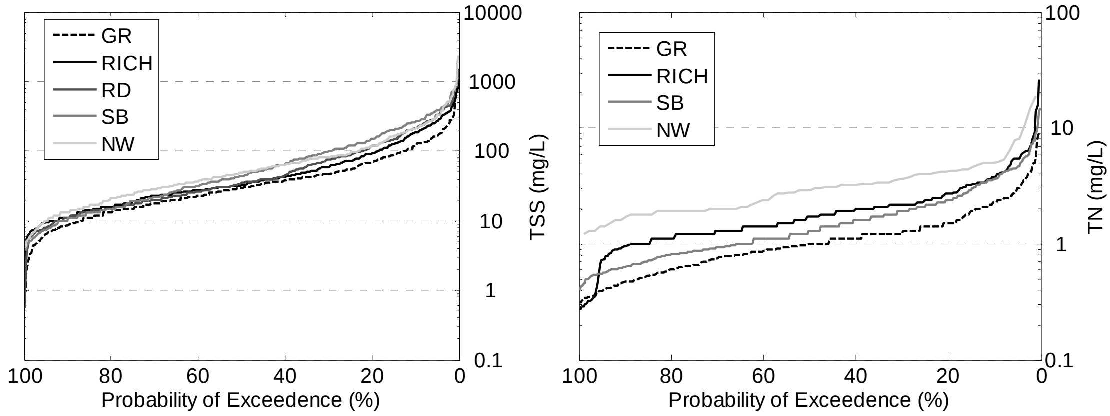


Figure 3.5 Probability of exceedence plots for TSS (left) and TN (right) concentrations measured during wet weather events at each site (mg/L).

3.5 Models used in the study

This research focuses on modelling flows and pollution generation from urban areas located in Melbourne, Australia. For such purpose, the selected models should be able to predict flow rate (rainfall-runoff models) and TSS and TN concentrations (quality models) during wet weather events.

Stormwater models with different levels of complexity were used in order to evaluate their performance and applicability to different domains. Flows and stormwater quality were simulated separately with distinct models. After a thorough literature review (see Chapter 2), two conceptual models MUSIC (eWater CRC, 2012) and KAREN (Rauch and Kinzel, 2007) were selected for runoff modelling, while a process-based build-up/wash-off model (Sartor and Boyd, 1972) and a few empirical regression models (as used in SWMM - USEPA, 2007) were compared in terms of stormwater pollutant modelling. The following subsections summarise the rationale behind this selection and present the description of the selected models.

3.5.1 Rainfall runoff models

The choice of conceptual models over simple empirical or complex process based ones was founded on the advantages and disadvantages presented in Section 2.2.2. For example, empirical models are not suitable for daily and smaller timesteps, which is the focus of this research. In addition, no information on the catchment's hydrologic response is obtained. On the other hand, complex process based models require a large number of inputs and parameters that are not usually available in practice. Conceptual models describe the main processes in the catchment, providing not only reasonable runoff generation, but also information on the catchment's imperviousness and hydrological behaviour (e.g. flow regimes).

MUSIC

The algorithm in MUSIC – Model for Urban Stormwater Improvement (eWater CRC, 2012) is based on the originally daily urban SimHyd model developed by Chiew and McMahon (1999), which was initially developed for large natural catchments. This SimHyd model was modified to enable disaggregation of daily runoff into sub-daily temporal patterns. The model is a simplified description of the rainfall runoff processes in urban catchments and involves the concepts of the impervious area and soil moisture storage. For given rainfall and evapotranspiration time series, MUSIC continuously simulates catchment discharges. MUSIC was designed to operate at a range of temporal and spatial scales, suitable for catchment areas from 0.01 to over 100 km^2. The model operates at timesteps from 6 minutes to 24 hours to match the spatial scale of the catchment being modelled. Previous studies suggested that MUSIC is among the models suitable and recommended for prediction of flow rates from small catchments and also for conceptual or preliminary design at either a subdivision or catchment scale (McAlister et al., 2006; Elliott and Trowsdale, 2007). Flows from impervious and pervious areas are modelled separately (see Figure 3.6). The model components, as well as the main relationships, are explained in the following subsections.

The parameters involved in each of the modelled processes are summarised in Table 3.3 and a full description of the model is available in the MUSIC manual (eWater CRC, 2012). In addition to abbreviations and units, the table refers to the description of each parameter and also reports the default values. Default values were initially obtained based on results from manual calibration of the model with few catchments along the southeast of Australia (e.g. Chiew and McMahon, 1999; eWater CRC, 2012). In Melbourne, for example, MUSIC was calibrated for one single catchment. Table 3.3 also suggests ranges of values that are based on the manual and also on more recent studies that reported MUSIC parameters for different regions (Brisbane City Council, 2006; Gold Coast City Council, 2006; Macleod, 2008).

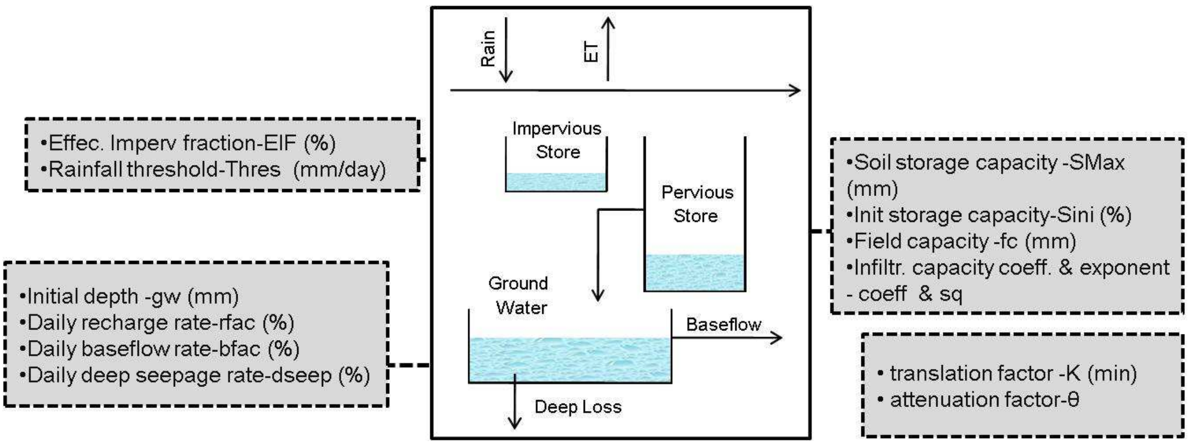


Figure 3.6 Schematic of MUSIC rainfall runoff model and its parameters (after CRCCH, 2005).

Impervious area component. The impervious area runoff is primarily a function of the proportion of catchment imperviousness with initial loss playing a small part. The effective impervious fraction (EIF), which corresponds to the areas that are directly connected to the drainage system, is a calibration parameter. Surface runoff from impervious areas occurs after the rainfall in the catchment exceeds the runoff threshold $(Thres)$ parameter, which defines the minimum daily rainfall before surface runoff occurs from the impervious area according to Equation 3.1:

$$Q_{imp} = max(0, Rain - Thres) \qquad \text{Equation 3.1}$$

where Q_{imp} is the surface runoff rate from impervious areas (mm); $Rain$ is the rainfall incident on the catchment (mm) and, $Thres$ is the initial loss parameter (mm/day).

Pervious area component. The pervious area represents the fraction of the catchment in which infiltration occurs. The infiltration rate (Inf) is defined as an exponential function of the soil moisture storage. Runoff from the pervious areas occurs when the pervious soil storage is either saturated $(SatEx)$ or its infiltration rate (Inf) is exceeded $(InfEx)$. The pervious area runoff related parameters are mainly described by: (1) soil storage capacities, the maximum soil storage $(SMax)$ and the initial storage level $(SIni)$, in which the second is expressed as a percentage of the first, and (2) infiltration factors ($coeff$ and sq). Water from the soil storage is lost due to actual evapotranspiration (ET), which is a function of the current day's potential evapotranspiration (PET) and the ratio between the water currently in the pervious store and its capacity $(S/Smax)$. Equations 3.2 to 3.5 describe the processes in the pervious area:

$$Inf = min(coeff,\ e^{\left(\frac{-sq \cdot SIni}{SMax}\right)}, Rain) \qquad \text{Equation 3.2}$$

$$SatEx = max(SIni - SMax, 0) \qquad \text{Equation 3.3}$$

$$InfEx = Rain - Inf \qquad \text{Equation 3.4}$$

$$ET = min\left(\frac{10SIni}{SMax}, PET\right) \qquad \text{Equation 3.5}$$

where Inf is the pervious soil storage infiltration rate (mm); $coeff$ and sq are the infiltration capacity coefficient and exponent, respectively; $SIni$ is the initial storage level (mm); $SMax$ is the maximum storage capacity of the pervious area store (mm); *Rain* is the rainfall incident on the catchment (mm); $atEx$ is the saturation excess (mm); $InfEx$ is the infiltration excess (mm); ET is the water lost to atmosphere by evaporation (mm); and, PET is the current daily evapotranspiration (mm/day).

Baseflow component. Groundwater is modelled as a store that is recharged when the level in the pervious soil storage exceeds the field capacity (fc). The rate of this recharge is a percentage of the water in the store, which is the calibration parameter $rfac$. This store is emptied via baseflow, which is modelled as a percentage of the water within the store, the model parameter (fc). The rate of this recharge is a percentage of the water in the store, which is the calibration parameter $rfac$. In similar way, deep seepage is set as a percentage of the groundwater store, which is the model parameter $dseep$. Equations 3.6 to 3.8 represent these groundwater processes. Baseflow becomes part of the catchment outflow, deep seepage, on the other hand, is permanently lost from the catchment.

$$Gr = max(0, rfac \cdot (S - fc)) \qquad \text{Equation 3.6}$$

$$Basfl = bfac \cdot gw \qquad \text{Equation 3.7}$$

$$Seep = dseep \cdot gw \qquad \text{Equation 3.8}$$

where Gr is the groundwater recharge; fc is the field capacity (mm); $rfac$ is the daily groundwater daily recharge from the soil store (expressed as a percentage of the volume above the fc in the store); $Basfl$ is the baseflow (mm); gw is the volume of the groundwater store at the start of the simulation (mm); $bfac$ is the daily baseflow rate (expressed as a percentage of the initial groundwater storage gw); $Seep$ is the deep seepage; and, $dseep$ is the daily deep seepage rate (also expressed as a percentage of the initial groundwater volume gw).

Table 3.3 MUSIC rainfall runoff model - Summary of model parameters.

Component	Parameter name	Description	Unit	Default value	Comments**
Impervious Area	Effective impervious fraction (*EIF*)	Fraction of areas that are directly connected to the drainage system	%	-	-
	Rainfall threshold (*Thres*)	Minimum daily rainfall before surface runoff occurs from the impervious	mm	1.0	Values from 0 to 5
Pervious Area	Soil storage capacity (*SMax*)	Maximum soil storage	mm	30*	Values from 30 to 500
	Initial storage (*SIni*)	Initial storage level as a percentage of the Soil storage capacity	%	30	Values from 0 to 50
	Field capacity (*fc*)	When the level in the pervious soil store exceeds the field capacity, groundwater store starts being recharged	mm	20*	Values from 10 to 200
	Infiltration capacity coefficient (*coeff*)	Maximum infiltration loss		200	Values from 0 to 400
	Infiltration capacity exponent (*sq*)	Infiltration loss exponent	-	1	Values from 0 to 7
Groundwater	Daily recharge rate (*rfac*)	Rate of groundwater recharge as percentage of the water in the store	%	25%	-
	Daily baseflow rate (*bfac*)	Rate that the groundwater empties via baseflow as a percentage of the groundwater store	%	5%	-
	Groundwater initial storage (*gw*)	Volume of the groundwater store at the start of the simulation	mm	10	Values from 0 to 100
	Daily deep seepage rate (*dseep*)	Deep seepage rate as a percentage of the groundwater store	%	0	-
Muskingum Cunge	Translation and factor (*K*)	Related to the travel time for the flood wave through the channel reach in minutes	min	30	Values larger than 1/3 of the chosen timestep
	Attenuations factor (θ)	Dimensionless flow weighting factor; lower the value, lager the attenuation	-	0.25	Values from 0.1 to 0.3

*MUSIC default values for Melbourne

**Compiled values from Brisbane City Council (2006), Gold Coast City Council (2006), Macleod (2008) and eWater (2012)

Routing routine. The Muskingum Cunge routing method (Cunge, 1969) is applied for the routing of flows through the drainage system. The method is based on the continuity of mass equation within a channel reach. The basic equations are presented in this subsection and complete discussion of the method can be found in Cunge (1969) and Bedient and Huber (1992). Given the inflow into the channel reach and the outflow from the channel reach, the variation of the volume of water in storage within the channel reach (as illustrated in Figure 3.7) is expressed as:

$$- = \qquad \text{Equation 3.9}$$

where Q_{in} is the inflow into the channel reach; Q_{out} is the outflow from the channel reach; S is the volume of water in storage within the channel reach; and,t is the time.

The total volume of water in the storage within the channel reach can be expressed as a function of the flow rates entering and leaving the reach:

$$Q_{out} = \frac{S - K\theta Q_{in}}{1 - \theta} \qquad \text{Equation 3.10}$$

where K is approximately equal to the travel time for the flood wave through the channel reach in minutes. It is recommended that K should not assume values less than one third of the chosen timestep. θ is a dimensionless weighting factor that has a value between 0 and 0.5, and is generally between 0.1 and 0.3 for natural channels. When θ is zero, the volume of water in storage is purely a function of the outflow. A value of 0.5 produces no attenuation and the flood wave is purely translated by a time value equals to K. MUSIC assumes that the values of K and θ remain constant within a reach throughout the simulation and are considered as model calibration parameters (Table 3.3).

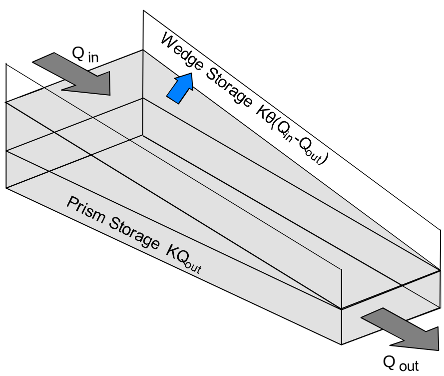


Figure 3.7 Schematic of storages in a channel reach (after CRCCH, 2005).

KAREN

KAREN is a simple linear reservoir model (Rauch and Kinzel, 2007). For a given rainfall time series, the model generates a series of flows originating from impervious areas only. The pervious components of the catchments are not considered. The main relationships are explained in this section. Table 3.4 presents a summary of the parameters included in the model. For a full description of the model, see the manual (Rauch and Kinzel, 2007). In addition, Table 3.4 presents a description of each parameter. Whenever available, the table reports the default values of parameters (according to the manual - Rauch and Kinzel, 2007). A schematic presentation of the rainfall runoff model implemented in KAREN is given in Figure 3.8.

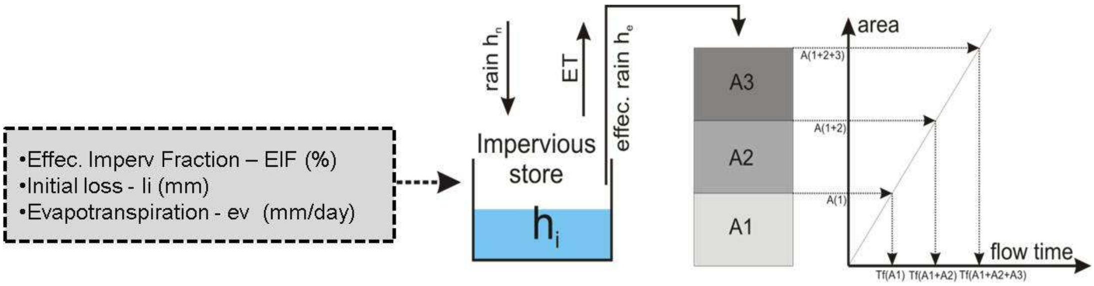


Figure 3.8 Schematic of KAREN rainfall runoff model and its parameters(Kleidorfer et al., 2009).

The model is similar to that found in MUSIC but neglects all processes/parameters relating to runoff from pervious areas. Again, the effective impervious area of the catchment is an important parameter (EIF) to be calibrated. Runoff from impervious areas is generated depending on whether a certain rainfall threshold has been exceeded. Such threshold is represented by the initial loss parameter (li) and it is modelled as a single reservoir. Furthermore, here the initial loss is not a minimum daily rainfall sum but a total (time-independent) value, which fills during rainfall and drains during dry weather periods depending on a permanent loss, which is the evaporation calibration parameter (ev). The effective rainfall is calculated as the difference between the measured rainfall and an initial loss:

$$h_e = h_m - li \qquad \text{Equation 3.11}$$

where h_e is the effective rainfall intensity (mm); h_m is the measured rainfall intensity (mm); and, li is the initial loss parameter (mm).

The initial loss is drained during dry weather periods, in each timestep *t* depending on a permanent loss:

$$li_t = li_{t-1} - ev \qquad \text{Equation 3.12}$$

where li_t is the initial loss at the timestep *t*; li_{t-1} is the initial loss at the previous timestep; and, ev is a permanent loss through evaporation (mm/day).

Surface runoff concentration is calculated using the linear time-area method, which is similar to the unit hydrograph method (Sherman, 1932). At the beginning of a rainfall event, the effective impervious area (EIF) is increased according to the flow time on the catchment surface until the whole catchment contributes to runoff after the time of concentration, which is a calibration parameter (TOC). The runoff is calculated as:

$$Q_n = \sum_{m=1}^{n \le M} I_{e,m}\ A_{n-m+1}\ 10^{-3} \qquad \text{Equation 3.13}$$

where Q_n is the runoff (m^3/s), n is the index of the runoff; I is the effective rainfall intensity (mm/s); m is the index of the rainfall; and, A is the current effective impervious area (m^2).

Table 3.4 KAREN rainfall runoff model - Summary of model parameters.

Component	Parameter name	Description	Unit	Default value
Impervious Area	Effective impervious fraction (*EIF*)	Fraction of areas that are directly connected to the drainage system	%	-
	Concentration time (*TOC*)	Related to the time that the whole catchment contributes to runoff	Min	-
	Initial loss (l_i)	Minimum rainfall before surface runoff occurs from the impervious	mm	1
	Evapotranspiration (*ev*)	Permanent loss during dry weather	mm/day	1.5

3.5.2 Water quality models

Contrary to the water quantity models, reliable stormwater pollution generation models are almost non-existent (Elliott and Trowsdale, 2007). Understanding the processes within pollution generation is very important for the development of better modelling approaches. In general, these processes are very complex and are influenced by a variety of factors, such as: rainfall, runoff, climatic, land use and surface characteristics (Deletic and Maksimovic, 1998; Vaze and Chiew, 2002; Egodawatta et al., 2007; McCarthy et al., 2011). This complex nature of pollutant accumulation and wash-off, together with high temporal and spatial variations, generates technical difficulties in the development of accurate and reliable models of pollutant processes. Few approaches are available for reproducing the catchment's response in terms of pollutants (e.g. Sartor and Boyd, 1972; Vaze and Chiew, 2003).

The approaches vary among the popular modelling packages (Elliott and Trowsdale, 2007). For example, MOUSE (DHI, 2002c; 2004) and SLAMM (Pitt, 1998) use the build-up/wash-off method (based on work of Sartor and Boyd, 1972). Empirical power rating curves for concentration as function of rainfall intensity and flow rate are also included in some of the models (e.g. SWMM -USEPA, 2007). However, they are difficult to calibrate and validate as they seem unable to accurately reproduce the pollutant's behaviour in the systems (Beck, 1987; Kanso et al., 2006; Egodawatta et al., 2007). MUSIC (eWater CRC, 2012), SLAMM (Pitt, 1998) and XP-SWMM (WP Software, 1995) come with a stochastic component, which has been widely used. This poses a challenge for calibration as the models always generate different values. Simple statistical models, such as investigated by Blasone et al. (2008) and Lee and Heaney (2003) cannot be used outside catchments for which they are developed. To advance these models it is important to understand the sources of their uncertainties. It has been recognised that one of the main problems in the establishment of water quality models is the high level of uncertainty in their calibrated parameters (Fletcher et al., 2004; Francey et al., 2010).

A process-based build-up/wash-off model and a few empirical regression models were compared. Although poor performance of these models is expected, they are commonly used in practice (Palmstrom and Walker, 1990; XP-SOFTWARE, 1999; USEPA, 2007), and were assessed in terms of parameter calibration and model sensitivity analysis in order to (i) guide future development of such models and indicate the data required to support their development and application, and (ii) test the applicability of different sensitivity analysis methods to 'ill-posed' models.

Build-up/wash-off model

The generation of pollutants in the runoff from an impervious surface is often described and modelled using the concepts of build-up and wash-off. The attempt to model these both processes was proposed by Sartor and Boyd (1972) and is summarized in Sartor et al. (1974). It has been tested (Deletic et al., 2000; Shaw et al., 2010) and derivations have been adopted in several stormwater softwares, such as SWMM (USEPA, 2007) and the US Army Corps's STORM model (XP-SOFTWARE, 1977). A modified version, of the one originally proposed by Sartor and Boyd (1972), was tested. The build-up of pollutants during dry weather periods is calculated as

$$\textbf{Build}-\textbf{up}:\quad M(t_d) = \quad M_0 \cdot (1 - e^{-\frac{6k_1}{1440}t_d}) \qquad \text{Equation 3.14}$$

where M is the amount of pollutant on the surface (in g/m² in Chapters 4 and 5 and kg in Chapter 8) and t_d is the dry period in, here being the 6 minutes timestep. Again, this equation has two calibration parameters: M_0, which is the maximum amount of solids expected at the surface (in g/m² in Chapters 4 and 5 and kg in Chapter 7) and k_1 that represents an accumulation constant (day^{-1}).

Two versions of the wash of model were used. The first represented by the Equation 3.15:

$$\textbf{Wash}-\textbf{off}: C(t) = \quad k_2 \cdot M(t) \cdot I(t)^{k_3} \cdot A \qquad \text{Equation 3.15}$$

where $\bar{C}$ is the concentration in runoff (mg/L), I is the rainfall intensity (mm/hr), and, A is the impervious area (m²). The calibration parameters are k_2, the wash-off coefficient and k_3, which is the wash-off exponent. In total, there are four calibration parameters (summarised in Table 3.5).

Another modified wash-off model, which includes a transport component was also tested. It calculates the amount of pollutants washed from the surface and the concentration of pollutants in the runoff within a timestep as a power function of the catchment's runoff (runoff rate and volume):

$$\textbf{Wash}-\textbf{off}: C(t) = \quad k_2 \cdot M(t_d) \cdot \frac{q}{RC}(t-r)^{k_3} \qquad \text{Equation 3.16}$$

$$W(t) = 10^{-6} * C(t) * Vol(t) \qquad \text{Equation 3.17}$$

where C is the concentration of pollutants (mg/L); q is the modelled runoff (mm/hr) and RC is the catchment runoff coefficient, here assumed as the effective impervious fraction of the catchment obtained from the rainfall runoff model, EIF). RC was included to represent wash-off only from impervious surfaces, which is a safe assumption because the majority of runoff from urban catchments originates from impervious surfaces (Chiew and McMahon, 1999). Modelling water quality with modelled q values is widely used in practice where measured data are scarce or not available. If instead of q, rainfall intensities were used, the model would have to include a routing algorithm (e.g. linear reservoir routing method) which means an additional equation and at least one extra parameter to attenuate the rainfall. In addition, the uncertainty in the model would increase due to the errors in extra input data (rainfall records). And finally the use of modelled q and Vol accounts for the changes in the time of concentration existing between different events (Vezzaro et al., in press). As in the original approach, two calibration parameters are involved in the wash-off process: k_2 as the washoff coefficient, and k_3 which is the washoff exponent. In addition, a transport related parameter (r) was added into the model to represent the small lag time which is often noted between the hydrographs and the pollutographs (Vaze and Chiew, 2003). The runoff is translated by a number of timesteps in min. The amount of pollutants washed from the surface (W in kg) is then calculated in function of the predicted concentration and the runoff volume (Vol in L). Table 3.5 presents a summary of the parameters included in the model.

Table 3.5 Build-up/wash-off model versions 1 and 2- Summary of model parameters.

Component	Parameter name	Interpretation	Unit	Comments
Build-up	Maximum amount of pollutant (M_0)	Maximum amount of pollutant on the surface before a rain event	g/m² (Version 1) kg (Version2)	• Values from 3 and 18 g/m² for suspended solids in urban stormwater (Sriananthakumar and Codner, 1992; Tomanovic and Maksimovic, 1996; Hossain et al., 2010) • Values from 200 to 560 kg for suspended solids and from 2.6 to 3.5 kg for TN in stormwater alley (Alley and Smith, 1981)
	Accumulation constant (k_1)	Accumulation rate of pollutant during dry weather period	day^{-1}	• Values of 0.098 and 0.38 for suspended solids in combined sewers (Kanso et al., 2003) • Values from 0.015 to 0.2 for sediments in stormwater (Alley and Smith, 1981; Sriananthakumar and Codner, 1992; Tomanovic and Maksimovic, 1996) • Values of 0.05 and 0.08 for TN in stormwater (Sriananthakumar and Codner, 1992)
Wash-off	Wash-off coefficient (k_2)	Related to the sources of pollutants in the catchment	-	• Values of 0.049 and 0.073 for concentration of suspended solids in combined sewers (Kanso et al., 2003) • Values from 0.002 to 0.2 for loads of suspended solids (from impervious surfaces) in urban stormwater (Hossain et al., 2010)
	Wash-off exponent (k_3)	Related to the kinetic energy of the rainfall in detaching pollutant from the surface or to the shear stress generated by flow	-	• Values of 1.3 and 1.2 for concentration of suspended solids in combined sewers (Kanso et al., 2003) • Values from 0.3 to 0.7 for loads of suspended solids (from impervious surfaces) in urban stormwater (Hossain et al., 2010) • Value of approximately 0.29 for sediments in stormwater. (McCarthy et al., 2011)
	Translation factor (r in Version 2)	Represents the lag between the hydrographs and the pollutographs	min x number of timesteps	-

Regression models

The simple regression models adopted in this study estimate concentrations within a timestep as a power function of either the catchment's runoff or rainfall intensity. Derivations of this regression models are used in practice in several stormwater models, such as XP-AQUALM (XP-SOFTWARE, 1999), SWMM (USEPA, 2007) and P8-UCM (Palmstrom and Walker, 1990). Three different equations were tested:

$$C(t) = a \ I(t)^b \qquad \text{Equation 3.18}$$

$$C(t) = a \ I(t)^b_{Routed} \qquad \text{Equation 3.19}$$

$$C(t) = a \ q(t)^b \qquad \text{Equation 3.20}$$

where C is the pollutant concentration (mg/L) at time *t*. I is the rainfall intensities (mm/hr). The calibration parameters are a and b. In which a relates the amount of pollutants on the surface of the catchment and b relates to the kinetic energy of the rainfall in detaching pollutant from the surface (Equation 3.18 and Equation 3.19) or to the shear stress generated by flow (Equation 3.20). Because the build-up is not considered, these wash-off curves assume that the amount of available material in the catchment before every event is constant. I_{Routed} is the routed rainfall intensities (mm/hr). The intensities were translated and attenuated with the Muskingum Cunge routing method described in the Section 3.5.1. q is the measured or modelled runoff (mm/hr), for which K and θ parameters have to be calibrated. Table 3.6 presents a summary of the model parameters.

Table 3.6 Regression models - Summary of model parameters.

Parameter name	Description	Unit	Comments
Water quality scale coefficient (a)	Related to the sources of pollutants in the catchment	-	
Water quality shape coefficient (b)	Related to the kinetic energy of the rainfall in detaching pollutant from the surface or to the shear stress generated by flow	-	Value of approximately 0.29 for sediments in stormwater. (McCarthy et al., 2011)
Translation and factor (K)	Related to the translation of the pollutograph through the channel reach	min	-
Attenuations factor (θ)	Related to the attenuation of the pollutograph through the channel reach	-	-

As described in this section, the models were arranged to predict pollutant concentrations from urban catchments. In the literature, however, the models have been often calibrated for pollutant loads (e.g. Vaze and Chiew, 2003; Francey, 2010). Very little is available for concentrations, mainly in separate sewer systems.

3.6 Chapter summary

This chapter focused on presenting the data and models used for this research. Five study catchments, located in Melbourne Australia, were used to represent the different land-uses (industrial/commercial and residential), catchment areas (from 10 to 100 ha) and levels of development/imperviousness (total impervious fraction varying from 0.2 to 0.8). Data on rainfall, runoff flows and concentrations of TSS and TN recorded from 2004 to 2007 were used for this research.

Two rainfall runoff models were selected for the study. MUSIC was chosen mainly due to its widespread application in Australia and KAREN was adopted due to its simpler design. Both MUSIC and KAREN are used to estimate the runoff generated from urban areas continuously and require a series of rainfall and the catchment area as the main inputs. The difference in the number of calibration parameters and the processes that are simulated in each of the models exemplify the difference in their complexity. MUSIC presents thirteen parameters to be calibrated, while KAREN presents only four. MUSIC estimates flows from impervious and pervious areas separately as a series of reservoirs, while KAREN predicts the runoff only from impervious areas using a single reservoir model. The evaluation of such models will define their applicability to different domains.

Two conceptual process-based build-up/wash-off models and three empirical regression models to quantify stormwater pollutant concentrations were described. They were chosen for this study because they are usually adopted in the most used stormwater modelling packages. The process-based models account for both build-up and wash-off processes, while the regressions estimate only wash-off. Evaluation of these models will help to guide their future improvement and indicate the data required to support new model development and application. The described models will be used through the thesis for a variety of tasks.

3.7 References

Beck, M. B. (1987). Water Quality Modeling: A Review of the Analysis of Uncertainty. *Water Resources Research* **23**(8): 1393-1442.

Bedient, P. B. and Huber, W. C. (1992). *Hydrology and Flood Plain Analysis*. New York, Addison-Wesley.

Blasone, R.-S., Madsen, H. and Rosbjerg, D. (2008). Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* **353**(1-2): 18-32.

BOM (2012). Bureau of Meteorology (www.bom.gov.au). Accessed in the year 2012.

Brisbane City Council (2006). Estimation of MUSIC parameters for use in Brisbane city.

Chiew, F. H. S. and McMahon, T. A. (1999). Modelling runoff and diffuse pollution loads in urban areas. *Water Science and Technology* **39**(12): 214-248.

CRCCH (2005). MUSIC Version3 - User guide, Cooperative Research Centre for Catchment Hydrology.

Cunge, J. A. (1969). On the Subject of a Flood Propagation Computation Method (Muskingum Method). *Journal of Hydraulic Research* **7**(2): 205-230.

Deletic, A., Ashley, R. and Rest, D. (2000). Modelling Input of Fine Granular Sediment into Drainage Systems via Gully-Plots. *Water Research* **34**(15): 3836-3844.

Deletic, A. and Maksimovic, C. T. (1998). Evaluation of water quality factors in storm runoff from paved areas. *Journal of Environmental engineering - ASCE* **124**: 869-879.

DHI (2002c). MOUSE pipe flow reference manual. Software, D. Horsolm, Denmark.

DHI (2004). MOUSE surface runoff models reference manual. *DHI Software*. Horsolm, Denmark.

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

Elliott, A. H. and Trowsdale, S. A. (2007). A review of models for low impact urban stormwater drainage. *Environmental Modelling and Software* **22**(3): 394-405.

eWater CRC (2012). Model for urban stormwater improvement conceptualisation (MUSIC) (Version 5.0), Canberra: eWater Cooperative Research Centre.

Fletcher, T. D., Wong, T. H. F., Duncan, H. P., Deletić, A., Coleman, J. R. and Jenkins, G. A. (2004). Recent advances in decision-supprt for stormwater management: the MUSIC approach. *6th international conference on sustainable techniques and strategies in urban water management - NOVATECH 2004*. Lyon, France.

Francey, M. (2010). Characterising urban pollutant loads. *Civil Engineering*. Melbourne, Monash University. **PhD thesis:** 248.

Francey, M., Fletcher, T. D., Deletic, A. and Duncan, H. (2010). New insights into water quality of urban stormwater in South Eastern Australia. *Journal of Environmental Engineering* **136**(4): 381-390.

Gold Coast City Council (2006). MUSIC modelling guidelines.

HACH, C. (2008). Sigma 950 Flow Meter - Instrument Manual. 2009, from http://www.hach.com/fmmimghach?/CODE%3A3314-2008-1117081%7C1.

Kanso, A., Chebbo, G. and Tassin, B. (2006). Application of MCMC-GSA model calibration method to urban runoff quality modeling. *Reliability Engineering and System Safety* **91**(10-11): 1398-1405.

Kleidorfer, M., Deletic, A., Fletcher, T. D. and Rauch, W. (2009). Impact of input data uncertainties on stormwater model parameters. *Water Science and Technology* **60**(6): 1545 - 1554.

Lee, J. G. and Heaney, J. P. (2003). Estimation of Urban Imperviousness and its Impacts on Storm Water Systems. *Journal of Water Resources Planning and Management* **129**(5): 419-426.

Leecaster, M. K., Schiff, K. and Tiefenthaler, L. L. (2002). Assessment of efficient sampling designs for urban stormwater monitoring. *Water Research* **36**(6): 1556-1564.

Macleod, A. (2008). MUSIC calibration based on soils conditions. *SW and QLD Stormwater Industry Association*. Surfers Paradise, QLD, Australia.

McAlister, T., Mitchell, G., Fletcher, T. and Phillips, B. (2006). Modelling Urban Stormwater Management Systems. In: Wong, T. H. F. *Australian Runoff Quality*. 153, Institution of Engineers, Australia.

McCarthy, D. T. (2008). Modelling microorganisms in urban stormwater. *Civil Engineering*. Melbourne, Monash University, Australia. **PhD thesis:** 488.

McCarthy, D. T., Deletic, A., Mitchell, V. G. and Diaper, C. (2011). Development and testing of a model for Micro-Organism Prediction in Urban Stormwater (MOPUS). *Journal of Hydrology* **409**(1-2): 236-247.

Palmstrom, N. and Walker, W. W. J. (1990). P8 urban catchment model: User's guide. Program documentation, and evaluation of existing models, design concepts, and Hunt-Potowomut data inventory. NBP-90-50, T. N. B. P. R. N.

Pitt, R. (1998). Unique features of the Source Loading and Management Model (SLAMM). *James, W. (Ed.), Modeling the Management of Stormwater Impacts* **6** (Computational Hydraulic International): 13-35.

Rauch, W. and Kinzel, H. (2007). KAREN - User Manual, hydro-IT GmbH, Innsbruck, (in German).

Sartor, J. D. and Boyd, G. B. (1972). Water pollution aspects of street surface contaminants. *Rep. EPA-R2-72-081*. Washington, D. C., U.S. Environ. Prot. Agency.

Sartor, J. D., Boyd, G. B. and Agardy, F. J. (1974). Water pollution aspects of street surface contaminants. *Journal of the Water Pollution Control Federation* **46**(3): 458-467.

Shaw, S. B., Stedinger, J. R. and Walter, M. T. (2010). Evaluating Urban Pollutant Buildup/Wash-Off Models Using a Madison, Wisconsin Catchment. *Journal of Environmental Engineering* **136**(2): 194-203.

Sherman, L. K. (1932). Stream-flow from Rainfall by the Unit-Graph Method. *Engineering News-Record 108: 501-505*. **108:** 501-505.

Taylor, G. D. (2006). Improved effectiveness of nitrogen removal in constructed stormwater wetlands. Civil Engineering Department. Monash University, Australia.

USEPA (2007). Storm Water Management Model (SWMM) - User Manual version 5.0. US Environmental Protection Agency (USEPA). *US Environmental Protection Agency - EPA/600/R-05/039*. Washington DC.

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

Vaze, J. and Chiew, F. H. S. (2003). Comparative evaluation of urban storm water quality models. *Water Resources Research* **39**(10): 1280.

Vezzaro, L., Eriksson, E., Ledin, A. and Mikkelsen, P. S. (in press). Quantification of uncertainty in modelled partitioning and removal of heavy metals (Cu, Zn) in a stormwater retention pond and a biofilter. *Water Research*(0).

WP Software (1995). *XP-SWMM stormwater management model with XP graphical interface, User's manual, Version 2*. Tampa, Forida and Belconnen, USA.

XP-SOFTWARE (1977). XP-Storm - Technical Description. *XP Software*. Australia.

XP-SOFTWARE (1999). XP-AQUALM - Technical Description, XP Software, Australia.

Chapter 4

Sensitivity and uncertainty analysis: methods for stormwater models

DECLARATION FOR THESIS CHAPTER 4

Declaration by candidate

In the case of Section 4.2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Initiation, ideas, methodology set up, data preparation and interpretation, analysis of the results, synthesis of the results from different groups, and leading write-up.	35

The following co-authors contributed to the work:

Name	Nature of contribution	Extent of contribution (%)
Giorgio Mannina	Ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	20
Manfred Kleidorfer	Ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	10
Luca Vezzaro	Ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	10
Malte Henrichs	Ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	10
Gabriele Freni	Ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	10
David T. McCarthy	Initiation, ideas and reviewing	n/a
Wolfgang Rauch	Initiation, ideas and reviewing	n/a
Ana Deletic	Initiation, ideas and reviewing	n/a

Candidate's Signature	[redacted]	Date 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location	Dipartimento di Ingegneria Civile, Ambientale e Aerospaziale, University of Palermo, Italy	
Signature Giorgio Mannina	[redacted]	**Date** 29/10/2012
Location	Unit for Environmental Engineering, University of Innsbruck, Austria	
Signature Manfred Kleidorfer	[redacted]	**Date** 29/10/2012
Signature Wolfgang Rauch	[redacted]	**Date** 29/10/2012
Location	Department of Environmental Engineering (DTU Environment), Technical University of Denmark, Denmark	
Signature Luca Vezzaro	[redacted]	**Date** 29/10/2012

Location	Muenster University of Applied Sciences, Department of Civil Engineering, Laboratory of Water Resources Management, Germany	
Signature Malte Henrichs	[redacted]	**Date** 29/10/2012
Location	Facolta` di Ingegneria ed Architettura, Kore di Enna University, Italy	
Signature Gabriele Freni	[redacted]	**Date** 29/10/2012
Location	CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus	
Signature David T. McCarthy	[redacted]	**Date** 29/10/2012
Signature Ana Deletic		**Date** 29/10/2012

Declaration by candidate

In the case of Section 4.3, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up.	70

The following co-authors contributed to the work:

Name	Nature of contribution	Extent of contribution (%)
Ana Deletic	Initiation, ideas and reviewing	n/a
Tim D. Fletcher	Initiation, ideas and reviewing	n/a

Candidate's Signature	[redacted]	**Date** 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location	CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus

Signature Ana Deletic	[redacted]	**Date** 29/10/2012

Location	Waterway Ecosystem Research Group. Melbourne School of Land & Environment. The University of Melbourne

Signature Tim D. Fletcher	[redacted]	**Date** 29/10/2012

4.1 Introduction

As the literature review demonstrated (see Sections 2.3 and 2.4), different methods have been applied for parameter calibration, model sensitivity and uncertainty analysis in stormwater management practice and related fields (e.g. environmental and hydrological modelling). Although, some of these approaches have been used in assessment of urban drainage models, the application and comparison among methods have not been systematically investigated. Therefore, the first aim of this research was to ***identify suitable global uncertainty analysis method(s) to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models.***

The objectives were to address the following key research questions and hypotheses:

- Given that all global uncertainty analysis methods have advantages and limitations, which methods generate more comprehensive results with respect to model calibration and predictive uncertainty (e.g. results from more formal mathematical based methods or more information about parameter interaction)?
    - Different uncertainty methods lead to different results with respect to model parameter sensitivity and predictive uncertainty because they rely on different formulations (e.g. formal probabilistic or not) and assumptions (e.g. assumption about the model errors structure, such as the assumption that the residuals are independent and normally distributed).
- What are the main requirements of different methods in terms of computational resources and modeller skill/knowledge level?
    - There is a complex interaction between the complexity of the method used, the computational time required and the knowledge/skill level of the modeller.

The aims, methods and results of this assessment have been published as two separate journal papers. In order to address the knowledge gap that the application and comparison among uncertainty methods have not been systematically investigated in the urban drainage field, an international research project commenced in 2008 to address this gap. The main objective was to compare different methods often used for uncertainty assessment of the parameters in urban water related fields. The project was led by the candidate who coordinated work of research teams from Australia, Italy, Austria, Denmark and Germany. The project went for two years and the results, *Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling*, are now published in *Water Research.* This paper is the body of text of Section 4.2. Subsequently, a preliminary application of the uncertainty method (a Bayesian approach) for stormwater flow and quality modelling is provided. This work, *Analysis of parameter uncertainty of a flow and quality stormwater model*, was first presented at the *11th International Conference on Urban Drainage,* held in Edinburgh, Scotland in 2008 and it was subsequently selected for publication in *Water Science and Technology*. This paper was published in 2009 (cited 14 up-to-date) and it is included in Section 4.2.

4.2 Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling

WATER RESEARCH 46 (2012) 2545–2558

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling

Cintia B.S. Dotto[a,*], Giorgio Mannina[b], Manfred Kleidorfer[c], Luca Vezzaro[d], Malte Henrichs[e], David T. McCarthy[a], Gabriele Freni[f], Wolfgang Rauch[c], Ana Deletic[a]

[a] Centre for Water Sensitive Cities, Department of Civil Engineering, Monash University, Australia
[b] Dipartimento di Ingegneria Civile, Ambientale e Aerospaziale, University of Palermo, Italy
[c] Unit for Environmental Engineering, University of Innsbruck, Austria
[d] Department of Environmental Engineering (DTU Environment), Technical University of Denmark, Denmark
[e] Muenster University of Applied Sciences, Department of Civil Engineering, Laboratory of Water Resources Management, Germany
[f] Facoltà di Ingegneria ed Architettura, Kore di Enna University, Italy

ARTICLE INFO

Article history:
Received 15 March 2011
Received in revised form
17 January 2012
Accepted 5 February 2012
Available online 11 February 2012

Keywords:
Urban drainage models
Uncertainties
Parameter probability distributions
Bayesian inference
GLUE
SCEM-UA
MICA
AMALGAM
MCMC
Multi-objective auto-calibration

ABSTRACT

Urban drainage models are important tools used by both practitioners and scientists in the field of stormwater management. These models are often conceptual and usually require calibration using local datasets. The quantification of the uncertainty associated with the models is a must, although it is rarely practiced. The International Working Group on Data and Models, which works under the IWA/IAHR Joint Committee on Urban Drainage, has been working on the development of a framework for defining and assessing uncertainties in the field of urban drainage modelling. A part of that work is the assessment and comparison of different techniques generally used in the uncertainty assessment of the parameters of water models. This paper compares a number of these techniques: the Generalized Likelihood Uncertainty Estimation (GLUE), the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), an approach based on a multi-objective auto-calibration (a multialgorithm, genetically adaptive multi-objective method, AMALGAM) and a Bayesian approach based on a simplified Markov Chain Monte Carlo method (implemented in the software MICA). To allow a meaningful comparison among the different uncertainty techniques, common criteria have been set for the likelihood formulation, defining the number of simulations, and the measure of uncertainty bounds. Moreover, all the uncertainty techniques were implemented for the same case study, in which the same stormwater quantity and quality model was used alongside the same dataset. The comparison results for a well-posed rainfall/runoff model showed that the four methods provide similar probability distributions of model parameters, and model prediction intervals. For ill-posed water quality model the differences between the results were much wider; and the paper provides the specific advantages and disadvantages of each method. In relation to computational efficiency (i.e. number of iterations required to generate the probability distribution of parameters), it was found that SCEM-UA and AMALGAM produce results quicker than GLUE in terms of required number of simulations. However, GLUE requires the lowest modelling skills and is easy to implement. All non-Bayesian methods have problems with the way they accept behavioural parameter sets, e.g. GLUE, SCEM-UA and AMALGAM have subjective acceptance thresholds, while MICA has usually problem with

* Corresponding author.
E-mail address: [redacted] (C.B.S. Dotto).

doi:10.1016/j.watres.2012.02.009

its hypothesis on normality of residuals. It is concluded that modellers should select the method which is most suitable for the system they are modelling (e.g. complexity of the model's structure including the number of parameters), their skill/knowledge level, the available information, and the purpose of their study.

1. Introduction

The application of stormwater models as design, planning and/or management tools has become common practice in the urban drainage field. These models, as any other mathematical models, represent only a fraction of reality leading to uncertain results; in fact uncertainties are inherent in all models and their total elimination is not possible (Beck, 1987; Harremoës, 2003). Therefore, uncertainty analysis must be performed for urban drainage models in order to quantify the level of reliability of the model results and provide a robust basis for their application in practice (e.g. to provide a level of confidence for a model used for risk analysis).

As pointed out by Beven (2009), there are many sources of uncertainty that interact non-linearly in the modelling process. Not all uncertainty sources can be 'quantified', and the fraction of uncertainty sources being 'ignored' might be significant in environmental investigations (Rauch and Harremoës, 1999). The following sources of uncertainties are commonly listed in the environmental modelling literature (e.g. Butts et al., 2004): (i) model parameters, (ii) input data, (iii) calibration data, and (iv) model structure. The impact of calibration methods, choice of objective functions, and calibration data availability are also recognised.

Many published studies investigated the impact of the uncertainties on urban drainage model predictions (e.g. Kanso et al., 2003; Lindblom et al., 2007; Willems, 2008; Kleidorfer et al., 2009; Dotto et al., 2010). For example, Freni and Mannina (2010a) used the variance decomposition method to understand the significance of different uncertainty sources in an integrated urban water quality model (composed of a sewer system, wastewater treatment plant and a river). According to this analysis, the uncertainty of the sewer water quality model was one order of magnitude higher than the uncertainty of the quantity model.

Assessing uncertainties in models is not wide spread in practice and is usually an academic exercise; indeed, urban drainage models are no exception. This is mainly because the techniques required for this analysis are so numerous, highly complex, poorly understood, and some are still highly underdeveloped. The need of step-by-step manuals is also recognised. Clear and comprehensive comparisons of these techniques when applied to typical drainage models would therefore be desirable. Some examples of such comparisons are now available in literature. Recently, Freni et al. (2009) compared the classical Bayesian Monte Carlo (i.e. Bayesian inference approach) and the pseudo-Bayesian approaches (i.e. Generalized Likelihood Uncertainty Estimation – GLUE) employing eight different stormwater quality models. They found that the methods performed similarly when GLUE is based on the same assumptions as the Bayesian Monte Carlo approach. However, GLUE is sensitive to the user-defined threshold for accepting parameter sets, which might lead to an under- or over-estimation of uncertainties.

The different approaches have also been compared in related fields other than urban drainage; for example, Makowski et al. (2002) compared GLUE and Markov Chain Monte Carlo (MCMC) methods (in particular the Metropolis–Hasting sampling approach) using a simplified crop model with 22 parameters. Both methods presented similar results, but the authors recommend the use of the Metropolis–Hasting algorithm. This is because this method converges to the true posterior distribution even if the model includes a large number of parameters, while GLUE found this challenging because it required a large number of simulation runs. However, MCMC procedures also have their own limitations, and a misspecification of the error structure (or likelihood function) in the Bayesian approach can lead to an erroneous quantification of the model predictive uncertainty (Beven et al., 2008). Although these insights are valuable, it is difficult to relate the results between different comparison studies, as they employ different models and different datasets.

The objective of this paper is to provide a comprehensive comparison of the most common methods used for assessing urban drainage model parameter uncertainties. This comparison highlights the advantages and disadvantages of each method when used in urban drainage practice. Four uncertainty techniques are compared:

1. The Generalized Likelihood Uncertainty Estimation (GLUE) developed by Beven and Binley (1992);
2. The Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) by Vrugt et al. (2003a, b), applied in combination with GLUE by Blasone et al. (2008);
3. A multialgorithm, genetically adaptive multi-objective method (AMALGAM) by Vrugt and Robinson (2007); and,
4. The classical Bayesian approach based on a Markov Chain Monte Carlo method and the Metropolis–Hastings sampler (implemented in the software MICA by Doherty, 2003).

As these techniques are based on different philosophies and hypotheses, an exact comparison is not possible. Therefore, common criteria for the comparison have been established, mainly with regards to practical applicability as the number of simulations required, the 90% probability bands and the likelihood measure. A simplified stormwater quantity and quality model was employed for the comparison using an Australian case study for which long-term high resolution data was available. The key finding is that all methods provided similar results on the rainfall/runoff model sensitivity to its parameters, but the probability distributions of model parameters and the prediction intervals were slightly different for the water quality model. Methods generally differed in efficiency (i.e. required less iterations and consequently less computing time).

2. Material and methods

2.1. Uncertainty techniques

2.1.1. The Generalized Likelihood Uncertainty Estimation – GLUE

GLUE is based on Monte Carlo simulations, where the random sampling of individual parameters from prior probability distributions is used to determine a set of parameter values. Parameter sets are compared with respect to their ability to reproduce available observations by the means of a user-defined likelihood measure. Specifically, sets with poor likelihood weights (e.g. lower than a selected threshold) are discarded and are classified as "non-behavioural". All other parameter sets are retained, and their likelihood weights, rescaled to, so that their cumulative probability sum is equal to 1. The uncertainty bands of model predictions are then calculated as per Beven and Binley (1992).

The GLUE methodology is attractive because there is no need for an assumption of the error distribution functions. However, GLUE can be very computationally demanding. The GLUE results can be affected by the definition of parameter variation ranges. A small parameter range can lead to unrealistic uncertainty in the model outputs, while wider ranges (by giving more significant information of the influence of parameters) can, for example, increase the computational cost of the analysis (Beven and Binley, 1992). However, the key criticism of GLUE is the fact that it relies on some subjective criteria that can impact the results (Freni et al., 2008, 2009; Li et al., 2010). For example, it employs a user-defined likelihood measure that often is a subjective acceptance threshold which has large impact on results. To overcome this problem, other likelihood measures have been introduced in literature (Freer et al., 1996) and formal probability distributions may be adopted (Romanowicz et al., 2000). However, it can be argued that this transforms GLUE into a more formal Bayesian method.

2.1.2. The Shuffled Complex Evolution Metropolis Algorithm – SCEM-UA

Optimization algorithms, such as MCMC methods, have been applied to reduce the computational burden by exploring the parameter space, identifying the region of higher likelihood and increasing the parameters sampling in these regions. In this context, the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) was initially developed by Vrugt et al. (2003a) in a Bayesian framework and it was subsequently coupled with GLUE by Blasone et al. (2008). The general procedure for the application of the SCEM-UA algorithm in a pseudo-Bayesian framework can be subdivided in three steps (see the step-by-step description in Vrugt et al., 2003b):

Step 1:Identification of the likelihood measure (e.g. objective function and acceptance threshold) and the model parameter space defined by prior distributions.

Step 2:Generation of a parameter set sample of dimension n (with n smaller compared to the sample size used in traditional sampling) and application of the optimization algorithm to identify the regions of the parameter space with higher likelihood. The algorithm develops as follows.

(a) Subdivision of initial parameter sample into k complexes.
(b) Evolution of each complex, which evolves independently according to the Sequence Evolution Metropolis (SEM) algorithm. This applies the Metropolis–Hastings method (Metropolis et al., 1953; Hastings, 1970) by accepting/rejecting candidate sets which are generated by jumping in wide regions of the parameter space. This feature avoids the collapse of sequence on local optima.
(c) Substitution of the worst member of the sequence, focussing it in the region of higher likelihood.
(d) Shuffling of the complexes (after a defined number of iterations) to avoid the collapse in local optima. The algorithm starts again from point (b)
(e) Evolution until the convergence criterion (defined by $\sqrt{R}$ criterion of Gelman et al., 1995) is satisfied.

Step 3:Application of the GLUE analysis to the new parameter set sample: rejection of the parameter sets below the acceptance threshold and estimation of the model uncertainty bounds.

The SCEM-UA method can overcome some of the GLUE limitations. For example, the initial range of parameter samples can be wide without necessarily increasing computational requirements. However, the issues surrounding the subjective acceptance criteria are still the same as in GLUE.

2.1.3. Multi-objective calibration using AMALGAM

This approach is based on the further analyses of multi-objective calibration results (Wohling et al., 2008). The parameter distributions (PDs) of the Pareto optimal solutions are estimated within a multi-event calibration run. In comparison to the other methods, each estimated parameter represents one optimal solution for the used objective function.

The AMALGAM method is a multialgorithm, genetically adaptive multi-objective method (Vrugt and Robinson, 2007; Huisman et al., 2010). It uses the four optimization algorithms: (i) non-dominated sorted genetic algorithm-II (NSGA-II, Deb et al., 2002), (ii) particle swarm optimization (PSO, Kennedy et al., 2001), (iii) adaptive Metropolis search (AMS, Haario et al., 2001), and (iv) differential evolution (DE, Storn and Price, 1997). The four algorithms share information through their common population, but methods with higher reproduction success are favoured in the calibration process (Huisman et al., 2010). AMALGAM uses the fast non-dominated sorting algorithm (Deb et al., 2002) and crowding distance for population ranking. The result of an AMALGAM calibration is a set of Pareto optimal solutions of the used objectives.

The AMALGAM algorithm is included in the software tool KALIMOD (Uhl and Henrichs, 2010), which was constructed as an interface between simulation models and optimization algorithms. The general procedure used in this study can be summarized as in the following:

Step 1:Identification/Selection of relevant rainfall–runoff events;
Step 2:Perform multi-event calibration with AMALGAM; and,
Step 3: Estimation of the Pareto optimal solutions.

As GLUE and SCEM-UA, this approach is sensitive to the subjective acceptance criteria. With a smaller threshold, larger uncertainty bands can be obtained. The AMALGAM method is also sensitive to the selection of rainfall–runoff events. The choice of different rainfall–runoff events with varying characteristics concerning rainfall, intensity, peak flow, and volume for multi-event calibration leads to a set of Pareto optimal solutions with different optima and parameter sets. Therefore, rainfall and runoff data have to be analysed carefully (Schmitt et al., 2008).

2.1.4. *Bayesian inference using MICA*

The fundamentals go back to Bayes' theorem (also often called Bayes' law by Bayes, 1763) for calculation of conditional probabilities. In terms of modelling Bayes' theorem can be written as:

$$P(\theta|D) = \frac{P(D|\theta)\cdot P(\theta)}{P(D)} \tag{1}$$

with *P*() as prior distribution of a set of model parameters , *P*(*D*) as distribution of observations (calibration data) and *P*(*D*|) as conditional probability of observing data *D* for a given parameter set (i.e. the likelihood function). Hence, *P*(|*D*) is the probability distribution for the parameter set for given the observed data *D* (often called posterior distribution). Furthermore, *P*(|*D*) is the updated parameter probability after imposing calibration constraints.

Using this formal learning strategy, the posterior distribution *P*(|*D*) can be updated by observing *P*(*D*|) to finally approach the true posterior probability distribution of the parameters. For example, Lindley (2006) suggests that probability is the only way to deal with uncertainty in models. Therefore, the Bayesian inference can be an efficient approach to deal with uncertainties in urban drainage modelling as long as the different components can be defined adequately (Beven, 2009).

In this study, the Bayesian inference was done using the software MICA (Doherty, 2003). It undertakes a MCMC analysis with the Metropolis–Hastings algorithm (Hastings, 1970) by sampling mainly in areas of high likelihood (to be as effective as possible), but also allowing some samples in areas of low likelihood (to have a wide scan of the parameter space). First, the model is run with initial parameter values sampled from a prior distribution. The estimated values are compared with the observed data, and the likelihood of such parameters is calculated. Subsequently, the model is run with a new set of parameters derived from proposal distributions and the likelihood of the new set of parameters is calculated. Bayes' theorem is used to calculate the posterior distribution and parameter sets are accepted or rejected (i.e. similar to the behavioural or non-behavioural parameter sets for non-Bayesian methods). The proposal distributions are updated using the information from previous accepted parameter sets. The acceptance of parameter sets is not influenced by a user-defined threshold, but instead it is implicitly included in the Metropolis–Hastings algorithm and the assumed likelihood function. Usually, this process is repeated until the $\sqrt{R}$ criterion of Gelman et al. (1995) is reached.

The main limitation of this specific MCMC procedure is that the distribution of the model errors has to be known and in MICA it is assumed to be normally distributed. This is a concern for many uncertainty/sensitivity procedures as this assumption is usually not satisfied (e.g. Feyen et al., 2007; Thyer et al., 2007), and therefore, erroneous quantification of the model predictive uncertainty is expected (Beven et al., 2008).

2.2. *Stormwater model – SIMPLE KAREN*

The stormwater model SIMPLE KAREN has been adopted to perform the comparison of the different techniques. The model was selected for its simple structure, as the comparison focuses on the features of the uncertainty analysis techniques rather than on the model behaviour. This software is divided into two different modules: a rainfall/runoff module, and a water quality module. SIMPLE KAREN is a simplified version of the commercial software KAREN (Rauch and Kinzel, 2007), which is commonly used in Austria to evaluate efficiency of combined sewer systems. A brief description of the two modules is provided leaving further details to literature (Kleidorfer, 2010).

The rainfall/runoff module is a simple linear reservoir model (Rauch and Kinzel, 2007). It requires the catchment area and a rainfall time series as inputs to generate a series of flows originating from impervious areas only. The pervious components of the catchments are not considered. The model has four calibration parameters (Table 1): the effective impervious fraction, *EIF* (%); the time of concentration, *TC* (min); initial loss, *li* (mm); and, evapotranspiration, *ev* (mm/d).

The water quality module is a simple concentration regression model which is based on the most common approach used in practice for decades (Francey et al., 2010) and is widely used in several stormwater models, such as SWMM5 (Rossman, 2008) and P8-UCM (Palmstrom and Walker, 1990):

$$C_t = W\cdot R_t^b \tag{2}$$

where: C_t is the pollutant concentration at time *t* (mg/L); *Rt* is measured flow/catchment area (mm/hour over the time-step); and, W and *b* are calibration parameters (Table 1).

Table 1 – SIMPLE KAREN parameters.

Parameter name	Unit	Prior distribution (min–max)
The rainfall runoff module		
Effective impervious fraction (*EIF*)	%	0–100
Time of concentration (*TC*)	min	1–83
Initial loss (*li*)	mm	0–10
Evapotranspiration (*ev*)	mm/day	0–15
The water quality module		
Water quality scale coefficient (W)	$\frac{g\cdot s^b}{m^{3(1-b)}}$	10–300
Water quality shape coefficient (*b*)	–	0–2

For both rainfall/runoff and water quality modules a simulation time-step of six minutes was applied. Although it is known that this simple form of a stormwater quality model is not able to sufficiently predict Total Suspended Solids (TSS) concentrations (Dotto et al., 2010), it was chosen to test the performance of the different uncertainty methodologies on a poor performing model but characterized by limited computational burden.

2.3. *Case study*

The study used the comprehensive stormwater dataset collected in Melbourne, Australia, at Richmond (Francey et al., 2010). This urban catchment is drained by a separate stormwater system and has a total area of 89.20 ha with a total imperviousness of 74%. The dataset consists of two years of continuous rainfall and flow measurements with a temporal resolution of 1 min (the data was then aggregated in a 6 min time-step for modelling). The rainfall totals for wet weather events range from 1.2 to 40.8 mm, and the mean maximum event runoff rate is 547 L/s. The data also includes 12 wet weather events monitored for TSS. The water quality samples were collected at the outlet of the catchments using a discrete sampling methodology, ensuring that pollutographs are monitored throughout each event. For more information on the site and monitoring program see Francey et al. (2010).

The model was applied for calculations of flow rates and TSS concentrations, both at 6 min intervals. Because the models are explicitly independent of each other (i.e. water quality does not rely on the outputs of the water quantity module – see Eq. (2)), the uncertainty analyses were run sequentially: first, the uncertainty in the parameters of the water quantity module was estimated, then the same procedure was followed for the quality module.

2.4. *The comparison of the uncertainty methods*

The tested uncertainty analysis methods outlined above were applied according to their original formulations. However, the following was kept constant for all methods to ensure an unbiased comparison: (i) likelihood measure and acceptance threshold, (ii) range and prior distribution of parameters, and (iii) criterion for the definition of the number of simulations.

The Nash and Sutcliffe (E) efficiency index (Nash and Sutcliffe, 1970) was selected as the likelihood measure for all tested methods. In GLUE, SCEM-UA, and AMALGAM parameter sets with E values smaller than the selected threshold value were discarded and the other parameter sets were accepted as behavioural. As discussed previously, results of the uncertainty analysis are sensitive to this threshold; there is a trade-off between the level of acceptance of behavioural parameter sets to achieve informative results (i.e. representative distributions, uncertainty bands, etc) and the computational cost of the analysis (see Freni et al., 2008). The following threshold values were selected considering this trade-off and the fact that all methods will be compared using the same threshold values (thereby reducing the impacts of this sensitivity):

- Water quantity module: measured and predicted flow rates were compared and parameter sets that achieved $E > 0.6$ were behavioural; and,
- Water quality module: measured and predicted TSS concentrations were compared and parameters sets that achieved $E > 0$ were behavioural.

In the Bayesian approach (MICA), the evaluation of uncertainty is based on appropriate formal likelihood measures (here based on the normality of the residuals between observed and modelled values), and E was only used for comparison of accepted/behavioural simulations with the other three methods.

A uniform prior distribution was considered for all parameters and for all methods. This was because there was insufficient prior information on parameters' behaviour. Indeed, as recently demonstrated by Freni and Mannina (2010b), a uniform prior distribution of model parameters is preferred unless relevant prior parameter information is available. The ranges of values are listed in Table 1.

The number of simulations was selected through a separate process (similar to the one described in Bertrand-Krajewski et al., 2002). For each method, the uncertainty analysis was carried out by means of a variable number of simulations starting from 500 and increasing with steps of 500. At each step, the cumulated parameter distributions were compared to those obtained in the previous step. The number of simulations adopted for each method was reached when the differences between the two posterior distributions were not appreciable according to Kolmogorov–Smirnov maximum distance test with a significance level equal to 0.01. Consequently, this means that both distributions come from the same population and that the sample size is high enough to represent this population.

The comparison of results between the different uncertainty techniques was defined similarly to the criteria adopted in previous studies (Yang et al., 2008; Li et al., 2010):

- Performance of the model, assessed by the Nash and Sutcliffe efficiency criterion;
- The best parameter estimate, their posterior distributions, the estimated parameter uncertainty, and the correlation coefficients between model parameters;
- Model prediction uncertainty, or more precisely the derived 90% probability bands defined as the Average Relative Interval Length ARIL (as proposed by Jin et al., 2010):

$$\mathrm{ARIL} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{Limit}_{\mathrm{Upper},i} - \mathrm{Limit}_{\mathrm{Lower},i}}{\mathrm{X}_{\mathrm{obs},i}} \tag{3}$$

where, $\mathrm{Limit}_{\mathrm{Upper},i}$ and $\mathrm{Limit}_{\mathrm{Lower},i}$ are respectively, the upper and the lower boundary values of the 90% confidence interval; N is the number of measured values; and Xobs, i is the observed ith value.

- Percentage of observations within the prediction bands (i.e. coverage of observations by the probability bands). A combination of a lower ARIL value and a larger percentage of observations within the prediction bands would indicate the better performance of a certain method/model.
- Computational requirements, expressed as the number of required model runs.

3. Results and discussion

For each method, matrix plots were produced for the quantity (Fig. 1) and the quality module (Fig. 2). Each plot provides the parameters' posterior distribution and information on the correlations between model parameters. The figures also show the scatter plots between the behavioural parameter values and the Nash and Sutcliffe efficiency. Each dot represents one behavioural model run. The diagonal histograms represent the density function of the modelling efficiency (in the left upper corner of the figures) and the likelihood densities for all the parameters. The importance of the various model parameters can be seen from these plots (Ratto et al., 2001). Model sensitivity is identified by the clear peaks (for the influential parameters) or flat shapes (for the non-influential parameters) in the parameter PDs. Tables 2 and 3 report the mean, standard deviation, and correlation matrix of the posterior distribution for the water quantity and water quality module parameters, respectively.

The hydrograph and pollutograph for one selected event (recorded on 23th of April 2004) were presented for all methods in Fig. 3 for water quantity, and Fig. 4 for water quality. This event was selected for being an example of a large event observed in the studied catchment. The frequency plots were also created to show the uncertainty bands against the measured data for the whole time series in Fig. 5 for water quantity and Fig. 6 for water quality.

The efficiency of each method is presented in Table 4, which includes information on the maximum Nash and Sutcliffe values, percentage of the measured data points that fall within the prediction bands, the ARIL, and the number of simulations needed to produce reliable results.

3.1. *Rainfall/runoff module*

From Figs. 4 and 5 it is clear that the model is able to predict flow rates. The maximum *E* values were around 0.8 for all methods (Table 4). However, for flows higher than approximately 0.5 m^3/s, the model regularly under predicted, as shown in Fig. 5 where these measured data points are above the bisector line. For flows above 1.5 m^3/s the measured flows fall outside the 95% uncertainty bands. This is most likely due to a change in the system behaviour that cannot be simulated by the SIMPLE KAREN model as pervious catchment areas are not represented (which is most likely to contribute to runoff during large rainfall events).

Fig. 3 and Table 4 both show that all four methods provide similar results in terms of producing prediction bands which cover the observed data (ranging from 45% to 48%). Although not the specific focus of this paper, this low percentage also suggests that the modelling approach should be questioned and that additional sources of uncertainties exist (e.g. model structure, input and calibration data errors, etc.) which are not covered in the analyses (Beven, 2009). The addition of the residual term to these prediction bands to provide the total

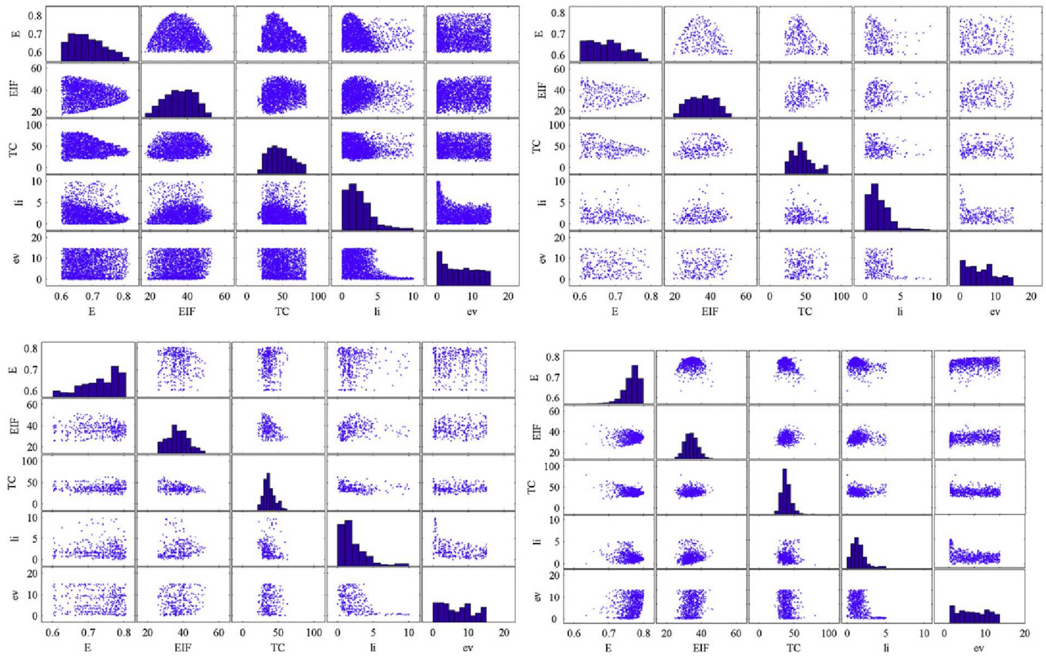


Fig. 1 – Matrix plot of efficiency scatter plots and posterior histograms (efficiency density in the upper left corner, parameters posterior likelihood densities in the other diagonal places) of water quantity module parameters; GLUE (top left), SCEM-UA (top right), AMALGAM (bottom left) and MICA (bottom right).

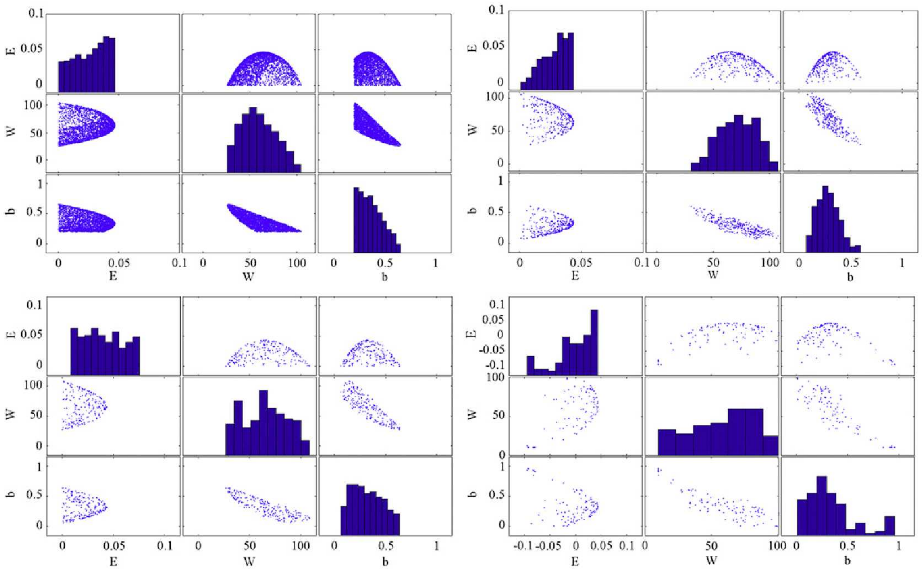


Fig. 2 – Matrix plot of efficiency scatter plots and posterior histograms (efficiency density in the upper left corner, parameters posterior likelihood densities in the other diagonal places) of water quality module parameters; GLUE (top left), SCEM-UA (top right), AMALGAM (bottom left) and MICA (bottom right).

error in the prediction may have changed these results (as described in Feyen et al., 2007). This was not conducted here since we are specifically interested in parameter uncertainties and comparing methods used to develop parameter distributions and associated prediction bands.

Considering the average relative interval length, AMALGAM presented the lowest ARIL (1.17) among the methods (Table 4), while still maintaining the highest percentage of observations within its prediction bands. GLUE presented the highest ARIL value (2.33), suggesting that GLUE

Table 2 – Mean, standard deviation (σ), Coefficient of Variation (CoV) and correlation matrix of the posterior distributions for the flow module parameters.

Method	Parameter	Mean	σ	CoV	Correlation coefficient, R			
					EIF	*TC*	*li*	*ev*
	EIF (%)	35.08	7.86	22%	1			
GLUE	*TC* (min)	47.03	16.01	34%	0.15	1		
	li (mm)	2.47	1.83	74%	0.08	−0.15	1	
	ev (mm/d)	6.59	4.53	69%	0.08	−0.08	−0.14	1
	EIF (%)	35.33	7.43	21%	1			
SCEM-UA	*TC* (min)	45.22	14.4	32%	0.30	1		
	li (mm)	1.99	1.45	73%	0.07	−0.17	1	
	ev (mm/d)	6.17	4.12	67%	0.16	−0.04	−0.26	1
	EIF (%)	37.03	5.68	15%	1			
AMALGAM	*TC* (min)	37.22	7.80	21%	−0.32	1		
	li (mm)	2.19	1.81	83%	0.12	−0.21	1	
	ev (mm/d)	6.72	4.54	68%	0.01	0.04	−0.44	1
	EIF (%)	34.20	3.54	10%	1			
MICA	*TC* (min)	38.51	6.68	17%	0.12	1		
	li (mm)	1.48	0.91	61%	0.10	−0.04	1	
	ev (mm/d)	4.73	2.96	63%	0.23	−0.13	−0.36	1

Table 3 – Mean, standard deviation (σ), coefficient of variation (CoV) and correlation matrix of the posterior distributions for the water quality module parameters.

Method	Parameter	Mean	σ	CoV	Correlation coefficient, R: b	Correlation coefficient, R: W
GLUE	b	0.302	0.41	136%	1	
	W	67.87	21.12	31%	−0.70	1
SCEM-UA	b	0.281	0.108	38%	1	
	W	71.5	16.9	24%	−0.82	1
AMALGAM	b	0.32	0.149	47%	1	
	W	64.92	21.2	33%	−0.86	1
MICA	b	0.33	0.24	73%	1	
	W	67.33	30.60	45%	−0.84	1

usually had higher uncertainty bands with respect to the other methods (yet its coverage percentage was not significantly different).

For all four methods, the effective impervious fraction of the catchment (*EIF*) was found to be the most important calibration parameter, followed by the time of concentration (*TC*) and the initial loss (*li*) (Fig. 1 and Table 2). The model was insensitive to the evapotranspiration (*ev*) parameter with all the tested methods. This is because *ev* only controls the drainage of the initial loss volume during dry weather periods and therefore it does not have a significant influence when the initial loss is only a small fraction of the total event rainfall volume or when the dry weather periods are long (i.e. it does not matter if the initial loss is drained within couple of hours or days). The reduced sensitivity is also compounded by the least squares objective function used here, which focuses on these larger events (see Dotto et al., 2011 for further discussion). The calibrated parameter sets did not differ much. For instance, the EIF parameter did not vary more than 5% between methods (Table 2). However, it appears that GLUE and SCEM-UA tend to flatten the confidence region response surface so that sharp peaks and valleys are not as clear for EIF as they are for the other two methods (Fig. 1). This is also reflected in Table 2, where the standard deviations (and coefficient of variations) for the *EIF* and *TC* parameter PDs were generally larger for GLUE and SCEM-UA. This finding is consistent with previous studies (e.g. Beven et al., 2008 and Yang et al., 2008) which related this behaviour to the equifinality philosophy that is behind GLUE. However, it is also possible that these findings are dependent on the threshold selection (that influences GLUE and SCEM-UA results; Freni et al., 2008), which can influence the shape of the parameter PDs. In the case of MICA, the data used to create the parameter PDs generally had the highest *E* values (usually >0.7, compared with, for example, >0.6 for GLUE; Fig. 1), meaning that PDs from MICA are constructed only with the best parameter sets which subsequently produces more pronounced peaks in the PDs and therefore smaller ARIL values (yet the coverage of observations was similar to that of the other methods).

All four methods found weak (in fact often non-existent) correlation between hydrological parameters (Table 2). However, GLUE showed consistently the lowest level of correlations, possibly a function of the number of data points used in the comparisons. These low correlations found from

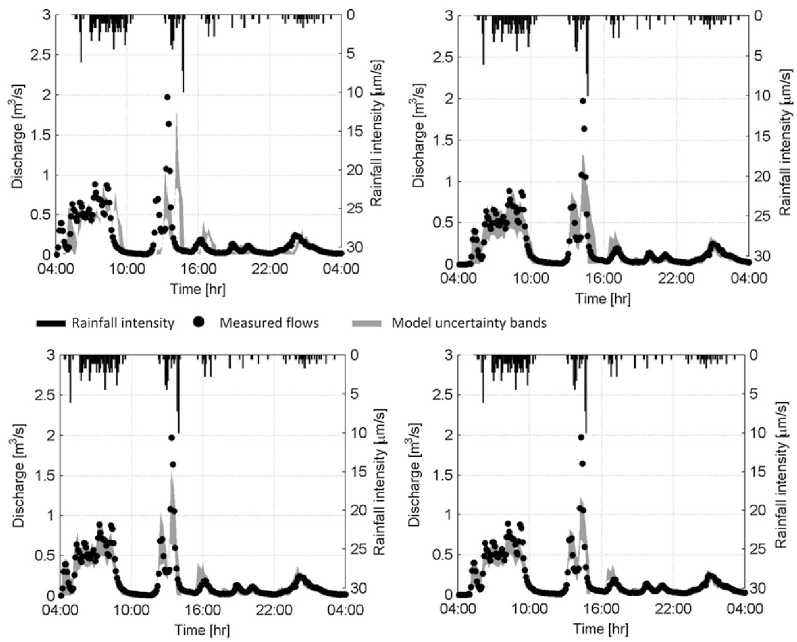


Fig. 3 – Model uncertainty bands for flows during the rainfall event recorded on 2004/04/23; GLUE (top left), SCEM-UA (top right), AMALGAM (bottom left) and MICA (bottom right).

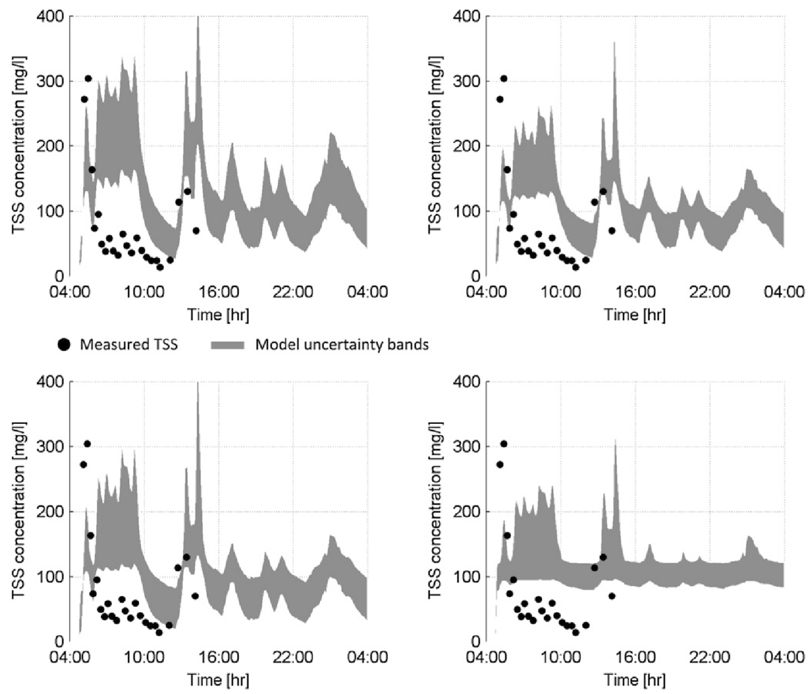


Fig. 4 – Model uncertainty bands for TSS concentrations during the rainfall event recorded on 2004/04/23; GLUE (top left), SCEM-UA (top right), AMALGAM (bottom left) and MICA (bottom right).

the GLUE results confirm the fact that this method tends to flatten the response surface. Such results are in line with previous studies (among others, Beven et al., 2008; Yang et al., 2008). The other three methods had a tendency to show a weak negative correlation between *ev* and *li*, both of which were non-sensitive parameters (R being between −0.26 and −0.44). Other correlations existed, but they were also very weak and never in the same direction across the four methods. As pointed out by Yang et al. (2008), these correlations are likely related to the behavioural parameter sets with significant weight that are quite uniformly dispersed over the parameter space.

3.2. *Water quality module*

All the techniques provided similar optimised model efficiency results, each having a maximum recorded E value of around 0.04 (Table 4). Furthermore, Fig. 6 shows that, for all four methods, the model over-predicts low measured TSS values and heavily under-predicts values for measured TSS above 130 mg/L. The independence of method type indicates that this low efficiency is a model or data error (not a method error) and this has repeatedly been reported in the literature (e.g. Kanso et al., 2003; Dotto et al., 2010). The model structure is simple and is not able to simulate some characteristics which have been recognised to influence TSS behaviour in stormwater (e.g. the first flush effect – Deletic, 1998; variability in concentrations between wet weather events; etc.).

These model limitations are reflected by the low number of observations covered by the prediction bands (Table 4), which ranged from 23% to 29% for the four methods. MICA and AMALGAM provided the narrowest bands (ARIL of 1.57 and 2.21, respectively) and the lowest percentage of observations within these bands, while GLUE and SCEM-UA had the highest ARILs and the highest percent coverage of observations.

The four procedures gave different suggestions about sensitivity to the two parameters W and b (Fig. 2). As compared with the water quantity results, the standard deviations and coefficient of variations of W and b (showing the peakiness of the PDs) are now lowest for SCEM-UA and AMALGAM (Table 3). This could be related to the number of runs needed to create the plots (see number of simulations in Table 4). It is possible that for an ill-posed model, MICA cannot focus around the best parameter sets, thereby limiting the update of the proposal distributions which may imply the process becomes similar to that of GLUE (hence the reasonably similar results seen in Table 3 for GLUE and MICA).

The matrix plots (Fig. 2) also show how some methods are limited in the way they explore the parameter space. For example, it appears that GLUE did not accept any parameter values below a certain threshold for b (e.g. $b < 0.2$) while other methods had relatively high levels of acceptance below the same threshold. This suggests that GLUE may not have explored this area of the parameter space adequately – an interesting finding considering GLUE used a uniform distribution to randomly create parameter sets between $b = 0$ and 2.

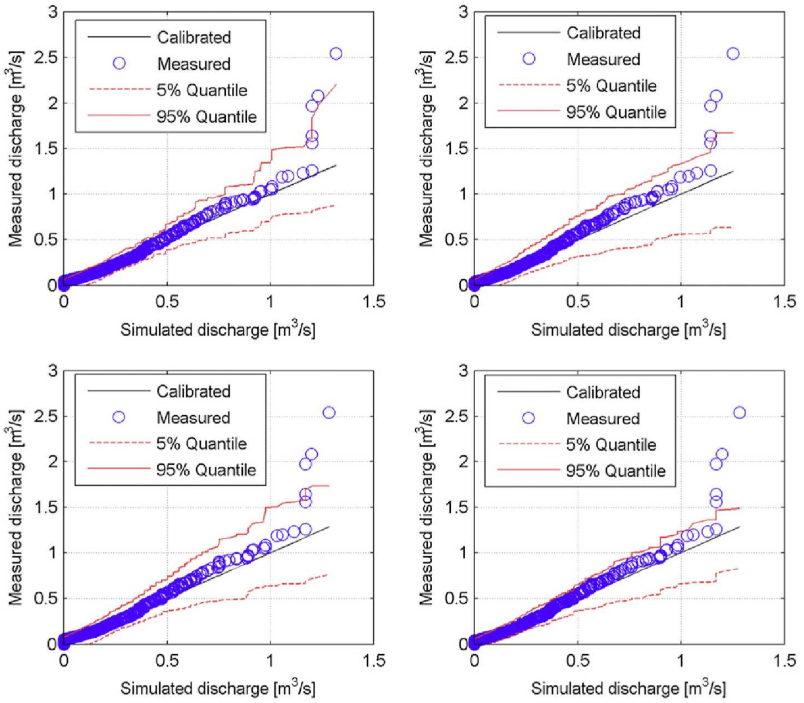


Fig. 5 – Frequency plots of the modelled and measured flows; GLUE (top left), SCEM-UA (top right), AMALGAM (bottom left) and MICA (bottom right).

All four methods found very high levels of cross-correlation between the two water quality model parameters (Tables 3). These results clearly indicate that the model is 'ill-posed' and needs to be reformulated in order to include some relevant processes not accurately simulated (e.g. first flush, sewer sediment erosion and transport, etc.). It may be argued that the model is so ill-posed that all the analyses on the parameter sensitivity are flawed and that the assessment of a single ill-posed model would not have been sufficient to compare uncertainty techniques. However, it was a specific objective of this paper to use one model which adequately represents the data (i.e. water quantity module) and one which cannot accurately represent the measured data (i.e. water quality module) so that comparisons between the four methods can be made on both ends of the spectrum.

3.3. *Computational requirements*

The application of GLUE requires the lowest modelling skill, as it can be simplified in three main steps: (1) generation of parameter set samples, (2) simulation of the model and estimation of the likelihood for each set, and (3) identification of behavioural parameter sets. SCEM-UA increases the complexity of GLUE by using an MCMC optimization algorithm. This is also used by MICA, which requires additional assumptions regarding the structure of the model error. This is not required by AMALGAM, which, although requires a prior analysis of the input data (identification of relevant rainfall events), this process does not add significant computational requirements.

The number of runs required from each method is listed in Table 4; for the water quantity model GLUE and MICA required the highest number of simulations (3500), while the lowest was needed by SCEM-UA (<2350). For the ill-posed water quality model, the same pattern was observed except that GLUE was now a clear outlier and required 30,000 model runs. The methods employing advanced optimization algorithms (SCEM-UA and AMALGAM) require smaller samples to identify water quality parameters than water quantity. This is explained by the higher sensitivity of the two parameters W and b, resulting in a steep response surface that facilitates the evolution of the search algorithms. SCEM-UA is the most computationally efficient method, with a saving in computational time of >6% for the quantity module and >36% for the quality module when compared to the second most efficient methods.

3.4. *User expertise requirements*

A modeller's objective and expertise often vary dramatically; some are only interested in the outputs of a model with a very superficial understanding of the model's methods/processes,

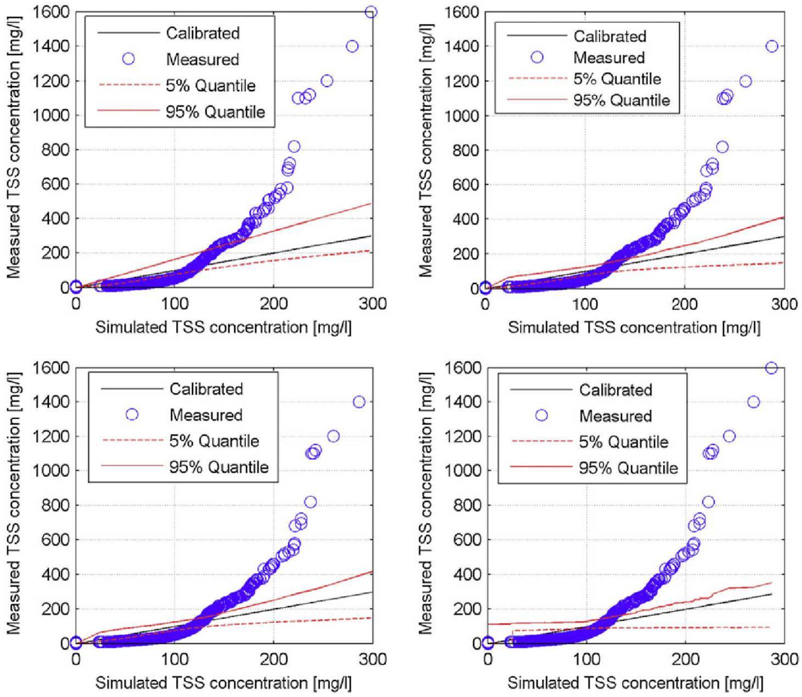


Fig. 6 – Frequency plots of the modelled and measured TSS; GLUE (top left), SCEM-UA (top right), AMALGAM (bottom left) and MICA (bottom right).

while others are experts who fully understand the modelling methods and use the model to, for example, examine interactions of complex processes in the environment. No matter their expertise, at least some knowledge of uncertainty analysis, and a basic understanding of computer programming, is required to interrogate and analyse the output data from these methods and to properly interpret these findings. The following further explains the requirements for implementation of each method.

GLUE is the easiest method to implement, but it can be very computationally demanding and it might be limited in application because of the subjective acceptance threshold (which often requires expert judgement to define). MICA is the most difficult method to implement as it requires fundamental understanding of Bayesian methods. Although it is not limited by the acceptance threshold, it does require the user to have an understanding of the resultant error distribution. GLUE and MICA are both available in free software packages, both with detailed user manuals allowing even a novice programmer to use these tools (see Wagener et al., 2004 and Doherty, 2003 for MICA). Similarly to GLUE, SCEM-UA is easy to implement, but similarly to MICA it does require

Table 4 – The efficiency of the four different methods.

		GLUE	SCEM-UA	AMALGAM	MICA
Flow module	Max E	0.815	0.79	0.805	0.8
	Observations within prediction band (%)	47.5	47	48[a]	45
	ARIL	2.33	1.42	1.17[a]	1.3
	Number of simulations	3500	<2350	3000	3500
TSS module	Max E	0.0436	0.0436	0.0435	0.043
	Observations within prediction bound (%)	28.8	27.23	22.8	23.18
	ARIL	2.91	2.60	2.21	1.567
	Number of simulations	30000	<1600	2500	4500

a Only values with >0.01 m^3/s (Measured data) considered.

Table 5 – Comparison of advantages and disadvantages of methods.

		GLUE	SCEM-UA	AMALGAM	MICA
Flow module	Coverage of observations	+	+	+	+
	Computational requirements	+	++	+	+
TSS Module	Coverage of observations	+	+	+	+
	Computational requirements	−	++	+	−
Ability to identify ill-posed models (correlated parameters, poor efficiency and low coverage)		+	+	+	+
Ability to identify sensitive parameters		+	+	+	+
Availability		Software package	Code for use in MATLAB and OCTAVE	Software package	Software package
Required programming skills		Min	Min	Max	Max
Limitations		Subjectivity of acceptance threshold	Subjectivity of acceptance threshold	Subjectivity of acceptance threshold	Knowledge about error distribution

+ better/equal performance; − worse performance.

background understanding of Bayesian methods (albeit to a lesser extent). Furthermore, SCEM-UA is only available as MATLAB and Octave code, thereby requiring that the user be sufficient in programming. Although AMALGAM requires more skills to use than SCEM-UA (e.g. background on the different optimisation algorithms used for the multialgorithm approach), it is available in a free software package.

3.5. *Advantages and disadvantages of the tested methods*

Table 5 summarizes the performance of the four methods. The results presented in the previous sections demonstrate that the four investigated methods provide similar results in terms of model performances, with similar levels of maximum efficiency for both the water quantity and water quality modules. Differences between the methods were observed for the parameter distributions, although all methods were capable of distinguishing sensitive from non-sensitive parameters. GLUE and SCEM-UA produced the widest/flattest parameter distributions for the rainfall runoff module, which although subsequently produced the largest ARILs did not increase the percentage of observations within the prediction bands. For the water quality module, there was not a direct relationship between the width/peakiness of the parameter distributions (Fig. 2, Table 3) and the ARIL values (this might be related to the ill-posed nature of this model). However, the methods which produced the highest ARIL for the water quality module (GLUE and SCEM-UA) covered a higher proportion of observations in the prediction bands.

All methods also provided useful information about parameter correlation and were able to determine that the water quality module is ill-posed. There were some differences between the methods when investigating parameter cross-correlations, but all methods showed similar results for the most significant correlation (i.e. b and W in the water quality model).

The most significant differences between the four methods were computational effort, required skills/method complexity and reliance on prior expert knowledge. SCEM-UA and AMALGAM were the most computationally efficient methods for both the water quantity and water quality modules. However, AMALGAM provided a lower coverage for the observations of the ill-posed model (water quality). Both SCEM-UA and AMALGAM have the limitation of the subjective acceptance threshold, making it only suitable to those who have sufficient prior expert knowledge in this area.

One of the key problems that GLUE, SCEM-UA and AMALGAM face is in the subjectivity of the acceptance criteria (e.g. there was no objectivity in selecting the threshold value of $E > 0.6$ for rainfall/runoff model). The methods are very sensitive to different threshold values. MICA has an objective way of selecting ‘good’ parameter sets (more user independent), but again the problem arising due to the normality assumption of the residuals presents another challenge.

GLUE and MICA had the highest computational requirements for both the water quantity and quality modules. GLUE provided better coverage for the ill-posed model (water quality) but it did require over six times the number of simulations.

4. Conclusions

A comparison among four different uncertainty techniques was carried out by employing simplified urban stormwater quantity and quality models to an Australian case study. For the well-posed rainfall–runoff model, all the tested techniques provided similar results. The methods generated *similar* posterior parameter distributions and predictions’ uncertainty. The influential parameters were likewise identified and valuable information on parameter interactions was derived. Search algorithms, such as AMALGAM and SCEM-UA, were the most efficient methods in terms of computational requirements.

All the methods highlighted the limitations of the ill-posed water quality model (low maximum efficiency, low coverage of observations, and high parameter interaction). Although the Bayesian approach (MICA), and especially GLUE, required a higher number of simulations, it was difficult to explicitly compare computational requirements because of the model’s ill-posed nature.

The identification of the most appropriate method for uncertainty estimation is a trade-off between the need for a strong theory-based description of uncertainty (but limited by the requirements on prior knowledge – Bayesian approach), simplicity (and subjectivity – GLUE) and computational efficiency (also affected by subjectivity – AMALGAM and SCEM-UA). It is also suggested that different evaluation scenarios should be analysed (i.e. different catchments, models, data, etc).

Based on these conclusions, modellers should select the method which is most suitable for the system they are modelling (e.g. complexity of the model's structure including the number of parameters), their skill/knowledge level, the available information, and the purpose of their study. If a modeller understands the subjectivity of the acceptance threshold and is applying a simple model (like the ones presented here), SCEM-UA might be preferred because it: is most computationally efficient, achieves similar coverage of observations as other methods, is able to determine sensitive parameters, can identify ill-posed model structures and is not complex or difficult to implement. One limitation of SCEM-UA (and of its recent evolution, i.e. Vrugt et al., 2009) is that it is available only in MATLAB or Octave code, but its theoretical framework is provided in literature (Vrugt et al., 2003a), and Vrugt et al. (2003b) provided a step-by-step description of how this method can be coded in any programming language. However, it is stressed that the applicability of each method is also highly case-specific. For example, the Bayesian approach may be more suitable for highly parameterised models, while the other methods would demand a greater number of model runs.

REFERENCES

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Phil. Trans. R. So. Lond. 53 (A), 370–418.

Beck, M.B., 1987. Water quality modelling: a review of the analysis of uncertainty. Water Resour. Res. 23, 1393–1442.

Bertrand-Krajewski, J.L., Barraud, S., Bardin, J.-P., 2002. Uncertainties, performance indicators and decision aid applied to stormwater facilities. Urban Water 4, 163–179.

Beven, K.J., 2009. Environmental Modelling: An Uncertain Future? Routledge, London, UK.

Beven, K.J., Binley, A.M., 1992. The future of distributed models – Model calibration and uncertainty prediction. Hydrol. Process 6, 279–298.

Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? J. Hydrol. 354, 15–32.

Blasone, R.S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., Zyvoloski, G.A., 2008. Generalized Likelihood Uncertainty Estimation (GLUE) using adaptative Markov Chain Monte Carlo sampling. Adv. Water Resour. 31, 630–648.

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242–266.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evolution. Comput. 6, 182–197.

Deletic, A., 1998. The first flush load of urban surface runoff. Water Res. 32, 2462–2470.

Doherty, J., 2003. MICA: Model Independent Markov Chain Monte Carlo Analysis. Watermark Numerical Computing, Brisbane, Australia.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T., Rauch, W., 2010. Stormwater quality models: performance and sensitivity analysis. Water Sci. Technol. 62, 837–843.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., Fletcher, T.D., 2011. Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. Environ. Modell. Softw. 26 (10), 1225–1239.

Feyen, L., Vrugt, J.A., Nualláin, B.Ó, van der Knijff, J., De Roo, A., 2007. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. J. Hydrol. 332, 276–289.

Francey, M., Fletcher, T.D., Deletic, A., Duncan, H., 2010. New insights into water quality of urban stormwater in South Eastern Australia. J. Environ. Eng. ASCE 136, 381–390.

Freer, J., Beven, K., Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resour. Res. 32, 2161–2173.

Freni, G., Mannina, G., 2010a. Uncertainty in water quality modelling: the applicability of variance decomposition approach. J. Hydrol. 394, 324–333.

Freni, G., Mannina, G., 2010b. Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. J. Hydrol. 392, 31–39.

Freni, G., Mannina, G., Viviani, G., 2008. Uncertainty in urban stormwater quality modelling: the effect of acceptability threshold in the GLUE methodology. Water Res. 42, 2061–2072.

Freni, G., Mannina, G., Viviani, G., 2009. Urban runoff modelling uncertainty: comparison among Bayesian and pseudo-Bayesian methods. Environ. Modell. Softw. 24, 1100–1111.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, New York, USA.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7, 223–242.

Harremoës, P., 2003. The role of uncertainty in application of integrated urban water modeling. In: Proceedings of the International IMUG Conference, Tilburg, The Netherlands, 23–25 April 2003.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Huisman, J.A., Rings, J., Vrugt, J.A., Sorg, J., Vereecken, H., 2010. Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion. J. Hydrol. 380, 62–73.

Jin, X., Xu, C.Y., Zhang, Q., Singh, V.P., 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. J. Hydrol. 383 (3–4), 147–155.

Kanso, A., Gromaire, M.C., Gaume, E., Tassin, B., Chebbo, G., 2003. Bayesian approach for the calibration of models: application to an urban stormwater pollution model. Water Sci. Technol. 47, 77–84.

Kennedy, J., Eberhart, R.C., Shi, Y., 2001. Swarm Intelligence. Morgan Kaufmann Publishers, San Francisco, USA.

Kleidorfer, M., 2010. Uncertain Calibration of Urban Drainage Models – A Scientific Approach to Solve Practical Problems. Innsbruck University Press, Innsbruck, Austria.

Kleidorfer, M., Deletic, A., Fletcher, T.D., Rauch, W., 2009. Impact of input data uncertainties on stormwater model parameters. Water Sci. Technol. 60, 1545–1554.

Li, L., Xia, J., Xu, C., Singh, V.P., 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. J. Hydrol. 390, 210–221.

Lindblom, E., Madsen, H., Mikkelsen, P.S., 2007. Comparative uncertainty analysis of copper loads in stormwater systems using GLUE and grey-box modeling. Water Sci. Technol. 56, 11–18.

Lindley, D.V., 2006. Understanding Uncertainty. John Wiley & Sons, Hoboken, New Jersey, USA.

Makowski, D., Wallach, D., Tremblay, M., 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. Agronomie 22 (S1), 191–203.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. J. Hydrol. 10 (3), 282–290.

Palmstrom, N., Walker, W.W.J., 1990. The P8 Urban Catchment Model for Evaluating NonpointSource Controls at the Local Level. Enhancing States' Lake Management Programs, US EPA.

Ratto, M., Tarantola, S., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. Comput. Phys. Commun. 136, 212–224.

Rauch, W., Harremoës, P., 1999. On the potential of genetic algorithms in urban drainage modeling. Urban Water 1, 79–89.

Rauch, W., Kinzel, H., 2007. KAREN – User Manual. hydro-IT GmbH, Innsbruck.

Romanowicz, R., Higson, H., Teasdale, I., 2000. Bayesian uncertainty estimation methodology applied to air pollution modelling. Environmetrics 11, 351–371.

Rossman, L.A., 2008. Storm Water Management Model – User's Manual Version 5.0. National Risk Management Research Laboratory – USEPA, Cincinnati, Ohio, USA.

Schmitt, G., Becker, M., Flores, C., Pfeiffer, E., Sitzmann, D., Uhl, M., 2008. Model calibration for quality assurance in urban drainage modeling. KA 55 (12), 1306–1313 (in German).

Storn, R., Price, K., 1997. Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11, 341–359.

Thyer, M.A., Renard, B., Kavetski, D.N., Kuczera, G.A., Srikanthan, S., 2007. Bayesian total error analysis for hydrological models: preliminary evaluation using multi-site catchment rainfall data. In: Proceeding of MODSIM07, International Congress on Modelling and Simulation, Christchurch, New Zealand, December 2007.

Uhl, M., Henrichs, M., 2010. KALIMOD: Calibration of Hydrological Process Models, Manual Version 2.0. Muenster University of Applied Sciences, Laboratory of Water Resources Management, Germany (in German).

Vrugt, J.A., Braak, C.J.F.t., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int. J. Nonlin. Sci. Num. 10 (3), 271–288.

Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. Proc. Natl. Acad. Sci. U S A 104, 708–711.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003a. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour. Res. 39 (8). SWC1–1SWC1-16.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003b. Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resour. Res. 39 (8). SWC5–1SWC5-19.

Wagener, T., Wheater, H.S., Gupta, H.V., 2004. Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London, UK.

Willems, P., 2008. Quantification and relative comparison of different types of uncertainties in sewer water quality modeling. Water Res. 42, 3539–3551.

Wohling, T., Vrugt, J.A., Barkle, G.F., 2008. Comparison of three multiobjective optimization algorithms for inverse modeling of vadose zone hydraulic properties. Soil Sci. Soc. Am. J 72 (2), 305–319.

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. J. Hydrol. 358, 1–23.

4.3 Preliminary application of uncertainty method for stormwater flow and quality modelling

4.3.1 Errata

It should be noted that Figure 6 in Section 4.3.2 is not completely accurate. It presents an incorrect positive value for one of the Nash and Sutcliffe coefficients (E) in the vertical *y* axis of the validation figure (on the right) when 6 months were used for model calibration. The correct value is -1.48, instead of the erroneous +1.48 as presented in the figure. The correct figure is presented below.

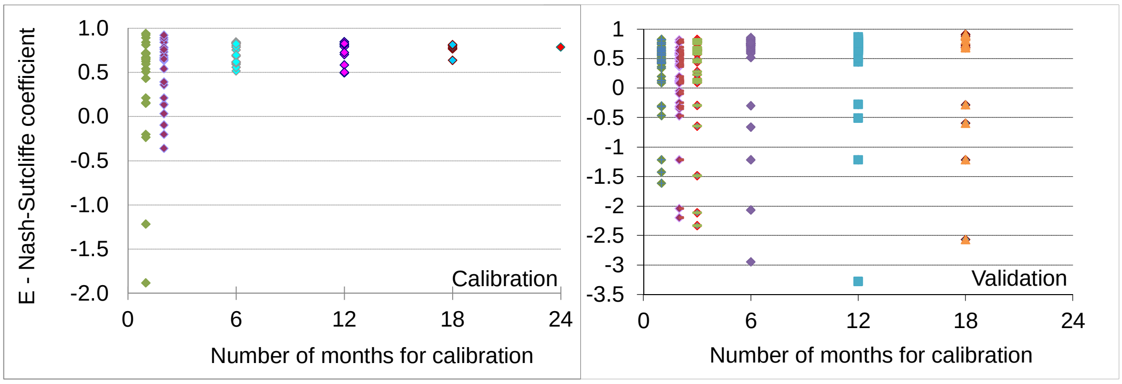


Figure 4.1 Figure 6. The effect of different length of data period on calibration and validation.

717 © IWA Publishing 2009 Water Science & Technology—WST | 60.3 | 2009

Analysis of parameter uncertainty of a flow and quality stormwater model

C. B. S. Dotto, A. Deletic and T. D. Fletcher

ABSTRACT

Uncertainty is intrinsic to all monitoring programs and all models. It cannot realistically be eliminated, but it is necessary to understand the sources of uncertainty, and their consequences on models and decisions. The aim of this paper is to evaluate uncertainty in a flow and water quality stormwater model, due to the model parameters and the availability of data for calibration and validation of the flow model. The MUSIC model, widely used in Australian stormwater practice, has been investigated. Frequentist and Bayesian methods were used for calibration and sensitivity analysis, respectively. It was found that out of 13 calibration parameters of the rainfall/runoff model, only two matter (the model results were not sensitive to the other 11). This suggests that the model can be simplified without losing its accuracy. The evaluation of the water quality models proved to be much more difficult. For the specific catchment and model tested, we argue that for rainfall/runoff, 6 months of data for calibration and 6 months of data for validation are required to produce reliable predictions. Further work is needed to make similar recommendations for modelling water quality.

Key words | calibration, MUSIC, parameter sensitivity, rainfall/runoff, stormwater model, validation, water quality

C. B. S. Dotto (corresponding author)
A. Deletic
T. D. Fletcher
Department of Civil Engineering and eWater CRC,
Institute for Sustainable Water Resources,
Monash University,
Victoria 3800,
Australia
E-mail: [redacted]

INTRODUCTION

The poor quality and altered flow regime of urban stormwater are major threats to aquatic ecosystems (U.S. Environmental Protection Agency 2000; Commonwealth of Australia 2002). The cumulative effect of increased frequency, volume, and rate of stormwater runoff, together with elevated pollution concentrations and loads result in the accelerated degradation of receiving water bodies. Restoring pre-development stormwater flow and quality is a driving goal of stormwater management practices in urban areas. Incorrect estimates of stormwater flows and pollution concentrations can easily lead to an inadequate design of stormwater management systems (Vaze & Chiew 2003). Thus, robust modelling of stormwater discharges and their associated pollutants is critical. Runoff generation and flow routing models are now well developed and widely adopted. However, stormwater quality models are less well developed. While, successful event based pollutant loads modelling have been published in the literature (e.g. Rodríguez *et al.* accepted). However, continuous modelling, in which one most probable set of parameters is found, still remains a challenge.

Uncertainty is intrinsic to all monitoring programs and all models. It cannot realistically be eliminated, but attempts can be made to minimise it. For effective management of uncertainties it is necessary to understand their sources and consequences. Considerable attention has been given to the development of global calibration procedures that estimate a best set of parameters, whereas less attention has been given to the assessment of the significance of the optimal set of parameters, and to the quantification of models' uncertainty.

Improving models and their effectiveness requires the use of robust methodologies for model calibration and

doi: 10.2166/wst.2009.434

validation. Such methods should be able to provide not only an assessment of the uncertainties in the model's parameter values, but also an evaluation of the confidence level of the model's predictions (Kanso *et al.* 2005). High levels of parameter uncertainty can result from a poor model structure, including a high level of parameter correlation, and insensitivity on the part of certain parameters.

The main approaches used for calibration and quantification of the uncertainty in the estimated parameter values are the frequentist and Bayesian (Ferguson 1973) methods. They are fully explained in Gallagher & Doherty (2007). The Metropolis algorithm (Metropolis *et al.* 1953), a general Monte Carlo Markov Chain (MCMC) sampling method, has been widely used for model calibration and sensitivity analysis (Kuczera & Parent 1998; Kanso *et al.* 2003). For example, Kanso *et al.* (2003) carried out a successful application of the Metropolis algorithm for parameter sensitivity analyses of lumped stormwater quality models. Contrary to frequentist approaches, the Metropolis algorithm identifies not only a best parameter set, but a probability distribution of parameters according to measured data; it estimates the true posterior probability distribution of parameters, which may differ significantly from the multinormal distributions used in classical parameter uncertainty estimation methods.

Selection of an appropriate objective function is a fundamental consideration when estimating model parameters. Previous studies (Diskin & Simon 1977; Servat & Dezetter 1991) demonstrated that objectives functions based on the sum of the squared deviations (e.g. sum of square errors, Nash and Sutcliffe coefficient (Nash & Sutcliffe 1970) produce the best results for the majority of applications.

Within this broad issue of model uncertainty there is a specific problem of the relationship between model uncertainty and the data availability for calibration and validation. While some work has been now published on calibration parameter uncertainties in stormwater models (Beven & Binley 1992; Gaume *et al.* 1998; Kanso *et al.* 2005) very little is known about the impacts of calibration and validation data sets on model performance.

The aim of this work is to increase our understanding of the uncertainties of the parameters in models that are currently being used for assessment of stormwater quantity and quality. We focus on the most commonly used stormwater model in Australia, the Model for Urban Stormwater Improvement Conceptualization, MUSIC, (CRCCH 2005). Although the model has been widely used by the industry and researchers, no work has been done so far on its uncertainties. This paper presents results of the uncertainty analysis in MUSIC due to the model parameters for the rainfall/runoff and pollution generation modules, and to the availability of data for calibration and validation for the rainfall/runoff module. In order to achieve the desired parameter evaluation two different approaches for model calibration were tested.

METHODS

The data

Urban stormwater quality and quantity data of two urban catchments, with different sizes and land use, were used. These catchments are located in the eastern suburbs of Melbourne, Australia. The data set contains around 200 events for which a range of pollutants was monitored. Table 1 summarizes the catchment characteristics and Figure 1 illustrates the sites and their respective rainfall and flow gauges. Complete information about the catchments characteristics is available in (Francey *et al.* in press).

For both catchments, flows below certain depth could not be measured; therefore a flow threshold was set, in which the flows below the value were regarded as non reliable. Data were thoroughly checked and only reliable data were included.

The model

MUSIC was developed for Australian conditions (CRCCH 2005). It predicts stormwater flows from urban catchments with separate stormwater and sanitary sewers, along with concentrations of key pollutants (Total Suspended Solids (TSS), Total Phosphorus (TP) and Total Nitrogen (TN)), as well as the performance of specific stormwater treatment measures. The rainfall-runoff algorithm is based on the SimHyd model developed by Chiew & McMahon (1997).

Table 1 | Summary of site details (Francey *et al.*, in press)

Characteristics	Madden Grove, Richmond	Ruffeys Wetland, Doncaster
Catchment size (ha)	89.2	105.7
Total fraction impervious	0.74	0.51
Land use	High-density residential	Medium-density residential
Catchment average slope (%)	3.5	5.0
Distance to rainfall gauge (from catchment centroid)	600 m	700 m
Mean annual rainfall	650 mm per year	650 mm per year
Peak flow (min/max)	65/2,850 L/s	140/4,600 L/s
Event duration (min/max)	2/25 hours	0.5/36 hours
Flow threshold	3 L/s	30 L/s
Parameters sampled	TSS, TP, TN	TSS

The original daily model was modified to enable disaggregation of daily runoff into sub-daily temporal patterns (Figure 2). Flow from impervious and pervious areas are modelled separately, with impervious area runoff being primarily as a function of the proportion of catchment imperviousness, with an initial loss term. Runoff from the pervious areas will only occur for large or intense storm events, when the pervious soil storage is saturated, and depends on soil properties, provided as a series of calibration parameters in the model. The model has 13 calibration parameters, listed in Figure 2.

Whilst MUSIC currently simulates pollution generation using a stochastic approach with dry weather and wet weather event mean (and standard deviation) concentrations, uncertainty analysis on such a model is not really feasible, given its stochastic nature. However, there is also a trial module, Deterministic Pollution Loads Generation (DPLG), that estimates the loads within a timestep as a power function of the rainfall intensity (Sartor & Boyd 1972):

$$L_t = aI_t^b$$

where L_t is the pollutant load at time t (Kg/timestep); I_t is the rainfall intensity (average, in mm/hr, over the timestep, t); a and b are parameters to be calibrated. Once the load has been calculated, the DPLG back-calculates concentration. Uncertainty analysis was undertaken on this

Figure 1 | Aerial photos of the catchments with the rainfall and flow gauges.

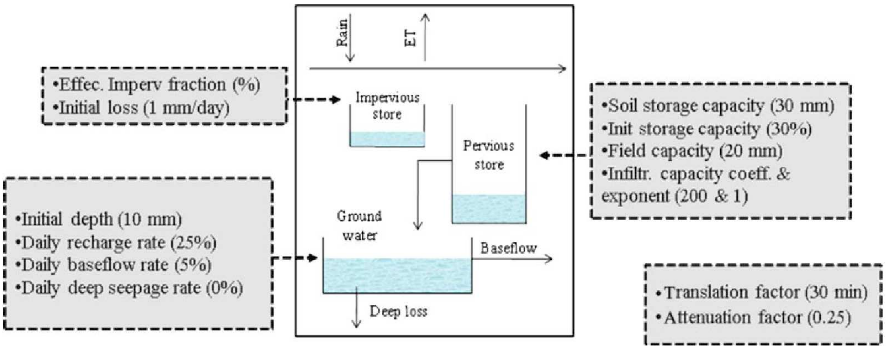


Figure 2 | MUSIC rainfall/runoff parameters and their default values for Melbourne between brackets (after CRCCH 2005).

Deterministic Pollution Loads Generation module, along with the rainfall-runoff module.

Calibration and sensitivity analysis

Parameters in urban drainage models can be highly correlated (commonly the case for water quality models, e.g. Dotto *et al.* (2009)), therefore it essential to perform a global sensitivity of parameters where all parameters are varied simultaneously. This study included two approaches for model calibration and sensitivity analysis. The frequentist and Bayesian approaches were adopted, applying two different tools, as follows.

Bayesian

The search of the probability density functions of the parameters was undertaken with the software MICA (Doherty 2003*a*,*b*). MICA undertakes a simplified MCMC analysis and uses a particularly flexible version of the Metropolis-Hastings algorithm (Hastings 1970). A uniform prior distribution was assumed for all parameters. The lower and upper limits of these distributions are established according to values reported in the literature and/or any previous knowledge about the parameters. Subsequently repeated model (rainfall/runoff and DPLG) runs using the obtained parameter samples were accomplished. In this manner, several Markov Chains were run in parallel, with 10,000 iterations each. The first 1,000 iterations were removed, assuming to be the "burn in" period (Berliner 1994).

Frequentist

The rainfall/runoff model automatic calibration and initial sensitivity analysis was undertaken using the optimization tool PEST (Doherty 2004). The problem of the objective function surface in parameter space being pitted with local minima was overcome using one of the PEST applications (PD_MS2, (Doherty 2003*a*,*b*). In this application, calibration runs are started from different points in the parameter space, which are selected in a manner that minimizes the chance of finding the same local minimum twice. PEST results include optimized parameter values, their 95% percent confidence limits, the sum of squared weighted residuals, the parameter correlation coefficient matrix and the simulated values with the calibrated parameters. PEST was initially applied by allowing all 13 parameters of rainfall/runoff model to float within given ranges. However this process failed and the calibration was repeated, once the most sensitive parameters were determined from the Bayesian method (see above); PEST in conjunction of PD_MS2 was run by floating only the sensitive parameters (EIA and *K*, Figure 1), and fixing the less important ones. This way both approaches could be assessed and compared in terms of efficiency and runtime.

The misfit between observed and modelled values was assessed with the Nash and Sutcliffe coefficient (*E*) for both approaches. While both MICA and PEST were used to calibrate and evaluate rainfall/runoff model, for both catchments, only MICA was used to evaluate the two parameters of DPLG model for the Richmond site, due to problems discussed below.

Impact of calibration and validation data sets

The last step was to assess the impact of data used in calibration and validation of the runoff model for only Richmond catchment. This means that the 'success' of calibration using PEST (and procedure explained above) and validation were assessed, for different lengths of calibration and validation data. Different tests are available for splitting the data for calibration and validation (Xu 1999). The split-sample test consists of dividing the data set in two and use half for calibration and half for validation. Some other tests (e.g. differential split-sample test) divide the data set according to rainfall rates or some other variable in order to demonstrate the ability of the model to predict general conditions. Richmond has two years of continuous rainfall and flow data available: with an average of 65 mm per month, the months between June and November characterized the wet period, while the remaining months had and average of 45 mm per month and were selected as the dry period In this study, the calibration was done using records of 1, 2, 3, … 24 months long. At the same time, validation was performed with the remaining 23, 22, 21, … 1 month long records. A differential split-sample test was naturally generated once in several occasions the model was calibrated with one extreme condition (wet or dry months) and validated against the opposite extreme condition or even against periods in which both wet and dry months were included. The calibration objective function (i.e. the Nash–Sutcliffe Coefficient) was plotted against the calibration/validation length of record.

RESULTS AND DISCUSSION

Calibration and sensitivity analysis

Rainfall/runoff modelling

Table 2 summarizes the results from Richmond and Ruffeys calibration with the two approaches. The values at the global minimum and the approximations to the 95% confidence intervals (CI) resulting from PEST are presented, along with the mean (μ) values and the standard deviation (σ) obtained with MICA runs.

Table 2 | Summary of PEST and MICA runs for flow modelling

	Frequentist (PEST)		Bayesian (MICA)	
	Value	95% CI	μ	σ
Madden Grove, Richmond				
EIA	0.32	0.318–0.326	0.30	0.036
K	20.51	20.30–20.72	20.64	2.68
E	0.80		0.77	
Ruffeys wetland, Doncaster				
EIA	0.246	0.237–0.255	0.26	0.005
Thres	0.285	0.122–0.449	0.51	0.057
S1Max	117.70	117.17–118.23	106.55	2.344
SInit	37.48	−2209.20–2284.16*	31.80	3.985
Fc	100.0	−451.387–651.387*	50.05	10.434
Coeff	100.0	89.4173–0.583	183.36	17.490
Sq	1.083	0.920–1.246	1.04	0.016
Rfac	0.028	−0.155–0.211*	0.84	0.055
Bfac	1.0	−62.512–64.512*	0.69	0.045
K	13.80	13.703–13.8973	8.41	2.514
E	0.40		0.50	

*Out of the upper and/or lower bounds.

Figure 3 presents the histograms for selected rainfall/runoff model parameters (see Figure 2 for their definition). Such histograms were generated according to the parameters mean values generated along with the MICA runs. The model is very sensitive to effective impervious area, EIA, for both catchments. For Richmond, the model was also sensitive to K (the Muskingum Cunge translation factor), but insensitive to the other 11 parameters. By contrast, the rainfall/runoff model for Ruffeys catchment appeared to be very sensitive also to the parameters related to the pervious area runoff (e.g. S1Max). This was expected because of the high proportion of pervious areas. It should be noted that some of the parameter distributions obtained with the MICA were non-normal (e.g. K), contrary to the hypothesis of the classical parameter uncertainty estimation methods (see Kanso *et al.* 2003 for further details).

Using MICA, the rainfall/runoff model calibrates well for Richmond, with the best E values being 0.77 (Table 2). To illustrate the model accuracy, the 5 and 95% prediction

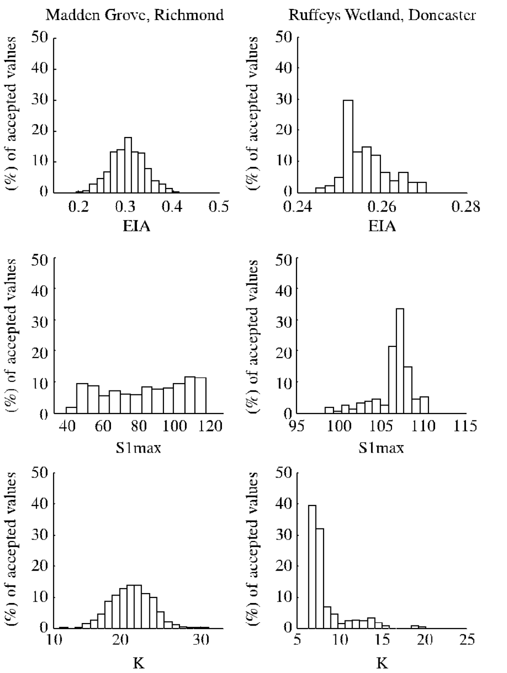


Figure 3 | Sample histograms for selected model parameters based on MICA runs with a Nash–Sutcliffe Coefficient (E) higher than 0.5.

limits due to the uncertainty in rainfall/runoff parameters EIA and *K*, are presented in Figure 4, for a subset of the simulation period. For the continuous simulation during 2005, the *E* between the predicted and observed values was 0.76, 0.81 and 0.69 for the mean, lower and upper 95% confidence limits, respectively. For Ruffeys, *E* was only 0.50 using MICA (Table 2). It is hypothesised that this rather low value may have surged from some periods of null-rainfall

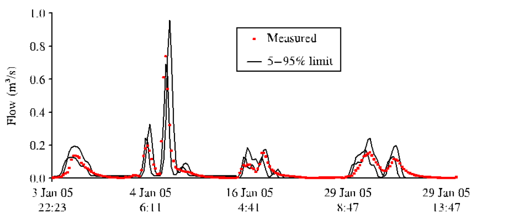


Figure 4 | The 5 and 95% prediction limits due to the uncertainty in rainfall/runoff parameters EIA and K for a period of rainfall for continuous simulation at Richmond.

and flows in the data set, as well as the complexity of the processes in this catchment with high levels of impervious areas, and potential extraneous wastewater inputs.

Calibration of the model's most important parameters (EIA and *K*) using PEST resulted in an *E* of 0.80 for Richmond (even when the other 11 parameters were fixed). The same procedure for Ruffeys Wetland catchment resulted in *E* values of 0.40 (Table 2).

The model calibrated well for Richmond, and results from both approaches were similar. However for Ruffeys, the model did not calibrate very well, with MICA being slightly more efficient. This is not surprising, since in the PEST calibration exercise, pervious surface parameters were kept constant, while they were floated in MICA. The high correlation between some of the parameters further complicated the calibration process. For example, the infiltration capacity coefficient (Coeff) and its exponent (sq) are highly correlated ($R^2 = 0.93$). Most importantly, it was found that MICA with MCMC methods can be applied successfully even to models with a large number of parameters. This approach has the advantage that it shows the probability structure of parameter space, while PEST only provides information on a single set of values. On the other hand, when dealing with a small number of non-correlated parameters, PEST can achieve the optimal set with far fewer model runs than MICA, minimizing the computational time. In the specific case of MUSIC's rainfall/runoff model (a rather complex model with 13 parameters), it is suggested that the Bayesian approach is the most appropriate for calibration and sensitivity analysis.

Water quality modelling

Results are only presented for the DPLG's application to Richmond (Table 3 and Figure 5).

Contrary to results for the rainfall/runoff model, the DPLG module did not calibrate well using the MICA approach. The DPLG for TSS seemed to be very sensitive to both *a* and *b*, and the global minimum is clearly defined by

Table 3 | Summary of results obtained with MICA runs for water quality modelling

Madden Grove, Richmond	TSS	TP	TN
E	0.24	− 17.76	− 2.9

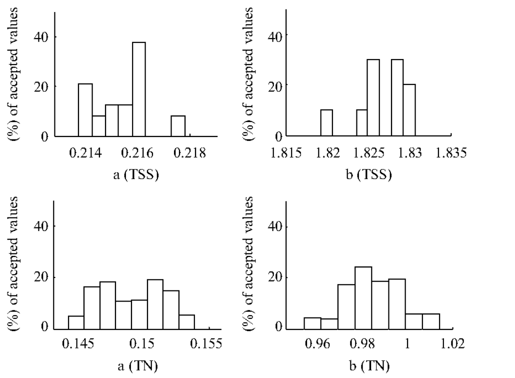


Figure 5 | Histograms for DPLG parameters for TSS and TN, based on the MICA runs.

the peaks in the histograms (Figure 5). In this case, the non-normal behaviour of parameters may be possible explained by the presence of multiple local minimum in the histograms. However, the *E* obtained with these parameters mean values was low (Table 3). Despite the long time consumed to run the chains with several iterations, MICA runs for TP and TN did not produce satisfactory results. It is possible that the limits assumed to the uniform prior distributions were too wide. Parameters *a* and *b* are highly correlated and different combination of these can lead to the same results (although they are not always realistic values); this is believed to be the principal problem. Further work is being carried to better investigate the global minimum for DPLG calibration and parameters' assessment, however at this stage it may be speculated that the DPLG is an 'ill-posed model'.

Further work on the choice of different objective functions approaches, in which the analysis of the predictive uncertainty and its consistency with the observed data is considered, would be able to improve the model calibration and performance (Thyer *et al.* 2009).

Impact of calibration and validation data sets

Calibration and validation with different length data sets were carried out for Richmond (Figure 6). The efficiency of calibration is highly related to the length of the calibration data period. The *E* value varied from −1.88 to 0.94 when calibration was undertaken with one month data. This variation decreases considerably when six months are used for calibration; *E* ranged from 0.51 to 0.84 with an average of 0.73. It is possible to obtain reasonable values of *E* by calibrating with a larger number of months, although the improvement over calibration conducted with 6 months of data is not great (calibration with the whole dataset, 24 months resulted in an *E* of 0.80, Table 2). However, it is not just the length of data that is important; the nature of data within the chosen subset is also critical. For example, calibration resulted in better values of *E* when the wet months were included (eg. June, July and August) due to the larger number of rainfall events with higher volumes.

Figure 6(b) shows that validation results are dependent on the size of calibration dataset. For instance, when 2 months were used to calibrate, validation was undertaken with 1 up to 22 months. Results showed that although reasonable values of *E* may be achieved in some cases, the majority of cases produced low *E* values, with an average of

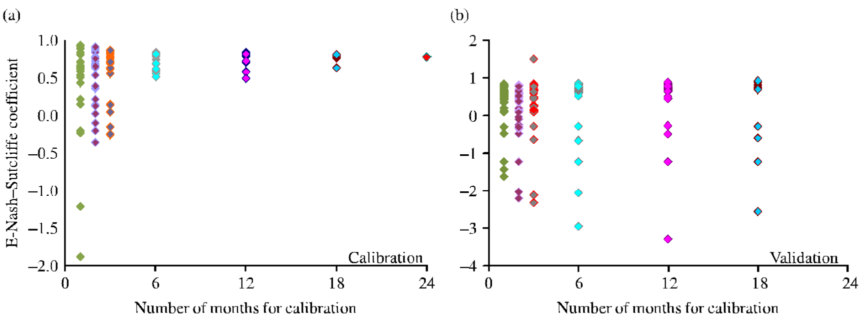


Figure 6 | The effect of different length of data period on calibration and validation.

only 0.18. E values for validation are on average 0.42, when 6 months data sets were used for calibration and between 1 and 19 months for validation. It is clearly not advisable to use a single month to validate the model, given the variations in storm events from month to month.

We conclude that for this catchment, MUSIC rainfall/runoff model, and Melbourne climate, 6 months of data is needed for calibration and a further 6 months of data for validation. Further work is needed to identify the data lengths required for the DPLG water quality module.

CONCLUSIONS

This study investigated uncertainty in a flow and water quality stormwater model due to the model parameters and the availability of data for calibration and validation of the flow model. Frequentist and Bayesian methods were tested for calibration and sensitivity analysis for the rainfall/runoff model used in MUSIC. It was clear that this model, with its 13 calibration parameters, is not sensitive to its pervious area parameters when applied to urbanised catchments. Therefore, it was suggested that only Effective Impervious Area (EIA) and routing parameter K need to be calibrated, with others fixed as defaults. The same does not apply when the level of urbanization and impervious area fraction is lower; in this case some of the pervious area parameters (e.g. S1Max and Fc) also become very important. However, even including them into the calibration process may not result in good model performance.

To better understand the performance of stormwater models and the sensitivity of these models to their parameters, a comparison between models with different levels of complexities will be carried out. In which, the same datasets and calibration/sensitivity analysis approach will be used.

Considering the model and the catchments dataset limitations, the rainfall/runoff model was satisfactory calibrated to both catchments when the Bayesian approach was applied, while the frequentist approach experienced some problems. Therefore it is suggested the Bayesian approach is more appropriate for calibration and sensitivity analysis of complex stormwater flow models. However, even this approach was not able to deliver satisfactory calibration of the stormwater quality DPLG model. Although the model is a simple regression, its parameters are highly correlated and the initial condition boundaries need to be further explored. It is not clear whether the model structure (perhaps over-simplified) or the MICA method is failing. Further analyses are underway to answer this question.

For the rainfall-runoff model applied to the Richmond case-study catchment, we argue that 6 months of data for each of calibration and validation are required. In a more general context it is suggested to divide the data in order to obtain a balance between calibration and validation efficiency. Further work is necessary to evaluate the impact of data on calibration and validation in the case of water quality models.

ACKNOWLEDGEMENTS

The authors acknowledge eWater CRC and CAPES for the scholarship provided.

REFERENCES

Berliner, S. N. M. L. M. 1994 Subsampling the Gibbs Sampler. *Am. Stat.* **48**(3), 188–190.

Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process* **6**(3), 279–298.

Chiew, F. H. S. & McMahon, T. A. 1997 Modelling daily runoff and pollutant load from urban catchments. *Water (J. Aust. Water Wastewater Assoc.)*, **24**, 16–17.

Commonwealth of Australia 2002 The value of water: Inquiry into Australia's urban water management. Report of the Senate Environment, Communications, Information Technology and the Arts Reference Committee. Canberra, Australia: The Parliament of the Commonwealth of Australia.

CRCCH 2005 MUSIC Version3 – User guide, Cooperative Research Centre for Catchment Hydrology.

Diskin, M. H. & Simon, E. 1977 A procedure for the selection of objective functions for hydrologic simulation models. *J. Hydrol.* **34**(1–2), 129–149.

Doherty, J. 2003*a* MICA – Model-Independent Markov Chain Monte Carlo Analysis – User Manual, Watermark Numerical Computing – U.S. EPA.

Doherty, J. 2003*b* PEST – Surface Water Utilities, Watermark Numerical Computing and University of Idaho.

Doherty, J. 2004 PEST – Model-Independent Parameter Estimation Version 10 – User Manual, Watermark Numerical Computing.

Dotto, C. B. S., Deletic, A., Fletcher, T. D. & McCarthy, D. T. 2009 Parameter sensitivity analysis of stormwater models. The 6th International Water Sensitive Urban Design Conference and 3rd International Hydropolis. Perth, Western Australia.

Ferguson, T. S. 1973 A bayesian anaysis of some nonparamteric problems. *Ann. Stat.* **1**(2), 209–230.

Francey, M., Fletcher, T. D., Deletic, A. & Duncan, H. in press. New insights into water quality of urban stormwater in South Eastern Australia. *J. Environ. Eng. – ASCE*.

Gallagher, M. & Doherty, J. 2007 Parameter estimation and uncertainty analysis for a watershed model. *Environ. Model. Softw.* **22**(7), 1000–1020.

Gaume, E., Villeneuve, J. & Desbordes, M. 1998 Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. *J. Hydrol.* **210**, 38–50.

Hastings, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B. & Chebbo, G. 2003 Bayesian approach for the calibration of models: application to an urban stormwater pollution model. *Water Sci. Technol.* **47**(4).

Kanso, A., Chebbo, G. & Tassin, B. 2005 Stormwater quality modelling in combined sewers: calibration and uncertainties analysis. *Water Sci. Technol.* **52**(3), 63–71.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* **211**(1–4), 69–85.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, A. H. 1953 Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**(6), 1087–1092.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *J. Hydrol.* **10**(3), 282–290.

Rodríguez, J. P., Achleitner, S., Möderl, M., Rauch, W., Maksimović, Č., McIntyre, N. & Rodríguez, M. A. D.-G. a. M. S. accepted. Sediment and pollutant load modelling using an integrated urban drainage modelling toolbox: An application of City Drain. 8th International Conference on Urban Drainage Modelling. Tokyo, Japan.

Sartor, J. D. & Boyd, G. B. 1972 Water pollution aspects of street surface contaminants. Rep. EPA-R2-72-081. Washington, DC., U.S. Environmental Protection Agency.

Servat, E. & Dezetter, A. 1991 Selection of calibration objective functions in the context of rainfall-runoff modelling in a sudanese savannah area. *Hydrol. Sci. J.* **36**(4), 307–330.

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

U.S. Environmental Protection Agency 2000 National Water Quality Inventory: 1998 Report to Congress (www.epa.gov/305b/98report). New York: U.S. Environmental Protection Agency.

Vaze, J. & Chiew, F. H. S. 2003 Comparative evaluation of urban storm water quality models. *Water Resour. Res.* **39**(10), 1280.

Xu, C.-y. 1999 Operational testing of a water balance model for predicting climate change impacts. *Agric. For. Meteorol.* **98–99**, 295–304.

4.4 Conclusions

This chapter presented two studies which demonstrated the application and comparison of different methods for the calibration, sensitivity and uncertainty analysis of stormwater quantity and quality models.

Firstly, four different sensitivity analysis techniques (GLUE, SCEM-UA, AMALGAM and MICA) were compared in terms of model performance, predictive uncertainty results and computational time among other criteria. They were used to evaluate a simple conceptual model rainfall runoff model with four calibration parameters and a simple regression pollution generation approach (two calibration parameters). The four different uncertainty analysis methods generated *similar* posterior parameter distributions and predictive uncertainty. The influential parameters were likewise identified and valuable information on parameter interactions was derived. The four methods also highlighted the limitations of the 'ill-posed' water quality model (low maximum efficiency, low coverage of observations, and high parameter interaction), **highlighting the importance of understanding structural uncertainties**. Search algorithms, such as AMALGAM and SCEM-UA, were the most efficient methods in terms of computational requirements.

It can be concluded that the identification of the most appropriate method for uncertainty estimation is a trade-off between the need for a strong theory-based description of uncertainty (but limited by the requirements on prior knowledge - Bayesian approach, MICA), simplicity (but limited by the subjectivity - GLUE) and computational efficiency (also affected by subjectivity - AMALGAM and SCEM-UA). It is also suggested that different evaluation scenarios should be analysed (i.e. different catchments, models, data, etc). In addition, modellers should select the method which is most suitable for the system they are modelling (e.g. complexity of the model's structure including the number of parameters), their skill/knowledge level, the available information, and the purpose of their study.

Subsequently, the application of MICA to a rainfall runoff model (with 13 parameters) and a water quality model verified the potential of the method to assess urban drainage models in urban catchments of different sizes and land-use types (one highly urbanised and flat and the other mostly pervious and very steep). MICA was able to calibrate rainfall runoff the models, while identifying their sensitivity to each of their parameters and producing reasonable predictive uncertainty bands. It can be concluded that the method can be recommended to explore parameter calibration, model sensitivity and predictive uncertainties in stormwater models.

Chapter 5

Exploring calibration, sensitivity and uncertainties of stormwater models

DECLARATION FOR THESIS CHAPTER 5

Declaration by candidate

In the case of Section 5.2, the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up.	70

The following co-authors contributed to the work:

Name	**Nature of contribution**	**Extent of contribution (%)**
Ana Deletic	Initiation, ideas and reviewing	n/a
David T. McCarthy	Initiation, ideas and reviewing	n/a
Tim D. Fletcher	Initiation, ideas and reviewing	n/a

Candidate's Signature	[redacted]	**Date** 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(7) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(8) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(9) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(10) there are no other authors of the publication according to these criteria;

(11) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(12) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location	CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus	
Signature Ana Deletic	[redacted]	**Date** 29/10/2012
Signature David T. McCarthy	[redacted]	**Date** 29/10/2012

Location	Waterway Ecosystem Research Group. Melbourne School of Land & Environment. The University of Melbourne	
Signature Tim D. Fletcher	[redacted]	**Date** 29/10/2012

Declaration by candidate

In the case of Section 5.3, the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up.	60

The following co-authors contributed to the work:

Name	**Nature of contribution**	**Extent of contribution (%)**
Manfred Kleidorfer	Ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	20
Ana Deletic	Initiation, ideas and reviewing	n/a
Tim D. Fletcher	Reviewing	n/a
David T. McCarthy	Initiation, ideas and reviewing	n/a
Wolfgang Rauch	Reviewing	n/a

Candidate's Signature	[redacted]	**Date** 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location	Unit for Environmental Engineering, University of Innsbruck, Austria	
Signature Manfred Kleidorfer	[redacted]	**Date** 29/10/2012
Signature Wolfgang Rauch	[redacted]	**Date** 29/10/2012

Location	CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus	
Signature Ana Deletic	[redacted]	**Date** 29/10/2012
Signature David T. McCarthy	[redacted]	**Date** 29/10/2012

Location	Waterway Ecosystem Research Group. Melbourne School of Land & Environment. The University of Melbourne	
Signature Tim D. Fletcher	[redacted]	**Date** 29/10/2012

Declaration by candidate

In the case of Section 5.4, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up.	65

The following co-authors contributed to the work:

Name	Nature of contribution	Extent of contribution (%)
Manfred Kleidorfer	Initiation, ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	20
Ana Deletic	Initiation, ideas and reviewing	n/a
Wolfgang Rauch	Reviewing	n/a
David T. McCarthy	Initiation, ideas and reviewing	n/a
Tim D. Fletcher	Reviewing	n/a

Candidate's Signature	[redacted]	Date 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location	Unit for Environmental Engineering, University of Innsbruck, Austria	
Signature Manfred Kleidorfer	[redacted]	**Date** 29/10/2012
Signature Wolfgang Rauch	[redacted]	**Date** 29/10/2012

Location	CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus	
Signature Ana Deletic	[redacted]	**Date** 29/10/2012
Signature David T. McCarthy	[redacted]	**Date** 29/10/2012

Location	Waterway Ecosystem Research Group. Melbourne School of Land & Environment. The University of Melbourne	
Signature Tim D. Fletcher	[redacted]	**Date** 29/10/2012

5.1 Introduction

As the literature review (see Section 2.2) demonstrated, a number of conceptual models have been used to predict stormwater discharges from urban environments. While the rainfall runoff models are well established, water quality models are far less developed. The performance of different urban stormwater model structures (mainly in sub-daily timesteps) has not been sufficiently investigated. In addition, the assessment of the models' structures and their associated predictive uncertainty are yet to be fully explored. Therefore, the main aim of this research was to ***explore parameter calibration, model sensitivity and the resulting predictive uncertainties in models with different levels of complexity by applying the methods to the same case-study.***

This chapter focuses on addressing the following key research questions and hypotheses:

- What are the key calibration parameters that govern the urban rainfall runoff and water quality models, and do they depend on the model structure?
    - A well-posed and well-calibrated model (which has influential and identifiable parameters, see Carrera and Neuman, 1986 for extended definitions) will have a higher model efficiency. Providing inadequate calibration for a well-posed model may neglect important processes represented by the model; and,
    - a well-posed and well-calibrated model will be sensitive to all calibration parameters.
- Do the physical parameters used in stormwater models require calibration (or can they be reliably determined via specific in situ measurements)?
    - While some parameters are purely conceptual (non-physical, Kuczera et al., 2006), some parameters are intrinsically related to the physical factors, thus they should be measured whenever possible (e.g. soil property related parameters).
- Can we use model parameter sensitivity and its associated predictive uncertainties to understand the appropriateness of the model structure for the given application?
    - Results from a sound model sensitivity analysis will inform if the model is *well* or 'ill-posed', as the identifiability of parameters, the confidence in the model results and the existence of model structure and conceptual errors will be determined.
- What is the model predictive uncertainty originated only from parameter uncertainty (without taking into account other sources of uncertainties such as measurement errors in input and calibration data), and how does this uncertainty compare to the total uncertainties in the predicted results?
    - The assessment of the uncertainty originating from model parameters allows a comprehensive analysis of model structure and parameter interaction.

Nevertheless, other sources of uncertainties (e.g. input measured data, model formulation and assumptions and selected objective function) should be investigated because they impact on the total uncertainties in the modelled results.

The work has been published in three separate journal papers. The first paper presents results from calibration and sensitivity analysis of MUSIC and a simple regression water quality model, both presented in Chapter 3. This paper was presented at the *6th International Conference on Water Sensitive Urban Design and Hydropolis* held in Perth, Australia, in 2009, and was subsequently selected for publication in the *Australian Journal of Water Resources*. The peer reviewed version, *Calibration and sensitivity analysis of stormwater models,* published in 2011 is included in Section 5.2. The second paper was initially also a conference paper. It was presented at *8th International Conference on Urban Drainage Modelling jointly with the 2nd International Conference on Rainwater Harvesting and Management* held in Tokyo, Japan, in 2010. The work explored interactions between parameter sensitivity and model structure uncertainties in three water quality models. It was recommended for publication in *Water Science and Technology* and after revision and updates, the paper, *Stormwater quality models: performance and sensitivity analysis,* was published in this journal in 2010. This paper forms the body of text of Section 5.3. The third paper, *Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data*, was published in *Environmental Modelling and Software* in 2011 as included in Section 5.4. This comprehensive paper compared parameter sensitivity of the selected rainfall runoff models (more complex MUSIC and simple KAREN) and water quality models (build-up/wash-off and simple regressions) and also explores the predictive uncertainty associated with the rainfall runoff models.

5.2 Parameter sensitivity analysis of stormwater models

Calibration and sensitivity analysis of urban drainage models: MUSIC rainfall/runoff module and a simple stormwater quality model*

CBS Dotto†, A Deletic, DT McCarthy and TD Fletcher
Centre for Water Sensitive Cities, Department of Civil Engineering and eWater CRC, Monash University, Clayton, Victoria

ABSTRACT: *Model calibration and sensitivity analysis of stormwater models are required to assess model performance; it is very unlikely that non-calibrated models will lead to reasonable results. The aim of this paper is to present results of the calibration and sensitivity analysis of the key parameters used in flow modelling by MUSIC and parameters of a simple stormwater quality model. The assessment of the models is undertaken using a Monte Carlo Markov Chain approach. We describe the models' performance, provide information on their sensitivity to parameters and also discuss the correlation between these parameters. This work will help practitioners to understand importance of the MUSIC parameters that they usually use without calibration. The information reported in the results will also help to guide future development of stormwater quality models and the data needed to support it.*

1 INTRODUCTION

Restoring pre-development stormwater flow and quality is a goal of stormwater management practices in urban areas. In order to provide reasonable estimates of stormwater flows and pollution concentrations, robust modelling of stormwater discharges and their associated pollutants is critical.

In Australia, the Model for Urban Stormwater Improvement Conceptualisation (MUSIC) (eWater CRC, 2009) is the most widely used water quality stormwater model (eg. Elliott & Trowsdale, 2007; Mitchell et al, 2008). Allowing the assessment of pollutant generation and performance of stormwater treatment measures, it is extensively used for the design of urban development proposals and the evaluation of their impact on the environment. As a continuous catchment model, MUSIC generates the runoff from impervious and pervious surfaces and simulates simplified channel/pipe flow. While MUSIC currently simulates pollution generation using a stochastic approach with dry weather and wet weather event mean (and standard deviation) concentrations, there is also a trial module that estimates the loads within a time-step as a power function of the catchment's runoff (as described in Sartor & Boyd, 1972). However, uncertainties in MUSIC modelling are not yet well understood, as is the case for almost all stormwater models used in practice (both in Australia and worldwide).

One of the key sources of uncertainties in modelling is caused by model parameters (eg. Lindenschmidt, 2006; Gallagher & Doherty, 2007). Understanding sensitivity of models to their parameters helps to understand how inadequately defined parameters impact upon modelled results. There are a number of tools that could be used to assess parameter sensitivity. For example, the Metropolis algorithm (Metropolis et al, 1953), a general Monte Carlo Markov Chain (MCMC) sampling method, has been widely used for calibration and sensitivity analysis of general hydrological models (eg. Kuczera & Parent, 1998; Kanso et al, 2003). The robustness of the Bayesian approach compared with other methods is apparent from several studies (eg. Makowski et al, 2002; Gallagher & Doherty, 2007; Blasone et al, 2008).

The aim of this paper is to present results of the sensitivity analysis of the key parameters used in flow modelling by MUSIC and parameters of a simple stormwater quality model. This work will help practitioners to understand importance of the parameters that they usually use without much calibration. The information will also help to guide

* Reviewed and revised version of a paper presented at the 6th International Water Sensitive Urban Design Conference (WSUD2009), Perth, Western Australia, May 2009.

† Corresponding author Cintia Dotto can be contacted at [redacted]

future development of stormwater models and the data needed to support it.

2 METHODS

2.1 Data

Urban stormwater quality and quantity data from urban catchments, with different sizes and land uses, were used. These catchments are located in the eastern suburbs of Melbourne, Australia. Table 1 shows a summary of the characteristics of the catchments and some descriptive statistics of the measured data (Francey et al, 2010). Most of the sites were monitored from January 2004 to December 2005. All catchments are serviced by separate stormwater and wastewater systems. Narre Warren is the only site that contains septic systems (tanks).

Rainfall data was monitored every minute using 0.2 mm tipping bucket rain gauges located as close as possible to the catchments' centroid. The mean annual rainfall in these catchments ranges from 580 to over 720 mm/year. Continuous flow data (recorded at the 1 minute interval) were measured at each catchment and around 300 wet weather events were monitored for a range of pollutants, including Total Suspended Solids (TSS) and Total Nitrogen (TN) (except for Ruffeys Lake, Doncaster (RD) catchment, where only TSS was monitored). The water quality samples were collected at the outlet of the catchments using a discrete sampling methodology.

2.2 Models

2.2.1 *Rainfall/runoff modelling*

The rainfall/runoff model implemented in MUSIC (eWater CRC, 2009), which is based on the SimHyd model developed by Chiew & McMahon (1997), was assessed in this study. In MUSIC, the original daily model was modified to enable disaggregation of

Table 1: Summary of sites and measured data details (Francey et al, 2010).

Site	Gilby Rd (GR)	Richmond (RICH)	Ruffeys Lake, Doncaster (RD)	Shepherds Bush (SB)	Narre Warren (NW)
Primary land use	Commercial	High density residential	Medium density residential	Medium density residential	Rural residential
Area (ha)	28.2	89.1	105.6	38	10.5
Total impervious fraction (TIF)	0.8	0.74	0.51	0.45	0.2
Catchment average slope (%)	1	3.5	5		4
Time of concentration (min)	23	31	14	14	16
Distance from catch centroid to rain gauge (m)	100	600	700	550	250
Mean annual rainfall (mm/year)	723	650	650	580	700
Median maximum event rainfall intensity (mm/h)	10	10	7	6	10
Range of event total rainfall (mm)	1.2-38.6	1.2-40.8	1.6-95.8	0.6-21.0	0.4-110
Mean maximum runoff rate (L/s)	408	547	723	214	44
Range of event maximum flow rates (L/s)	75-2241	67-3867	164-3069	29-1200	14-90
Number of events – TSS	49	40	54	19	41
Maximum TSS concentration (mg/L)	867	1600	1422	1545	2398
Mean of TSS event mean concentration (EMC) (mg/L)	71.6	125.1	77.0	94.8	91.9
Number of events – TN	47	39	–	17	18
Maximum TN concentration (mg/L)	9	26	–	15	19
Mean of TN EMCs (mg/L)	1.17	2.29	–	1.74	3.51

daily runoff into sub-daily temporal patterns. Flow from impervious and pervious areas are modelled separately, with impervious area runoff being primarily a function of the proportion of catchment imperviousness, with an initial loss term. Runoff from the pervious areas will only occur during large or intense storm events, when the pervious soil storage is saturated. Therefore, pervious area runoff depends on soil properties, provided as a series of calibration parameters in the model. The model has 13 calibration parameters, which are presented in figure 1 according to their relationship to the flow processes, along with their default values for Melbourne.

2.2.2 *Water quality modelling*

The simple regression model adopted in this study estimates concentrations within a time-step as a power function of the routed runoff:

$$C_t = aR_t^b$$

where C_t is the pollutant concentration at time t (mg/L); R_t is the routed runoff (average, in mm/h, over the 6-minute time-step, t); and a and b are parameters to be calibrated. Although this approach was proven as unsuitable (Kanso et al, 2004), derivations of it have been adopted in several stormwater models, such as XP-AQUALM (XP-Software, 1999), SWMM5 (Rossman, 2008) and P8-UCM (Palmstrom & Walker, 1990). Therefore, a detailed investigation of this model, in terms of processes and parameters, provides key information for the development of more suitable models.

2.3 Calibration and sensitivity analysis

Calibration and sensitivity analysis were undertaken with the software MICA (Doherty, 2003). MICA quantifies a parameter's uncertainty by assuming a prior and a proposal probability distributions and subsequently updating this proposal distribution on the parameter's samples into a posterior distribution (PD). PDs of parameters indicate the most probable value to generate the best model performance, as well as how the model results depend on this parameter. A PD with a flat shape shows the insensitivity of the model to the certain parameter; therefore for a number of parameter values, the model achieves similar accuracy. In contrast, if the PD presents a well-defined peak, it indicates that the model is very sensitive to the parameter. MICA undertakes a simplified MCMC analysis and uses a particularly flexible version of the Metropolis-Hastings algorithm (Hastings, 1970). A uniform prior distribution was assumed for all parameters.

2.4 Evaluation of the model parameters

The calibrated values (using MICA) for the five studied catchments (table 1) were compared with values previously reported in literature. MUSIC parameters have been reported in a number of studies: the catchment impervious area is suggested according to the site's land use in a Melbourne Water's (2004) MUSIC modelling guidelines, while values for the pervious and baseflow related parameters (based on the catchment's top and subsoil conditions) were suggested by Macleod (2008). Therefore, the comparison focused on effective impervious fraction (EIF), soil storage capacity (SMax), field capacity (fc), infiltration capacity coefficient and exponent (coeff and sq), daily recharge rate (rfac), daily baseflow rate (bfac) and daily deep seepage rate (dseep). In addition, a comparison and evaluation of MUSIC default values for Melbourne was attempted.

3 RESULTS

3.1 Calibration and sensitivity analysis

3.1.1 *Rainfall/runoff modelling*

The overall efficiency of the rainfall/runoff model assessed in this study is represented by the Nash and

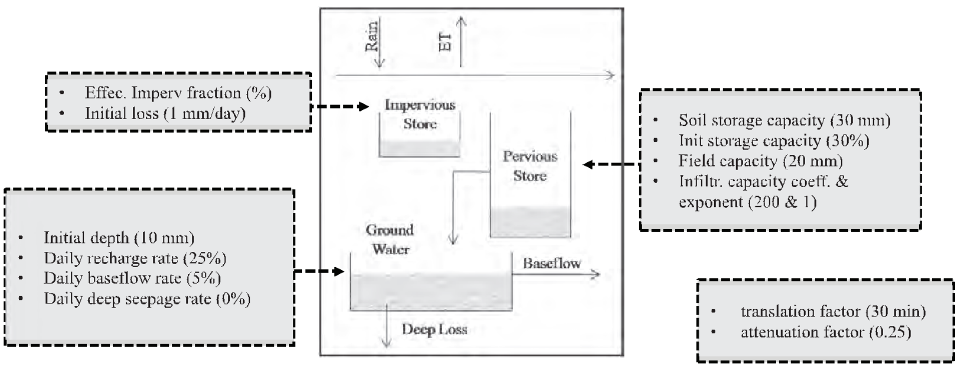


Figure 1: MUSIC rainfall/runoff parameters and their default values for Melbourne between brackets (after CRCCH, 2005).

Sutcliffe coefficient (E) (see Nash & Sutcliffe, 1970, for definition) in table 2. E_{best} stands for the best E obtained with the best set of parameters; E_{mean} indicates the mean E derived from all the accepted parameter sets, which were the sampled values within certain likelihood to generate the best model's outputs. The average efficiency coefficient was 0.61, the best one was 0.80 for RICH. For the four other catchments, MUSIC calibrations produced E_{best} from 0.49 (NW) to 0.62 (RD) (table 1). It is hypothesised that the lower values may be due to some gaps in the datasets, as well as the complexity of the processes in these catchments with different levels of impervious areas, and potential cross-connections with sewer system. MUSIC performance should still be investigated in the case of a dataset with rainfall events that actually produce significant pervious area flows, since this was not evident in any of the studied catchments' datasets. The poor performance for the most impervious catchment (GR) was unexpected, since the model was designed for urban conditions. Measurement errors in the input and calibration data, rainfall spatial distribution and some of the catchment's intrinsic characteristics, may be the cause of the poor fit.

Table 2: Rainfall/runoff models efficiency.

Catchment	TIF	MUSIC	
		E_{best}	E_{mean}
Gilby Rd, Mt. Waverley (GR)	0.80	0.54	0.51
Madden Grove, Richmond (RICH)	0.74	0.81	0.77
Ruffeys Lake, Doncaster (RD)	0.51	0.62	0.57
Shepherds Bush, Glen Waverley (SB)	0.45	0.57	0.44
Kilgerron Crt, Narre Warren Sth (NW)	0.20	0.49	0.46

The EIF and the Muskingum Cunge translation factor (K) were found to be very important to MUSIC's rainfall runoff model for all the different catchments. The "best fitting values" are clearly defined by the peaks in the histograms presented in figures 2(a) and 2(g), respectively.

EIF optimum values ranged from 0.11 (NW) to 0.45 (GR) for the five studied catchments. It is evident that the model is very sensitive to EIF and also that some other parameters are related to the EIF values; therefore specification of the EIF requires particular attention by the modeller. It is highly recommended that its value be calibrated. If there is no available data for model calibration, the use of satellite images to determine the TIF and a brief study of the drainage plan (to determine which impervious areas are directly connected to the drainage network) are required to give insights on the EIF value. K is the other very significant parameter, it reflects the travel time of the flood wave throughout the drainage system. All catchments differ in their intrinsic properties towards flows translation through the catchment (eg. EIF and time of concentration) and the model is very sensitive to it. K optimum values ranged from values around 5 (SB) to 24 min (GR).

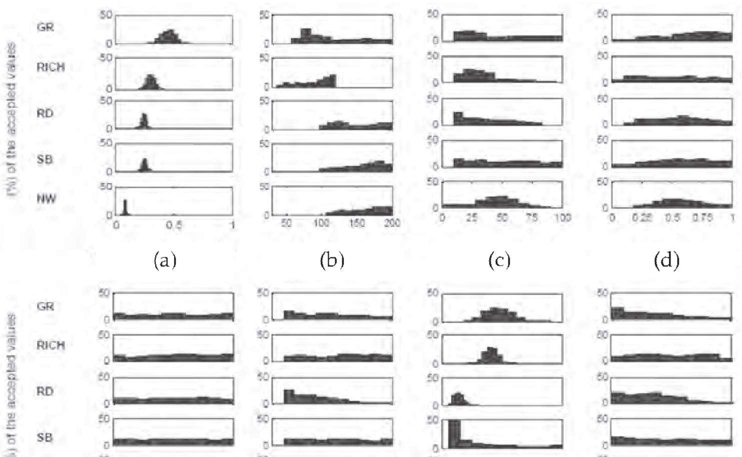


Figure 2: Sample histograms for selected model parameters based on MICA runs – (a) effective impervious fraction, EIF; (b) soil storage capacity, SMax; (c) field capacity, fc; (d) daily recharge rate, rfac; (e) groundwater initial storage, gw; (f) daily deep seepage rate, dseep; (g)translation factor, K; and (h) attenuation factor, θ.

MUSIC appeared to be sensitive also to the parameters related to the pervious area runoff (eg. SMax) and to the Muskingum Cunge attenuation factor for some other catchments. It should also be noted that some of the parameter distributions obtained with the MICA were not normal, contrary to the hypothesis of classical parameter uncertainty estimation methods. Moreover, not all of the parameters can be fitted to standard distributions.

According to figure 2(b), the SMax optimum values increased as the catchments' imperviousness decreased. It is apparent that the fc may be fixed in any valued between 10 and 40 mm, or even set to its default value (20 mm) for the two most urbanised catchments (GR and RICH). The same parameter did not show a standard behaviour for the remaining catchments. The model was sensitive to fc for RD and NW, but not for SB (figure 2(c)). From this sensitivity study, it is recommended that fc should be calibrated for catchments with EIF lower than 0.3. Similar patterns and recommendations are valid for the rfac, bfac and dseep.

It was verified that certain parameters presented a quite flat distribution, as in the case of the PDs obtained for the groundwater initial depth (gw), which are presented in figure 2(e); it is clear that any value between 10 and 30 may be assumed without compromising the results. MUSIC presented the same pattern for two more parameters. The model was also insensitive to the initial storage level in the pervious area storage (SIni) and to the infiltration capacity exponent (sq) for all catchments. These three parameters may be safely used in any value between their lower and upper limits adopted for the prior uniform distribution, or simply fixed to their default values when the model is applied to urban catchments with different levels of urbanisation (figure 1). It is suggested that the results found in this study can be used for modelling catchments with similar land use and hydrological behaviour. It is advised however, that MUSIC should be calibrated against local flow whenever data is available.

In summary, it was clear that the rainfall-runoff model MUSIC was not very sensitive to its pervious area parameters when applied to urbanised catchments. However, if the impervious area fraction was lower than 30%, the pervious area parameters (eg. soil storage and fc) became important. Such findings are not surprising, as runoff in urban catchments is driven mainly by impervious area flow and pervious runoff and/or baseflow were often minor or even non-existent in the studied catchments. In this context, the need of a model with such number of parameters in highly urbanised environments might be questioned. Dotto et al (2011) investigated this issue by comparing MUSIC to a single linear reservoir model, KAREN (Rauch & Kinzel, 2007) and found that the simple model can perform as well as MUSIC in terms of long time series simulation. However, peaks of high volume events are better captured by MUSIC due to its ability of modelling pervious area runoff.

3.1.2 *Water quality modelling*

The efficiency coefficients obtained for the power function model calibration were low (table 3). Despite the long time consumed to run the model along MICA, simulations did not produce satisfactory results for TSS and TN.

According to figure 4, for TSS, the optimum values for *a* ranged from 42 (GR) to 98 (NW), and *b* varied from 0.01 (RD and NW) to 0.69 (SB). For TN, the optimum values for *a* were between 1.3 (NW) and 2.86 (GR), and *b* ranged from 0.05 (SB) to 0.125 (NW).

The high correlation between parameters further complicated the calibration process. For example, R^2 between *a* and *b* ranged 0.43 (NW) and 0.91 (GR) for

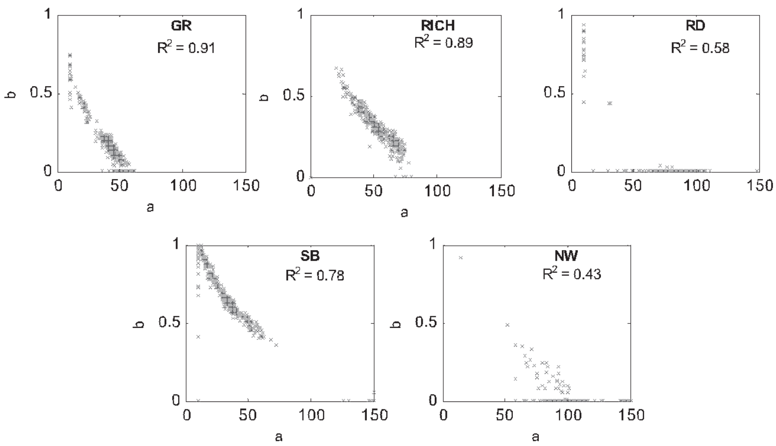


Figure 3: Simple regression model correlation between parameters for TSS.

Table 3: Simple regression model efficiency.

$C_t = aR_t^b$				
Catchment	TSS		TN	
	E_{best}	E_{mean}	E_{best}	E_{mean}
Gilby Rd, Mt. Waverley (GR)	0.03	0.01	–0.001	–0.035
Madden Grove, Richmond (RICH)	0.07	0.06	–0.04	–0.11
Ruffeys Lake, Doncaster (RD)	–0.06	–0.06	–	–
Shepherds Bush, Glen Waverley (SB)	–0.03	–0.04	–0.01	–0.11
Kilgerron Crt, Narre Warren Sth (NW)	–0.01	–0.02	–0.09	–0.41

TSS. The same high correlations were observed for TN. It means that different combination of *a* and *b* can lead to the same results. Examples of correlation graphs (figure 3) clearly illustrate the "ill-posed" nature of this model.

However, the model calibration failed, the obtained values suggested some interesting correlations. For example, *a*, *b* and *E* are correlated with the EIF with correlation coefficient (R^2) of 0.96, 0.31 and 0.58, respectively, for TN data. While *a* and *E* increased with EIF, *b* decreased. Same pattern was not observed for TSS (figure 4). In terms of water quality modelling approach, it is suggested that the power function model does not present any advantage than using mean concentration values and, that in fact, the use of mean values might represent an improvement.

3.2 Evaluation of the MUSIC flow model parameters

Melbourne Water (2004) suggests that the EIF should be calibrated whenever possible. In practice this is not the case and thus, values for the TIF according to the catchments' land use are presented in their report. It is usually the case that EIF is usually lower than the TIF and, therefore, it is advised that the use of the TIF will lead to an overestimation of the more frequent flows. According to the land use criteria, the TIF suggested values for the five catchments would ranged from 0.2 (NW) to 0.9 (GR) (table 4). The estimated values for TIF presented in table 2 ranged from 0.2 (NW) to 0.9 (GR), suggesting that the recommended values in the guidelines are often a little larger than the estimated one (expect for the most pervious catchment, NW). As expected, the EIF calibrated values presented in the histograms in figure 2 are significantly lower than the ones presented in table 4.

Regarding the pervious area-related parameters, the values suggested in the literature are generally slightly different from the calibrated ones, however, they are in the same range and most of them are within the 95% confidence interval (table 4). From the results obtained in this study, it is suggested that most parameters are really "calibration parameters" and some of them are related to the catchment's EIF (level of urbanisation). Further work is recommended for catchments where significant pervious flow is more apparent and also detailed soil profile studies are available.

The use of Melbourne's default values for the pervious area related parameters may be suitable (or at least not matter greatly) in the case of highly urbanised catchments (eg. GR and RICH). However, it is recommended that the values for some of these parameters be revised. For example, Melbourne's MUSIC default value for fc is 20 mm. The calibration conducted in this study suggested values ranging from 34 to 52 mm, and the reported literature values for fc in the respective soil conditions ranged from 69 to 79 mm. Similar applies for SMax: default value is 30 mm, literature values according to soil properties should be in a range between 97 and 175 mm, while the calibrated values are from 89 to 161 mm.

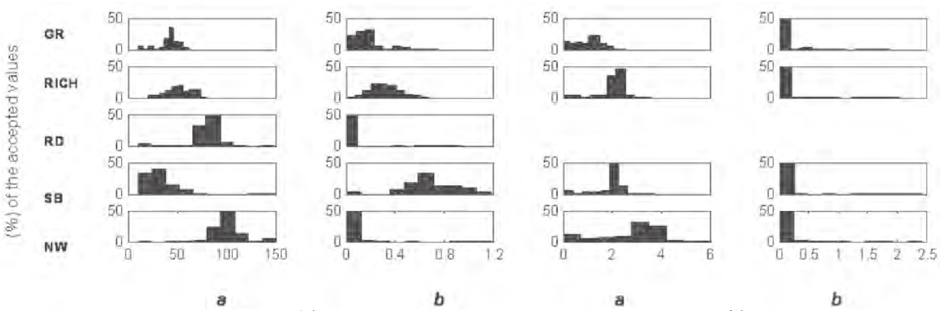


Figure 4: Histograms for the water quality model parameters based on MICA runs – (a) TSS and (b) TN.

Table 4: Parameters values suggested in literature and calibrated ones with their respective 95% confidence interval (literature values are from Melbourne Water (2004), CRCCH (2005) and Macleod (2008)). Values in bold indicate parameters outside of the 95% confidence level.

	GR				RICH			
Par	**Literature value**	**Calibrated values**			**Literature value**	**Calibrated values**		
		mean	**95% CI**			**mean**	**95% CI**	
EIF	**0.9**[a]	0.45	0.33	0.57	**0.80**[a]	0.30	0.24	0.38
S1Max	175	115.10	65.75	194.36	97	89.37	45.77	119.98
Fc	74	49.00	10.00	99.00	**79**	34.60	11.45	74.55
Coeff	**360**	152.69	100.00	197.73	**250**	160.44	105.67	198.01
Sq	0.50	1.04	0.91	1.19	**1.30**	1.04	0.90	1.19
Rfac	**1.00**	0.63	0.11	0.99	0.60	0.51	0.06	0.98
Bfac	1.00	0.62	0.10	1.00	0.45	0.47	0.02	0.98
Dseep	0.00	0.24	0.05	0.49	**0.00**	0.28	0.07	0.49
	RD				SB			
Par	**Literature value**	**Calibrated values**			**Literature value**	**Calibrated values**		
		mean	**95% CI**			**mean**	**95% CI**	
EIF	**0.60**[a]	0.25	0.21	0.29	**0.60**[a]	0.21	0.08	0.33
S1Max	97	149.06	102.22	197.89	139	157.10	100.88	197.44
Fc	**79**	35.77	11.00	77.00	69	51.96	10.00	100.00
Coeff	**250**	144.21	100.00	200.00	**360**	149.68	100.00	197.94
Sq	**1.30**	1.06	0.91	1.20	0.50	1.05	0.90	1.20
Rfac	0.60	0.57	0.20	0.97	**1.00**	0.56	0.06	0.98
Bfac	0.45	0.72	0.35	0.99	0.50	0.55	0.07	0.97
Dseep	**0.00**	0.19	0.05	0.45	**0.00**	0.26	0.05	0.50
	NW							
Par	**Literature value**	**Calibrated values**			**Melbourne default values**		**Units**	
		mean	**95% CI**					
EIF	**0.20**[a]	0.11	0.06	0.10	–			
S1Max	98	160.83	115.19	198.75	30		mm	
Fc	70	43.46	2.00	74.00	20		mm	
Coeff	**250**	156.62	116.23	197.36	200		–	
Sq	**1.30**	1.05	0.91	1.19	1.00		–	
Rfac	0.60	0.58	0.26	0.96	0.25		/100%	
Bfac	0.45	0.44	0.15	0.85	0.05		/100%	
Dseep	**0.00**	0.37	0.18	0.50	0.00		/100%	

[a] values in the literature are suggested for the TIF, there is no guidance on values for the EIF.

4 CONCLUSION

A Bayesian method was tested for calibration and sensitivity analysis for rainfall/runoff and water quality models. MUSIC rainfall/runoff module (with 13 parameters) was used in terms of catchment runoff, while a simple regression model (two parameters) was tested in terms of catchment pollution. A comprehensive stormwater dataset, containing data on stormwater flows and pollution concentrations from five urban catchments located in Melbourne, Australia, was used for the analysis.

The efficiency of the calibration and sensitivity analysis approach was verified in this work. The approach seems to be promising in generating the PDs and provided some valuable information on parameter correlation.

It was clear that the rainfall/runoff model is not very sensitive to its pervious area parameters when applied to highly urbanised catchments, in which pervious area runoff and baseflow are almost inexistent. For such areas, the need of a complex model with so many parameters related to the pervious flow of the catchment is questionable. The same does not apply when the level of urbanisation and impervious area fraction is lower than 30%; in this case some of the pervious area parameters (eg. soil storage and fc) become important. Preliminary results suggested that most of those parameters are in fact "calibration parameters" and not ones which relate to physical characteristics of the catchment. While some of these parameters are strongly related to the catchment's EIF, further work is required in catchments with significant pervious flows to understand their soil profile characteristics. It is suggested that such results can be used for modelling catchments with similar land use, climatic characteristics and hydrological behaviour. It is advised, however, that MUSIC should be calibrated against local flow whenever data is available.

Even with this robust calibration and parameter sensitivity approach, it is clear that the tested water quality model poorly represents reality and presents a high level of uncertainty. In terms of modelling approach, it is suggested that the power function model does not present any advantage than using mean concentration values; and that in fact the use of mean values might represent an improvement.

REFERENCES

Blasone, R.-S., Madsen, H. & Rosbjerg, D. 2008, "Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling", *Journal of Hydrology*, Vol. 353, No. 1-2, pp. 18-32.

Chiew, F. H. S. & McMahon, T. A. 1997, "Modelling daily runoff and pollutant load from urban catchments", *Water*, Journal of the Australian Water and Wastewater Association, pp. 16-17.

Cooperative Research Centre for Catchment Hydrology (CRCCH), 2005, *MUSIC Version 3 – User guide*.

Doherty, J. 2003, *MICA – Model-Independent Markov Chain Monte Carlo Analysis – User Manual*, Watermark Numerical Computing, US EPA.

Dotto, C. B. S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D. T. & Fletcher, T. D. (2011), "Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data", *Environmental Modelling & Software*, Vol. 26, No. 10, pp. 1225-1239.

Elliott, A. H. & Trowsdale, S. A. 2007, "A review of models for low impact urban stormwater drainage", *Environmental Modelling & Software*, Vol. 22, No. 3, pp. 394-405.

eWater CRC, 2009, "Model for urban stormwater improvement conceptualisation (MUSIC) (Version 4.0)", eWater Cooperative Research Centre, Canberra.

Francey, M., Fletcher, T. D., Deletic, A. & Duncan, H. 2010, "New insights into water quality of urban stormwater in South Eastern Australia", *Journal of Environmental Engineering*, Vol. 136, No. 4, pp. 381-390.

Gallagher, M. & Doherty, J. 2007, "Parameter estimation and uncertainty analysis for a watershed model", *Environmental Modelling & Software*, Vol. 22, No. 7, pp. 1000-1020.

Hastings, W. K. 1970, "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, Vol. 57, pp. 97-109.

Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B. & Chebbo, G. 2003, "Bayesian approach for the calibration of models: application to an urban stormwater pollution model", *Water Science and Technology*, Vol. 47, No. 4, pp. 77-84.

Kanso, A., Tassin, B. & Chebbo, G. 2004, "Managing uncertainties in urban runoff quality models: a branchmarking methodology", *5th Novatech*, Lyon, France, pp. 199-406.

Kuczera, G. & Parent, E. 1998, "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm", *Journal of Hydrology*, Vol. 211, No. 1-4, pp. 69-85.

Lindenschmidt, K. E. 2006, "The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling", *Ecological Modelling*, Vol. 190, pp. 72-86.

Macleod, A. 2008, "MUSIC calibration based on soils conditions", SW and QLD Stormwater Industry Association, Surfers Paradise, QLD, Australia.

Makowski, D., Wallach, D. & Tremblay, M. 2002, "Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods", *Agronomie*, Vol. 22, No. S1, pp. 191-203.

Melbourne Water, 2004, *MUSIC input parameters*.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, A. H. 1953, "Equation of state calculations by fast computing machines", *The Journal of Chemical Physics*, Vol. 21, No. 6, pp. 1087-1092.

Mitchell, V. G., McCarthy, D. T., Deletic, A. & Fletcher, T. D. 2008, "Urban stormwater harvesting – sensitivity of a storage behaviour model", *Environmental Modelling & Software*, Vol. 23, No. 6, pp. 782-793.

Nash, J. E. & Sutcliffe, J. V. 1970, "River flow forecasting through conceptual models part I – A discussion of principles", *Journal of Hydrology*, Vol. 10, No. 3, pp. 282-290.

Palmstrom, N. & Walker, W. W. J. 1990, *P8 urban catchment model: User's guide. Program documentation, and evaluation of existing models, design concepts, and Hunt-Potowomut data inventory*, T. N. B. P. R. N. NBP-90-50.

Rauch, W. & Kinzel, H. 2007, *KAREN – User Manual*, hydro-IT GmbH, Innsbruck, (in German).

Rossman, L. A. 2008, *Storm Water Management Model – User's Manual Version 5.0*, National Risk Management Research Laboratory, US Environmental Protection Agency, Cincinnati, Ohio.

Sartor, J. D. & Boyd, G. B. 1972, "Water pollution aspects of street surface contaminants",Rep. EPA-R2-72-081, US Environmental Protection Agency, Washington, DC.

XP-Software, 1999, "XP-AQUALM - Technical Description", XP Software, Australia.

837 © IWA Publishing 2010 Water Science & Technology—WST | 62.4 | 2010

Stormwater quality models: performance and sensitivity analysis

C. B. S. Dotto, M. Kleidorfer, A. Deletic, T. D. Fletcher, D. T. McCarthy and W. Rauch

ABSTRACT

The complex nature of pollutant accumulation and washoff, along with high temporal and spatial variations, pose challenges for the development and establishment of accurate and reliable models of the pollution generation process in urban environments. Therefore, the search for reliable stormwater quality models remains an important area of research. Model calibration and sensitivity analysis of such models are essential in order to evaluate model performance; it is very unlikely that non-calibrated models will lead to reasonable results. This paper reports on the testing of three models which aim to represent pollutant generation from urban catchments. Assessment of the models was undertaken using a simplified Monte Carlo Markov Chain (MCMC) method. Results are presented in terms of performance, sensitivity to the parameters and correlation between these parameters. In general, it was suggested that the tested models poorly represent reality and result in a high level of uncertainty. The conclusions provide useful information for the improvement of existing models and insights for the development of new model formulations.

Key words | Bayesian approach, calibration, parameter sensitivity, rainfall/runoff, stormwater quality model, urban pollutants

C. B. S. Dotto (corresponding author)
A. Deletic
T. D. Fletcher
D. T. McCarthy
Department of Civil Engineering and eWater CRC,
Institute for Sustainable Water Resources,
Monash University,
Victoria 3800,
Australia
E-mail: [redacted]

M. Kleidorfer
W. Rauch
Unit of Environmental Engineering,
Faculty of Civil Engineering,
University of Innsbruck,
Technikerstrasse 13,
A-6020 Innsbruck,
Austria
E-mail: [redacted]

INTRODUCTION

Stormwater models underpin decision-making processes in stormwater management. Runoff generation and flow routing models are now well developed and widely adopted. On the other hand, stormwater quality models are less well established. Due to limited data-availability, the performance of these models is mostly proven on few single events and the search for reliable models remains the object of current research endeavours. Understanding the processes within pollution generation is very important for better modelling approaches. In general, these processes are very complex and are influenced by a variety of parameters, such as: rainfall, runoff, climatic variables, land use and surface characteristics (Deletic & Maksimovic 1998; Egodawatta *et al.* 2007; McCarthy *et al.* 2007). This complex nature of pollutant accumulation and washoff, along with high temporal and spatial variations, generate technical difficulties in the development of accurate and reliable models of pollutant processes. Different approaches are available for reproducing the response of the sites in terms of pollutants: process-based buildup-washoff model as developed by Sartor *et al.* (1974), washoff rating curves as tested by Vaze & Chiew (2003), and Event Mean Concentration methods as proposed by Duncan (1999). In this context, model calibration is a "must do" exercise when pollution generation models are to be applied.

A number of studies have investigated the applicability and performance of the most popular water quality models. Vaze & Chiew (2003), for example, compared a number of simple regression equations and a process-based model, which estimate event pollutant loads from impervious

doi: 10.2166/wst.2010.325

surfaces. Between 14 and 20 events, with rainfall, flow and concentration data from three urban catchments in Australia were used in the study. Their results indicated that, once calibrated, both approaches estimated event pollutant loads satisfactorily. They also emphasized that the major problems in pollutant generation modelling are the lack of water quality data and the large variability in the pollutant concentration data. Sriananthakumar & Codner (1992) and (Kanso *et al.* (2004) are also examples of efficient estimation of fluxes and total loads from urban catchments. However, fluxes are driven by flow rates, which generates a degree of spurious correlation in load-based models and therefore masks the real predictive capability of the models (McCarthy 2008). Moreover, false conclusions about the intrinsic characteristics of pollutants and their sources could be easily derived. For these reasons, estimation of concentrations rather than fluxes would be indicated. Although few studies investigated pollutant concentrations modelling, they usually utilised a very limited number of events. Rodríguez *et al.* (2009), for instance, successfully tested different buildup-washoff models, nevertheless only two events were single calibrated.

Model sensitivity analysis is also crucial in order to estimate realistic stormwater pollutant concentrations and better understand the model parameters. Robust methodologies are currently available for global sensitivity analysis with the advantage of calibrating the model simultaneously. The most popular approaches are: (a) the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven & Binley 1992), known also as pseudo-Bayesian or informal Bayesian (Freni *et al.* 2009), and (b) the formal Bayesian methods, such as Monte Carlo Markov Chain (MCMC) methods (Kuczera & Parent 1998). The Metropolis algorithm (Metropolis *et al.* 1953), a general MCMC sampling method, has been widely used for model calibration and sensitivity analysis (e.g. Kuczera & Parent 1998; Kanso *et al.* 2003). The robustness of the Bayesian approach compared to other methods is apparent from several studies (e.g Makowski *et al.* 2002; Gallagher & Doherty 2007).

The aim of this paper is to present results of parameter calibration, sensitivity analysis and performance of three water quality models in order to define their reliability for different domains and applications. Two simple regressive equations and one buildup-wash off based model currently used to predict pollution generation in urban catchments were chosen. This study is based on comprehensive stormwater dataset which contains rainfall, stormwater flows and pollutant concentrations from five urban catchments, in Melbourne, Australia. The information from the paper will also guide future development of such models and indicate the data required to support their development and application.

METHODS

Data and models

Three empirical continuous concentration models derived from those used in practice, such as US Army Corps's STORM model (USACE 1977), SWMM (Rossman 2008), P8-UCM (Palmstrom & Walker 1990) and XP-AQUALM (1999) have been investigated. Their equations and parameters are summarised in Table 1.

Model 1 and Model 2 are simple regressive models, in which concentrations are estimated within a timestep as a power function of rainfall intensities and routed runoff, respectively. In these models, C_t is the pollutant concentration at time t (mg/L); I_t is the rainfall intensity (mm/hr, over the timestep, t), R_t is the routed runoff (average, in mm/hr, over the timestep, t); a and b are parameters to be calibrated. Model 3 is a buildup-washoff based approach (Sartor *et al.* 1974). Buildup is the process in which pollutants accumulate in the surface over a dry weather period, whereas washoff is the process of removing this accumulated pollution load by rainfall and incorporating it to the surface runoff. Buildup during dry periods was calculated after Sartor & Boyd (1972) and (Deletic *et al.* (2000) as presented in Table 1. Where $\overline{M(t)}$ is the amount of the pollutant available on the surface averaged over the area (g/m^2) during the dry weather period (t_d), M_0 is the

Table 1 | Water quality models

Models		Parameters
Model 1	$C_t = aI_t^b$	a and b
Model 2	$C_t = aR_t^b$	a and b
Model 3	Buildup : $\frac{d\overline{M(t)}}{dt_d} = k_1 \cdot (M_0 - \overline{M(t)})$	M_0 and k_1
	Washoff : $\frac{d\bar{C}(t+t_f)}{dt} = k_2 \cdot \overline{M(t)} \cdot I(t)^{k_3} \cdot A_i$	k_2 and k_3

maximum amount of the pollutant that can be stored at the surface (g/m^2) and k_1 is an accumulation constant (day^{-1}). Consequently, the calibration parameters for buildup are M_0 and k_1. Washoff during wet weather was calculated directly from rainfall intensity (not runoff) after an exponential function, in which $\bar{C}$ is the washoff concentration (mg/l), $\overline{M(t)}$ is the amount of solids available on the surface averaged over the area according to the buildup equation, $I(t)$ is the rainfall intensity (mm), A_i is the impervious area (m^2), k_2 is the washoff coefficient and k_3 is the washoff exponent. The models were calibrated using concentrations rather than loads given the reasons discussed in the introduction section.

Such models were applied to a comprehensive stormwater dataset, which contains rainfall, stormwater flows and pollutant concentrations from five urban catchments of different land uses and sizes located in Melbourne, Australia. Mean annual rainfall in the studied catchments ranged from 600 to 800 millimetres per year. Hundreds of events were available for which Total Suspended Solids (TSS) and Total Nitrogen (TN) were monitored (except for one catchment, where only TSS was monitored). TSS average Event Mean Concentration (EMC) between sites ranged from 71.6 to 125.1 mg/L, and TN was between 1.17 and 3.51 mg/L. All catchments are serviced by separate stormwater and wastewater systems, but some cross-connections between systems are expected. Narre Warren is the only site in that contains septic systems (tanks). Table 2 summarises the catchment characteristics and detailed description is available in Francey *et al.* (2010).

The Effective Impervious Fractions (EIFs) presented in Table 2 and the flow rates used in Model 2 were obtained from the calibration of a rainfall/runoff model; detailed description of the rainfall/runoff model and results for the specific catchments are available in Dotto *et al.* (2009).

Calibration and sensitivity analysis

Calibration and sensitivity analysis were undertaken with the software MICA (Doherty 2003). MICA undertakes a simplified MCMC analysis and uses a particularly flexible version of the Metropolis-Hastings algorithm (Hastings 1970). Parameter's uncertainty is quantified following a Bayesian approach by assuming a prior probability distribution and subsequently updating this prior on the parameter's samples into a Posterior Distribution (PD). Prior distributions specify the previous knowledge about the parameter values and/or ranges. This knowledge is often limited, mainly in the case of conceptual models, in which the variables are only empirical. A uniform distribution was assumed as prior distribution for all parameters and their lower and upper limits were chosen according to pre-calibration information. PDs of parameters indicate the most probable value that generates the best model performance (the expected value of the distribution), as well as how the model results depend on this parameter (the dispersion of the distribution). A PD with a flat shape reveals the insensitivity of the model to such parameter; therefore for a vast number of parameter values the model achieves similar accuracy. In contrast, if the PD presents a well defined 'peak', it indicates that the model is very sensitive to the parameter. Multiple local optima can be identified by multiple peaks in the PD.

For evaluating calibration performance (i.e. comparing the i measured (M) and simulated (S) data points) the Nash-Sutcliffe efficiency coefficient E (Nash & Sutcliffe 1970)

Table 2 | Summary of the catchment characteristics

Site	Primary land use	Area (ha)	TIF	EIF	No. of events	
					TSS	TN
Gilby Rd (GR)	Commercial	28.2	0.8	0.45	49	47
Richmond (RICH)	High density residential	89.1	0.74	0.30	40	39
Ruffeys Lake, Doncaster (RD)	Medium density residential	105.7	0.51	0.25	54	–
Shepherds Bush (SB)	Medium density residential	38	0.45	0.21	19	17
Narre Warren (NW)	Low density	10.5	0.2	0.11	41	18

TIF is the total impervious fraction of the catchment and EIF is the effective impervious fraction.

was chosen:

$$E = 1 - \frac{\sum_{i=1}^{n}(M_i - S_i)^2}{\sum_{i=1}^{n}(M_i - \bar{M}_i)}[-\infty|1]$$

Where: E is related to least square error, but for better appreciation it is normalized by $\sum_{i=1}^{n}(M_i - S_i)^2$ where $\bar{M}_i$ is the mean value of the observed data points.

RESULTS AND DISCUSSION

Contrary to rainfall and flow data, water quality data obtained by discrete sampling is not continuous, which makes pollution model calibration very complex and difficult; indeed, data availability represents one of the key obstacles to calibration of water quality models. Despite the long time consumed for the MICA runs, the models did not perform well and very low efficiency coefficients (E) were obtained (Tables 3 and 4). The buildup-washoff model generally performed slightly better than the others, simpler models, suggesting that there is some merit in accounting for the buildup-washoff processes. In general, calibration failed, suggesting that the model structure requires improvement. In addition, information on model efficiency is often limited in the literature. There are few studies, in which E values are presented, but they are not comparable to the present one as the models in those studies were calibrated against fluxes (e.g. Vaze & Chiew 2003; Kleidorfer *et al.* 2009).

Figure 1 presents the parameter histograms produced by the MICA runs for TSS (Figure 1(a)) and TN (Figure 1(b)). Such histograms belong to Model 2, in which pollutant concentrations are predicted as a power function of the runoff rate. The clear peaks in the histograms indicate that the model was sensitive to both a and b parameters. The same histograms suggested that parameters were generated with a high level of uncertainty as evidenced by their wide confidence intervals. The same outcome was observed in the results for both pollutants (TSS and TN) using Model 1 (not shown).

Table 3 | Nash and Sutcliffe coefficient (E) obtained for the models—TSS

Site	Model 1	Model 2	Model 3
Gilby Rd (GR)	−0.04	0.03	0.07
Richmond (RICH)	−0.10	0.07	0.12
Ruffeys Lake, Doncaster (RD)	0.06	0.14	0.22
Shepherds Bush (SB)	0.26	−0.03	0.06
Narre Warren (NW)	−0.04	−0.01	0.46

Table 4 | Nash and Sutcliffe coefficient (E) obtained for the models—TN

Site	Model 1	Model 2	Model 3
Gilby Rd (GR)	−0.07	−0.001	0.04
Richmond (RICH)	−0.30	−0.04	0.09
Ruffeys Lake, Doncaster (RD)	–	–	–
Shepherds Bush (SB)	−0.38	−0.01	0.26
Narre Warren (NW)	−0.01	−0.09	0.36

The optimum values of parameter a obtained for Model 2 during the TSS calibration ranged from 42 (GR) to 98 (NW) while parameter b varied from 0.01 (RD and NW) to 0.69 (SB) (Figure 1(a)). The changes in a between sites illustrate the effect of TSS sources on the model calibration. Higher TSS concentrations would be expected in more urbanised catchments. However, results in this study suggested the opposite; i.e. the inverse relationship between a and the catchments' imperviousness (as seen in Figure 1(a)). Few reasons might support this finding. For example, the minimum value achieved for GR may be explained by the fact that this is essentially a commercial area thus generating lower TSS concentrations than areas where residential activities are taken place (Duncan 1999). On the other hand, the maximum a value obtained for the most pervious catchment (NW) is probably due to the presence of some cross-connections between the storm-water and wastewater systems in this catchment. The optimum values for b varied from 0.01 (RD and NW) to 0.69 (SB) for TSS modelling. This indicates the effect of the kinetic energy of the rainfall in detaching TSS from the surface. For TN, the optimum values for a were between 1.3 (NW) and 2.86 (GR) and b ranged from 0.05 (SB) to 0.125 (NW). Parameter b values near zero indicate that TN modelling is mainly driven by parameter a, suggesting that TN concentrations are independent of the kinetic energy of the rainfall and highly dependent on the sources of such pollutant. This is in accordance to the literature, given the highly dissolved nature of nitrogen in urban stormwater (Taylor 2006). Other interesting correlations were identified. For instance, parameter a and the effective impervious fraction (EIF) were again inversely proportional to one

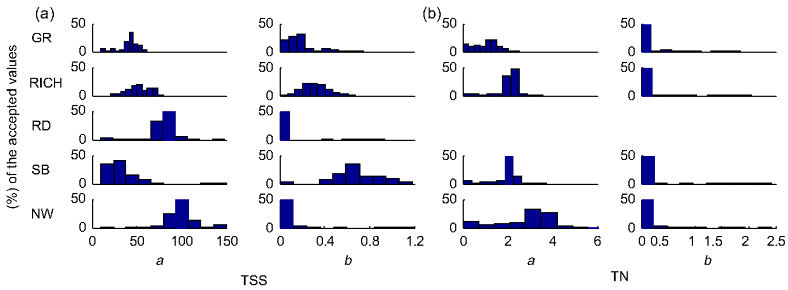


Figure 1 | Histograms for Model 2 with the parameters based on MICA runs—TSS and TN.

another for TN with Model 1 and 2. It is a rather interesting finding compared to Tong & Chen (2002). The authors found that the total amount of nitrogen was higher in impervious urban areas than in pervious ones. The contrast between findings could be possibly explained by the fact that Tong & Chen (2002) calibrated their water quality model against fluxes and therefore, areas with higher runoff produced higher loads.

Figure 2 reveals the correlation between parameters for TSS with Model 2. It is evident that the high correlation between parameters further complicated the calibration process. For example, R^2 between a and b ranged from 0.43 (NW) and 0.91 (GR) for TSS. The same high correlations were observed for TN, both for Models 1 and 2. This means that different combinations of a and b can lead to the same results. Figure 2 clearly illustrates the 'ill-posed' nature of these models.

Contrary to the simple regression models, the distributions obtained for the buildup-washoff model did not indicate one clear peak, instead resulted in multiple local optima for TSS as seen in Figure 3(a). Several peaks in these distributions suggest that several combinations of parameters are possible, confirming the problems of equifinality (Beven & Freer 2001) that may plague urban stormwater quality models. Distributions of calibration parameters for TN modelling are presented in Figure 3(b) and showed clearer peaks for three of the four calibration parameters. The model was insensitive to parameter k_1 for all catchments and therefore, indicates less influence of the dry wheatear processes. Parameter k_3 was close to zero, with all values lower than 1, once again implying that TN concentrations are independent of the kinetic energy.

It is interesting to note that there is no correlation between parameter sensitivity (seen in Figure 3) and maximum calibration performance (highest E–seen in Tables 3 and 4). Although, NW's calibration parameters seemed to be the least sensitive among the sites, it achieved the highest efficiency E.

The efficiency of the calibration and sensitivity analysis approach was also verified in this work. The approach seems to be promising in generating the posterior distri-

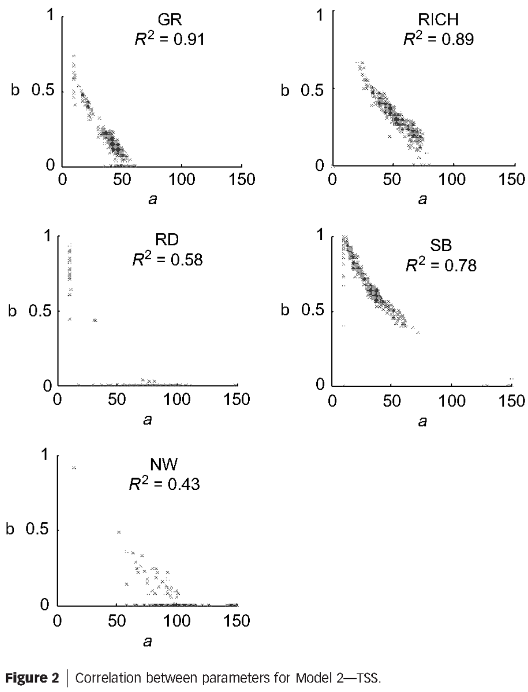


Figure 2 | Correlation between parameters for Model 2—TSS.

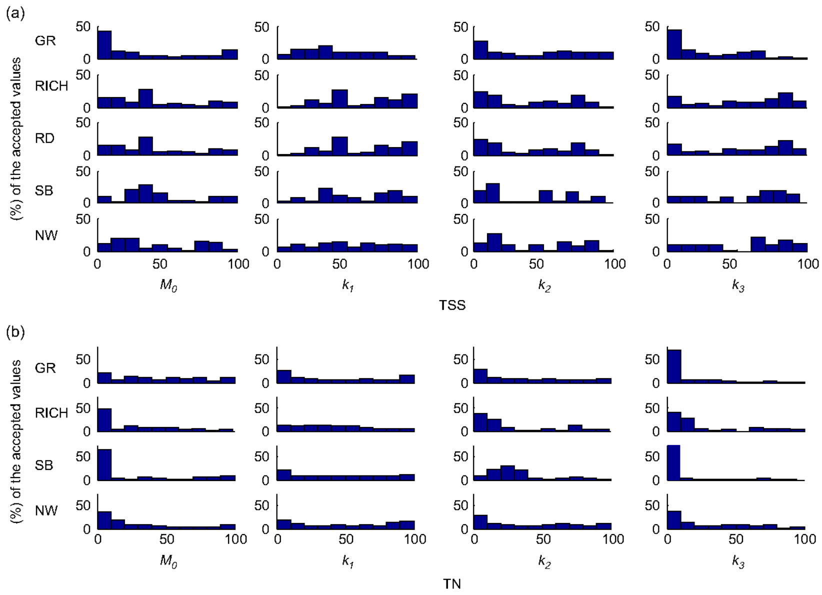


Figure 3 | Histograms for Model 3 with the parameters based on MICA runs—TSS and TN.

butions and gave some valuable information on parameter correlation. Further work should be done to investigate how the nature of the proposal distribution impact on the posterior distribution generated with the multiple runs.

CONCLUSIONS

Even with this robust calibration and parameter sensitivity approach, it is clear that the water quality models tested here poorly represent reality and result in a high level of uncertainty. It is not clear whether the structure of the power function models (perhaps over-simplified) is the main cause of their poor performance. However, the outcomes provide useful information for the improvement of existing models and also offered insights for the development of new model formulations. For example, it is recommended that future efforts be put into the development of models which use routed runoff or rainfall intensities, rather than the models which use 'unrouted' variables. Whilst routing essentially introduces extra model parameters, the temporal accuracy gained is likely to outweigh the calibration costs. The next step is to investigate the relationship between pollutants and some possible explanatory factors, such as: antecedent climatic variables, stormwater rainfall and flow variables in order to develop more efficient pollution generation model.

The study developed was also very important to verify the efficiency of the calibration and sensitivity analysis approach. The method presented seems to be promising in terms of generating the posterior parameter distributions and also gave some valuable information on parameter interaction. However, it is suggested that further work should be done to investigate the impact of the proposal distribution on the posterior distribution generated with the MCMC procedure, particularly in the case where highly correlated parameters are present.

ACKNOWLEDGEMENTS

The authors would like to acknowledge eWater CRC and CAPES for the scholarship provided and *hydro*-IT GmbH for software support.

REFERENCES

Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Processes* **6**(3), 279–298.

Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **246**, 11–29.

Deletic, A. & Maksimovic, C. T. 1998 Evaluation of water quality factors in storm runoff from paved areas. *J. Environ. Eng.–ASCE* **124**, 869–879.

Deletic, A., Ashley, R. & Rest, D. 2000 Modelling input of fine granular sediment into drainage systems via Gully-Plots. *Water Res.* **34**(15), 3836–3844.

Doherty, J. 2003 MICA–Model-Independent Markov Chain Monte Carlo Analysis–User Manual, Watermark Numerical Computing–U.S. EPA.

Dotto, C. B. S., Deletic, A., Fletcher, T. D. & McCarthy, D. T. 2009 Parameter sensitivity analysis of stormwater models. *The 6th International Water Sensitive Urban Design Conference and 3rd International Hydropolis*. Perth, Western Australia.

Duncan, H. P. 1999 Urban stormwater quality: a statistical overview. *Cooperative Research Centre for Catchment Hydrology*, Melbourne, Australia, p. 80.

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

Francey, M., Fletcher, T. D., Deletic, A. & Duncan, H. 2010 New insights into water quality of urban stormwater in South Eastern Australia. *J. Environ. Eng.–ASCE* **136**, 382–390.

Freni, G., Mannina, G. & Viviani, G. 2009 Urban runoff modelling uncertainty: comparison among Bayesian and pseudo-Bayesian method. *Environ. Modell. Softw.* **24**(9), 1100–1111.

Gallagher, M. & Doherty, J. 2007 Parameter estimation and uncertainty analysis for a watershed model. *Environ. Modell. Softw.* **22**(7), 1000–1020.

Hastings, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B. & Chebbo, G. 2003 Bayesian approach for the calibration of models: application to an urban stormwater pollution model. *Water Sci. Technol.* **47**(4), 77–84.

Kanso, A., Tassin, B. & Chebbo, G. 2004 Managing uncertainties in urban runoff quality models: a branchmarking methodology. *5th Novatech*, Lyon, France, pp. 199–406.

Kleidorfer, M., Deletic, A., Fletcher, T. D. & Rauch, W. 2009 Impact of input data uncertainties on stormwater model parameters. *Water Sci. Technol.–WST* **60**(6), 1545–1554.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* **211**(1–4), 69–85.

Makowski, D., Wallach, D. & Tremblay, M. 2002 Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie* **22**(S1), 191–203.

McCarthy, D. T. 2008 Modelling microorganisms in urban stormwater. *Civil Engineering*. Melbourne, Monash University. PhD Thesis: 488.

McCarthy, D. T., Mitchell, V. G., Deletic, A. & Diaper, C. 2007 *Escherichia coli* in urban stormwater: explaining their variability. *Water Sci. Technol.* **55**(11), 27–34.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, A. H. 1953 Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**(6), 1087–1092.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting trough conceptual models part I–a discussion of principles. *J. Hydrol.* **10**(3), 282–290.

Palmstrom, N. & Walker, W. W. J. 1990 P8 urban catchment model: User's guide. Program documentation, and evaluation of existing models, design concepts, and Hunt-Potowomut data inventory. T.N.B.P.R. N. NBP-90-50.

Rodríguez, J. P., Achleitner, S., Möderl, M., Rauch, W., Maksimović, Č., McIntyre, N., Díaz-Granados, M. A. & Rodríguez, M. S. 2009 Sediment and pollutant load modelling using an integrated urban drainage modelling toolbox: An application of City Drain. *8th International Conference on Urban Drainage Modelling jointly with the 2nd International Conference on Rainwater Harvesting and Management*. Tokyo, Japan.

Rossman, L. A. 2008 Storm Water Management Model–User's Manual Version 5.0. National Risk Management Research Laboratory–U.S. Environmental Protection Agency, Cincinnati, Ohio.

Sartor, J. D. & Boyd, G. B. 1972 Water pollution aspects of street surface contaminants. *Rep. EPA-R2-72-081*. Washington, D.C., U.S. Environ. Prot. Agency.

Sartor, J. D., Boyd, G. B. & Agardy, F. J. 1974 Water pollution aspects of street surface contaminants. *J. Water Pollut. Control Fed.* **46**(3), 458–467.

Sriananthakumar, K. & Codner, G. P. 1992 *Factors affecting pollutant washoff decay coefficients in urban catchments*. International Symposium on Urban Stormwater Management, Sydney, Australia.

Taylor, G. D. 2006 *Improved Effectiveness of Nitrogen Removal in Constructed Stormwater Wetlands*. Civil Engineering Department. Monash University, Australia.

Tong, S. T. Y. & Chen, W. 2002 Modeling the relationship between land use and surface water quality. *J. Environ. Manage.* **66**(4), 377–393.

USACE, U.C.E. 1977 Storage, treatment, overflow, runoff model (STORM): user's manual (No. Rep. NTIS AD-A955-817): Hydrologic engineers centre, Davis, Calif.

Vaze, J. & Chiew, F. H. S. 2003 Comparative evaluation of urban storm water quality models. *Water Resour. Res.* **39**(10), 1280.

XP-SOFTWARE 1999 XP-AQUALM–Technical Description, XP Software, Australia.

5.4 Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data

Environmental Modelling & Software 26 (2011) 1225–1239

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data

C.B.S. Dotto [a,*], M. Kleidorfer [b], A. Deletic [a], W. Rauch [b], D.T. McCarthy [a], T.D. Fletcher [a]

[a] Centre for Water Sensitive Cities, Department of Civil Engineering and eWater CRC, Monash University, Victoria 3800, Australia
[b] Unit of Environmental Engineering, Faculty of Civil Engineering, University of Innsbruck, Technikerstrasse 13, A-6020 Innsbruck, Austria

A R T I C L E I N F O

Article history:
Received 1 February 2010
Received in revised form
25 March 2011
Accepted 30 March 2011
Available online 6 May 2011

Keywords:
Rainfall/runoff model
Water quality model
Calibration
Sensitivity analysis
Monte Carlo Markov chain

A B S T R A C T

Stormwater models are important tools in the design and management of urban drainage systems. Understanding the sources of uncertainty in these models and their consequences on the model outputs is essential so that subsequent decisions are based on reliable information. Model calibration and sensitivity analysis of such models are critical to evaluate model performance. The aim of this paper is to present the performance and parameter sensitivity of stormwater models with different levels of complexities, using the formal Bayesian approach. The rather complex MUSIC and simple KAREN models were compared in terms of predicting catchment runoff, while an empirical regression model was compared to a process-based build-up/wash-off model for stormwater pollutant prediction. A large dataset was collected at five catchments of different land-uses in Melbourne, Australia. In general, results suggested that, once calibrated, the rainfall/runoff models performed similarly and were both able to reproduce the measured data. It was found that the effective impervious fraction is the most important parameter in both models while both were insensitive to dry weather related parameters. The tested water quality models poorly represented the observed data, and both resulted in high levels of parameter uncertainty.

1. Introduction and background

Stormwater models are important tools in stormwater management practice, being used to predict flow rates and water quality of discharges from urbanised areas. Stormwater flow models are currently well developed and widely adopted in practice. They range from simple models such as KAREN (Rauch and Kinzel, 2007) or CityDrain (Achleitner et al., 2007) that take into account only runoff from impervious surfaces, to very complex urban drainage models. Examples of the latter are MOUSE (DHI, 2002), Infoworks (Wallingford Software, 2009) or CANOE (INSA/SOGREAH, 1999) that model both complex surface runoff and channel/pipe flow, including transition from unpressurised to pressurised pipe flow. There are models that fall in between these two groups, such as the Australian tool MUSIC – Model for Urban Stormwater Improvement Conceptualisation (CRCCH, 2005) or HSPF – Hydrologic Simulation Program Fortran (Bicknell et al., 2001). These are distributed catchment models that include runoff generation from impervious and pervious surfaces and simulate simplified channel/pipe flow. However, all the runoff models, no matter how complex, contain calibration parameters that need to be determined for each specific catchment. It is recognised that the uncertainty in these parameters is one of the sources of error in the model's outputs (e.g. Lindenschmidt, 2006; Gallagher and Doherty, 2007b).

Contrary to the water quantity models, reliable stormwater pollution generation models are almost nonexistent (Elliott and Trowsdale, 2007). For example, conceptual build-up and wash-off models (mainly based on work of Sartor and Boyd, 1972) or simple regression equations (as used in SWMM; (Rossman, 2008)), seem unable to accurately reproduce pollutant behaviour (Beck, 1987; Kanso et al., 2005b; Egodawatta et al., 2007). Simple statistical models, such as investigated by Cohn et al. (1992) and Thierfelder (1999), cannot be used outside catchments for which they are developed. It has been recognised that one of the main problems in establishment of these models is the high level of uncertainty in their calibrated parameters (Kanso et al., 2005a).

Similarly to other modelling practices, understanding the uncertainties associated with the stormwater model parameters is crucial for advancing urban drainage modelling practice. The parameter uncertainty may result from: (1) a poor fit between

* Corresponding author. Department of Civil Engineering, Building 60, Room 153,
[redacted]

1364-8152/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2011.03.013

model outcomes and measured data (Yapo et al., 1996), (2) a high level of parameter correlation (Lindenschmidt, 2006), and/or (3) the insensitivity or identifiability problem due to certain parameters (Afonso and da Conceição Cunha, 2002). A comprehensive investigation on these sources is available in Deletic et al. (2009). Many methodologies are available to evaluate the model sensitivity while quantifying the uncertainty associated with the parameters. The most popular approaches are (a) the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992), known also as pseudo-Bayesian or informal Bayesian (Freni et al., 2009), and (b) the formal Bayesian methods, such as Monte Carlo Markov Chain (MCMC) methods (Kuczera and Parent, 1998).

GLUE has largely been applied to uncertainty assessment in general hydrological models (e.g. Montanari, 2005). The principle of GLUE is to generate parameters samples from a uniform distribution in order to provide a scan of the parameters' space. The method requires a large number of Monte Carlo simulations, while the criteria for accepting a parameter set is subjective and is defined by the user. Bayesian inference based on MCMC methods express the uncertainties associated with parameters and model outputs in terms of probability. Samples are generated from the Markov Chains, which will converge to the posterior distribution of the parameters. One of the most used MCMC methods is the Metropolis-Hasting algorithm (Hastings, 1970), which uses an adaptive proposal distribution to sample parameters and is thus better at finding the high posterior density region. Its effectiveness is now well established (e.g. Bates and Campbell, 2001). For example, a comparison of the GLUE and MCMC methods in regards to a simplified crop model with 22 parameters was demonstrated by Makowski et al. (2002). Both methods presented similar results, but the authors recommend the use of the Metropolis-Hasting algorithm. This method will converge to the true posterior distribution even if the model includes a large number of parameters, while in this case GLUE struggles due to a large number simulation runs required for discretisation of the parameter space. The fact that GLUE has a user-defined threshold for "accepting" parameter sets is also problematic and Freni et al. (2008) found that their results were very sensitive to the choice of this threshold.

MCMC procedures also have problems and a misspecification of the error structure (or likelihood function) in the Bayesian approach can lead to an erroneous quantification of the model prediction uncertainty (Beven et al., 2008). However, In a comprehensive analysis of the nature of structural noise Doherty and Welter (2010) demonstrated that model-to-measurement misfit in most real-world modelling circumstances will possess statistical properties of unknown type, with unknown amounts of serial correlation. This, unfortunately, is an outcome of the imperfect nature of any simulation model, and the fact that these imperfections are likely to introduce structural noise with a singular covariance matrix. They discuss the fact that this places some limitations on the degree to which model predictive uncertainty and confidence intervals can be subject to quantitative evaluation, with all such evaluation assuming a necessarily heuristic component. They also argue that the choice of objective function, rather than the exact nature of the statistical characterisation of model-to-measurement misfit, becomes an issue of critical importance in the model calibration process. In particular, the choice of an objective function that "tunes" a model to make predictions of a certain type can lower the uncertainty associated with predictions of that type. On the other hand, choice of an objective function that endows the model with an ability to make all types of predictions equally well is a luxury that is not available when dealing with imperfect models.

A number of studies used the above approaches to inspect parameter uncertainties in particular urban drainage models (e.g. Gallagher and Doherty, 2007a; Freni et al., 2009). However, very few studies have been published which compare the performance and parameter uncertainty of different quantity and quality models using long-term high resolution data. To fill this gap, this paper presents results on model sensitivity and parameter uncertainty of stormwater models with different levels of complexities, using a Bayesian approach. A rather complex model, MUSIC (CRCCH, 2005) and a simple model, KAREN (Rauch and Kinzel, 2007) were compared in terms of catchment runoff. It is expected to provide information not only on the model parameters but also on model choice and applicability. An empirical regression model was compared to a process-based build-up and wash-off model for stormwater pollutants. Although these approaches were previously proven as unsuitable, derivations of their equations have been adopted in several stormwater models, such as XP-AQUALM (XP-SOFTWARE, 1999) and SWMM (Rossman, 2008). Therefore, a detailed investigation of these models, in terms of processes and parameters, provides key information for the development of more suitable models. A large dataset, collected at five catchments of different land-uses in Melbourne, Australia, was used for the analysis.

2. Methodology

2.1. Rainfall runoff models

2.1.1. MUSIC rainfall/runoff model

For a given rainfall and evapotranspiration time series, MUSIC continuously simulates, at a user-specified timestep (from 6 min to 24 h) catchment discharges. The algorithm is based on the SimHyd model developed by Chiew and McMahon (1997), which was modified to enable disaggregation of daily runoff into sub-daily temporal patterns.

As shown in Fig. 1(a), flow from impervious and pervious areas are modelled separately. There are two parameters related to impervious area runoff: (1) the Effective Impervious Fraction (*EIF*) which corresponds to the areas that are directly connected to the drainage system, and (2) the initial loss, known as the rainfall threshold (*Thres*), which defines the minimum daily rainfall before surface runoff occurs from the impervious area. Runoff from the pervious areas occurs when the pervious soil storage is either saturated or its infiltration rate is exceeded. The pervious area runoff related parameters are mainly described by: (1) soil storage capacities, the maximum soil storage (*SMax*) and the initial storage level (*SIni*), (2) field capacity (*fc*), and (3) infiltration factors (*coeff* and *sq*). Water from the soil store is lost due to actual evapotranspiration, which is a function of the current day's potential evapotranspiration and the ratio of water currently in the pervious store to its capacity. Groundwater is modelled as a store which is recharged when the level in the pervious soil store exceeds the field capacity. The rate of this recharge (*rfac*) is a set percentage of the water in the soil store. This store is emptied via baseflow, which is modelled as a percentage of the water within the groundwater store, also a model parameter, *bfac*. In a similar way, deep seepage (*dseep*) is set as a percentage of the groundwater store. The Muskingum Cunge routing method is applied for the routing of flows through the drainage system; the translation (K) and the attenuation (θ) both require calibration. All the model's parameters are summarised in Table 1 and a full description is available in the MUSIC manual (CRCCH, 2005). All parameters described above require calibration (**13** in total).

2.1.2. KAREN

A simple linear reservoir model KAREN (Rauch and Kinzel, 2007) was adopted, due to its simple modular design. The model requires the catchment area and a rainfall time series as inputs to

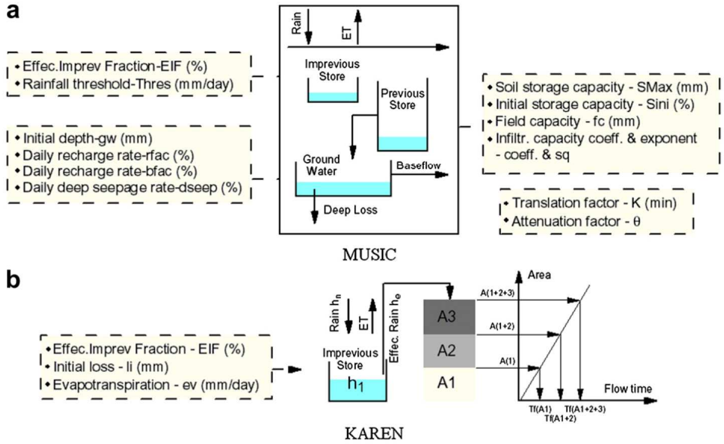


Fig. 1. MUSIC rainfall/runoff model (after CRCCH, 2005) (a), and KAREN (b).

generate a series of flows originating from impervious areas only. The pervious components of the catchments are not considered.

A schematic presentation of the rainfall/runoff model implemented in KAREN is given in Fig. 1(b). The effective impervious area of the catchment is calibrated as the *EIF* parameter. Runoff from impervious areas occurs after a rainfall threshold has been exceeded. This threshold is represented by the initial loss parameter (*li*) and it is modelled as a single reservoir, similar to that found in MUSIC. While MUSIC's impervious runoff threshold is calculated and reset on a daily basis, KAREN's initial loss is calculated continuously and fills during rainfall and is drained during dry weather by a permanent loss calibration parameter (*ev*). Surface runoff volume is calculated using the linear time-area method, which is similar to the unit hydrograph method (Sherman, 1932). At the beginning of a rainfall event, the effective impervious area is increased according to the flowtime on the catchment surface until the whole catchment contributes to runoff after time of concentration (another calibration parameter *TC*). The model's parameters are summarised in Table 1 and a detailed description is provided in the manual (Rauch and Kinzel, 2007).

2.2. Water quality models

A process-based build-up/wash-off model was compared to an empirical regression model for stormwater pollutants. Both models predict pollutant concentrations at the outlet of the catchment and none of them consider the transport processes in the catchment and drainage network.

2.2.1. Build-up/wash-off model

The generation of pollutants in the runoff from an impervious surface is often described and modelled using the concepts of build-up and wash-off. The attempt to model both of these processes was proposed by Sartor and Boyd (1972) and is summarised in Sartor et al. (1974). The key equations are:

Table 1
Rainfall/runoff model parameters.

Component	Parameter name	Unit	Prior distribution	Default value
MUSIC Parameters				
Impervious	Effective impervious fraction (*EIF*)	%	U[0 1]×100	–
Area	Rainfall threshold (*Thres*)	mm	U[0 3]	1.0
Pervious Area	Soil storage capacity (*SMax*)	mm	U[30 250]	30[a]
	Initial storage (*SIni*)	%	U[0 50]	30
	Field capacity (*fc*)	mm	U[0 100]	20[a]
	Infiltration capacity coefficient (*coeff*)		U[100 200]	200
	Infiltration capacity exponent (*sq*)	–	U[0.9 1.2]	1
Groundwater	Daily recharge rate (*rfac*)	%	U[0 1]×100	25%
	Daily baseflow rate (*bfac*)	%	U[0 1]×100	5%
	Groundwater initial storage (*gw*)	mm	U[0 30]	10
	Daily deep seepage rate (*dseep*)	%	U[0 0.5]×100	0
Muskingum	Translation and factor (*K*)	min	U[0 50]	30
Cunge	Attenuations factor (θ)	–	U[0.1 0.49]	0.25
KAREN Parameters				
Impervious	Effective impervious fraction (*EIF*)	%	U[0 1]×100	–
Area	Time of concentration(*TC*)	min	U[1 10,000]	–
	Initial loss (*li*)	mm	U[0 10]	1
	Evapotranspiration (*ev*)	mm/day	U[0 10]	1.5

[a] MUSIC default values for Melbourne.

Table 2
Water quality model parameters.

Component	Parameter name	Unit	Prior distribution
Build-up – washoff model			
Build-up	Maximum amount of pollutant M_0	g/m^2	U[0 100]
	Accumulation constant k_1	day^{-1}	U[0 100]
Washoff	Washoff coefficient k_2	–	U[0 100]
	Washoff exponent k_3	–	U[0 100]
Regression model			
Water quality scale coefficient (a)		$(g\cdot s^b)/m^{3(1\ b)}$	U[0 150]
Water quality shape coefficient (b)			U[0 2.5]

$$\text{Build–up}: \frac{d\overline{M(t)}}{dt_d} = K_1\cdot(M_0 - \overline{M(t)}) \tag{1}$$

$$\text{Wash – off}: \frac{d\overline{C}(t+tf)}{dt} - k_2\cdot\overline{M(t)}\cdot I(t)^{k_3}\cdot A_i \tag{2}$$

where $\overline{M}$ is the amount of pollutant available on the surface averaged over the area (g/m^2), $\overline{C}$ is the concentration in runoff (mg/l), I is the rainfall intensity (mm/hour over the timestep), and A_i is the impervious area (m^2). The calculated concentration is shifted by the flowtime tf, which is the average flowtime of concentration on catchment surfaces and in sewer pipes.

In this study tf was regarded as a constant which was not varied during Bayesian inference. Prior simulation runs showed that a variation of tf disturbs generation of posterior distributions of the other calibration parameters and the MCMC procedure does not converge. This means tf is not only a highly sensitive parameter but reasonable simulation results can only be achieved by fixing tf to certain value, which was found during calibration prior to Bayesian inference. The authors expect that this is caused by the timestep discretisation of the model. The pollutograph of model output can only be shifted by whole timestep of 6 min while intermediate values have no effect model results.

The build-up calibration parameters are M_0, which is the maximum amount of pollutant expected on the surface (g/m^2) and k_1, which is an accumulation constant (day^{-1}). According to Equation (2), the calibration parameter k_2 is the wash-off coefficient and k_3 is the wash-off exponent. In total, there are 4 calibration parameters (Table 2).

2.2.2. Regression model

The simple regression model adopted in this study was used in conjunction observed runoff data. Derivations of this regression model are used in practice in several stormwater models, such as XP-AQUALM (XP-SOFTWARE, 1999), SWMM5 (Rossman, 2008) and P8-UCM (Palmstrom and Walker, 1990). The model estimates concentrations, within a timestep, as a power function of the measured runoff:

$$C_t = aR_t^b \tag{3}$$

where C_t is the pollutant concentration at time t (mg/L); R_t is the measured runoff (average, in mm/hr, over the timestep, t); a and b are calibration parameters as outlined in Table 2.

2.3. Data set

A comprehensive stormwater dataset, containing data on stormwater flows and pollution concentrations from 5 urban catchments located in Melbourne, Australia (Fig. 2), was used for the analysis. Table 3 shows a summary of the characteristics of the catchments and some descriptive statistics of the measured data (McCarthy, 2008; Francey et al., 2010).

Rainfall data was monitored every minute using 0.2 mm tipping bucket rain gauges located as close as possible to the catchments' centroid. The mean annual rainfall in these catchments ranges from 370 to over 720 mm per year. Continuous flow data (recorded at the 1 min interval) were measured at each catchment and around 300 wet weather events were monitored for a range of pollutants, including Total Suspended Solids (TSS) and Total Nitrogen (TN) (except for RD catchment, where only TSS was monitored). The water quality samples were collected at the outlet of the catchments using a discrete sampling methodology and the variability of TSS and TN concentrations between sites was quite large as shown by their coefficient of variation (CV) (Table 3). In terms of pollutants, these large coefficient values also indicate the log nature of these pollutants. The average Event Mean Concentration (EMC) for TSS ranged from 72 to 125 mg/L between sites, and for TN this was between 1.17 and 3.51 mg/L.

Most of the sites were monitored from January 2004 to December 2007. Rain and flow data collected between 2004 and 2005 was used for calibration, while data from 2006 to 2007 was used for validation. Table 3 reports on the characteristics of events used for model calibration (general figures), and also on the events used or validation (between square brackets). The calibration of the water quality models resulted in very low performance during calibration and therefore, validation of these models would not make sense and was not carried out.

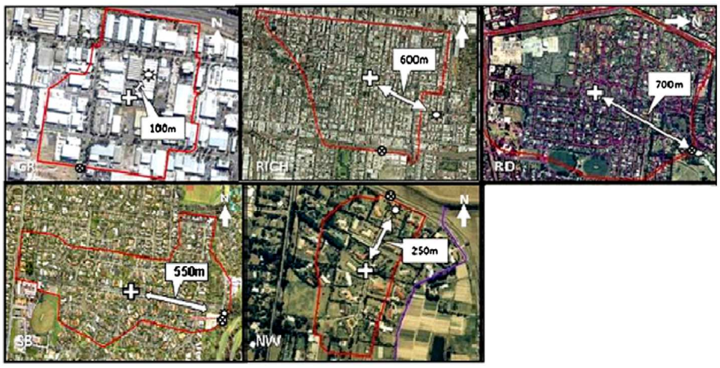


Fig. 2. Aerial photographs of each catchment showing the relative positions of the centroid, rainfall gauge and flow gauges (☼ rainfall gauge location, ✚ catchment centroid and ⊗ flow gauge location) (after McCarthy, 2008).

Table 3
Summary of sites and measured data details (Francey et al., 2010). The characteristics the events are presented while the characteristics of events used for validation are given in brackets [ ].

Site	Gilby Rd (GR)	Richmond (RICH)	Ruffeys Lake, Doncaster (RD)	Shepherds Bush (SB)	Narre Warren (NW)
Primary Land Use	Commercial	High Density Residential	Medium Density Residential	Medium Density Residential	Rural Residential
Area (ha)	28.2	89.1	105.6	38	10.5
TIF[a]	0.8	0.74	0.51	0.45	0.2
Catchment Average slope (%)	1	3.5	5		4
Time of concentration (min)	23	31	14	14	16
Distance from catch centroid to rain gauge (m)	100	600	700	550	250
Mean annual rainfall (mm/year)	723 [536]	650 [500]	650 [370]	580	700 [400]
Mean event maximum rainfall intensity (mm/hr)	10 [7.5]	10 [8.4]	9 [7]	6	10 [8]
Range of event maximum rainfall (mm/hr)	2–86 [2–36]	2–60 [2–44]	2–44 [2–28]	2–60	2.5–86.3 [2–32]
Mean event maximum runoff rate (L/s)	408 [50]	547 [212]	723 [165]	214	44 [20]
Range of event maximum flow rates (L/s)	75–2241 [30–200]	67–3867 [25–1430]	164–3069 [20–908]	29–1200	14–90 [10–58]
N. of events – TSS	49	40	54	19	41
Maximum TSS concentration (mg/L)	867	1600	1422	1545	2398
TSS CV[b] (%)	151.46	164.05	183.12	153.54	182.87
Mean of TSS EMC's[c] (mg/L)	71.6	125.1	77.0	94.8	91.9
N. of events – TN	47	39	–	17	18
Maximum TN concentration (mg/L)	9	26	–	15	19
TN CV[b] (%)	83.18	101.32		85.22	76.82
Mean of TN EMC's[c] (mg/L)	1.17	2.29	–	1.74	3.51

[a] Total impervious fraction (*TIF*).
[b] Coefficient of variation (CV).
[c] Event Mean Concentration (EMC).

Rainfall data used in this study was processed to cope with gaps and time drifts, which are inherent in any battery controlled logging device. Flow data was checked for any discrepancy (e.g. backflow effects indicated by negative velocities).

2.4. Calibration, parameter sensitivity analysis and validation

Calibration and sensitivity analysis were undertaken with the software MICA (Doherty, 2003). A least squares objective function was used for model calibration and sensitivity testing, which is based on the sum of the squared deviations and is the most adopted in the field (e.g. Feyen et al., 2007; Freni et al., 2009). Further, for evaluating calibration performance (i.e. comparing the measured and modelled data points) the Nash-Sutcliffe efficiency coefficient (E) (Nash and Sutcliffe, 1970) was used. Both the least squares objective function and E place emphasis on medium/large values, which are often the goals for stormwater management practices (i.e. high volumes (e.g. Chiew and McMahon, 1999)).

MICA quantifies a parameter's uncertainty following a Bayesian approach, in which the prior information about the parameters is updated to generate the Posterior Distribution (PD). It undertakes an MCMC analysis with the Metropolis-Hastings algorithm (Hastings, 1970). A full description of the MICA application is available in the software manual (Doherty, 2003), and is only briefly outlined below. Firstly, the model is run with the initial parameter values, which are sampled from a uniform initial prior distribution. The estimated values are compared with the observed data, and the likelihood of such parameters is calculated. Subsequently, the model is run with a new set of parameters derived from the proposal distributions and the likelihood of the new set of parameters is calculated. The Bayes' Theorem is used to calculate the PD and parameters are accepted or rejected according to the acceptance ratio. The ratio is calculated as a function of three ratios representing the model fit, the prior and the proposal densities. The process is repeated until the $\sqrt{R}$ criterion of Gelman et al. (1995) is reached: if the scale reduction score was less than 1.05, the Markov Chain was considered to be converged (Doherty, 2003). However, the process would not terminate before a minimum of 5000 iterations was achieved, which was set to guarantee that the extremities of the parameter PD are also sampled. The number of simultaneous Markov chains and the maximum number of iterations were chosen after some initial testing. Ten Markov chains, with a maximum of 10,000 iterations each, were sufficient to generate the parameter PDs for the rainfall/runoff model (i.e. 15 or 20 chains with 20,000 iterations did not improve the outcomes). The water quality models required more exploration of the parameter space and therefore, these models were run using 20 Markov chains, each with a maximum of 10,000 iterations. The parameter sets were updated for each chain and thus, each iteration required a number of runs equal to the number of chains. The proposal probability functions are updated by means of the statistics calculated from parameter samples taken up to the moment. A "burn-in" period was excluded from any analysis to ensure that the remaining values were true samples from the parameter PD. The number of iterations which was classified as the "burn-in" period was computed according to the statistics recorded by MICA as it implements the MCMC process. For instance the stabilization of means and standard deviations for all monitored parameters is a sign that the MCMC "burn-in" period may be over.

The likelihood function used in MICA assumes that the residuals between the measured and modelled values have a normal distribution (e.g.Gilks et al., 1996). This is a trait of many uncertainty/sensitivity procedures (see Bates and Campbell, 2001; Yang et al., 2008) and not checking such an assumption is common in the literature (e.g. Makowski et al., 2002; Varella et al., 2010). In cases when the assumptions are checked, they are usually not met (e.g. Feyen et al., 2007; Thyer et al., 2007). In this paper, the normality of the residuals was checked using normal probability plots and statistical inferences (Chakravarti and Roy, 1967). For one catchment (RICH), a Box-Cox (Box and Cox, 1964) transformation was applied to the measured and modelled flow rates (from KAREN), in order to meet the normality assumption of the residuals. The MCMC procedure was repeated using the transformed datasets, and PDs for each parameter were generated and compared to that found using the untransformed data.

Using the parameter PDs it was possible to verify the model sensitivity to each of its parameters according to the shape of their distributions. Whilst in theory in a well-posed model the PDs are expected to be unimodal, this will not be the case in poorly-posed models. A PD with a flat shape (uniform distributed parameter) shows the insensitivity of the model to this parameter since the model achieves similar accuracy for any value of the parameter. In contrast, if the model parameter PD has a well defined 'peak', the model is very sensitive to that parameter. Multiple local optima can be identified by multiple peaks in the PDs and are likely to occur in 'ill-posed' models representing some parameter interaction and/or compensation.

Initial calibration results showed that KAREN and MUSIC produced very similar model efficiencies, even though MUSIC explicitly represents pervious area runoff. Also, these initial results showed that impervious parameters were more sensitive than pervious parameters. To investigate these findings, individual event calibration was conducted on the five largest events where pervious runoff was known to occur (these events were taken from RICH and NW). This was used to determine whether MUSIC's pervious parameters were significantly contributing to the prediction during pervious runoff events. This method also showed whether KAREN could cope in these events without pervious area representation.

While the parameter distributions provide an indication of their sensitivity, an apparent lack of sensitivity could also be caused by parameter interaction and parameter identifiability issues. As such, each non-sensitive parameter was varied, one at a time, by sampling from its generated PD, while keeping all other parameters fixed at their optimised values. Model efficiency values from this procedure were then compared to that of the fully calibrated model to determine whether this parameter is truly non-sensitive or whether, for example, parameter interaction is occurring.

The rainfall/runoff models were validated with the second half of the datasets, in most cases including between 14 and 23 months; and the model predictions were obtained by running the model with the parameter values randomly sampled from the PDs obtained during calibration.

2.5. Modelling uncertainties

2.5.1. Parameter uncertainty

For both calibration and validation, the predictive uncertainty resulting from parameter uncertainty was obtained by running the models with 1500 parameter sets, randomly sampled from the generated parameter PDs. It is emphasised that the predictive uncertainty in this case originates only from the parameter uncertainty and does not include other sources of uncertainty. Currently, the procedures to include these other sources are not well developed, especially when dealing with those which originate from the model structure (Doherty and Welter, 2010).

2.5.2. Total uncertainty

A common approach for estimating the total uncertainty (i.e. parameter and all other sources) is adding a constant Gaussian error term to the model predictions (e.g. Feyen et al., 2007). However, this method does not take into account the serial correlation which exists between data points (e.g. Yang et al., 2008) and should only be applied when the residuals are normally distributed (which was not the case for most of these simulations). To overcome these issues, the total predictive uncertainty was calculated as follows.

1. The residuals between the measured and the simulated values were calculated.
2. The simulated values and their corresponding residuals were grouped in different bins according to the simulated values' magnitudes
3. The 2.5 and 97.5 percentiles of the residuals in each bin were plotted against their corresponding simulated values, and two linear relationships were verified (one for the 2.5 and one for the 97.5 percentile).
4. The error terms were calculated as functions of the simulated values and these error terms were added to the simulated values at each timestep.
5. This process was repeated for the 1500 parameter sets, randomly sampled from the generated parameter PDs. The 2.5 and the 97.5 percentiles of all 1500 simulations with the added error terms was obtained and used to represent the total uncertainty bounds.

3. Results and discussion

3.1. Rainfall/runoff modelling

3.1.1. Model performance – calibration

The overall efficiency of both rainfall/runoff models assessed in this study is represented by the Nash – Sutcliffe efficiency criterion (E) in Table 4. The models were run with the best parameter sets and the time series of the modelled outflows were plotted against the corresponding measured data (Fig. 3 shows this for three catchments). Coefficients of determination (R^2) and the best efficiencies between measured and modelled, E_{best}, are shown in these graphs.

Considering the fine temporal resolution used (6 min) and that all measured data have some uncertainty (see McCarthy et al., 2008), the runoff models calibrated well (all E_{best} values were either very close to, or over 0.5). It is hypothesised that the relatively low value ($E_{best} = 0.49$) achieved at NW with MUSIC may be due to the complexity of the processes in this catchment, which presents the lowest level of impervious area and potential extraneous wastewater inputs. There was also not much difference between the two models; calibration was performed with very similar efficiency for all catchments with MUSIC and KAREN.

Table 4
Rainfall/runoff models' efficiency: E_{best} is the best E obtained during the entire calibration/validation procedure; E_{mean} is the mean E derived from all the accepted parameter sets.

Catchment	TIF	MUSIC				KAREN			
		Calibration		Validation		Calibration		Validation	
		E_{best}	E_{mean}	E_{best}	E_{mean}	E_{best}	E_{mean}	E_{best}	E_{mean}
Gilby Rd (GR)	0.80	0.54	0.53	0.31	−0.02	0.53	0.26	0.41	0.04
Richmond (RICH)	0.74	0.81	0.75	0.70	0.60	0.75	0.71	0.71	0.61
Ruffeys Lake, Doncaster (RD)	0.51	0.62	0.55	0.32	−0.06	0.63	0.28	0.39	−0.56
Shepherds Bush (SB)	0.45	0.57	0.61	–	–	0.61	0.52	–	–
Narre Warren (NW)	0.20	0.49	0.60	−0.05	−0.69	0.60	0.12	−1.01	−5.19
Mean (E)		0.61	0.62	0.33	−0.04	0.62	0.38	0.13	−1.28

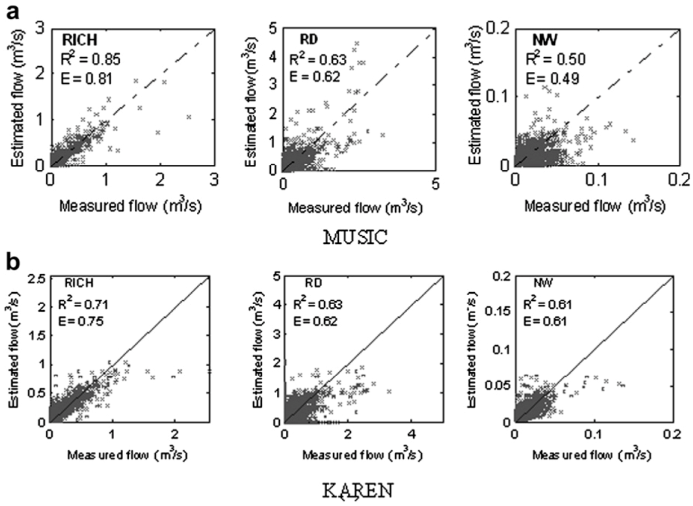


Fig. 3. Measured versus calibrated modelled flows: (a) MUSIC; (b) KAREN. R^2 is the correlation between measured and modelled, and E_{best} is the best model efficiency achieved.

However, for the relatively highly impervious GR and RICH catchments, the MUSIC model produced more accurate results than the KAREN model, whilst for the relatively more pervious RD, SB and NW it was the opposite. This is an interesting finding, considering that MUSIC has the ability to represent pervious areas. It was therefore concluded that the simple KAREN model, which does not simulate processes in pervious surfaces, could still be successfully calibrated for any urban development, since flow generation is mainly governed by impervious surfaces.

The above finding is further confirmed by the results from the single events calibration procedure. The model efficiency obtained for large events, which are known to include runoff from pervious surfaces, was similar for both models ($E_{KAREN} = 0.93$ and $E_{MUSIC} = 0.95$; Fig. 4). As expected, MUSIC was able to predict flow peaks during this event more accurately than KAREN. While this could be due to MUSIC's modelling of pervious surfaces, it is also possible that this is caused by the over-parameterisation of the MUSIC model for this event (i.e. the model is not representing reality and is instead behaving like a black-box). It is hypothesised that the reason why MUSIC's pervious area parameters were not contributing as much as others is due to a combination of factors: (1) model structural noise, (2) measured data errors and (3) selection of objective function.

3.1.2. Parameter sensitivity

The histograms (PDs) of the selected MUSIC parameters are presented in Fig. 5 and for KAREN parameters in Fig. 6. It can be noticed that some of the parameter distributions were non-normal, and that not all of the parameters can be fitted to standard distributions.

3.1.2.1. MUSIC. The effective impervious fraction (*EIF*) and the Muskingum Cunge translation factor (*K*) were the most sensitive for all catchments, since their PDs had clear peaks.

The PDs obtained for the groundwater initial depth (*gw*) were very flat (as shown in Fig. 5(e)) and any value between the specific range may be assumed without compromising the results. MUSIC presented the same pattern for two more parameters: initial storage level in the previous area storage (*SIni*) and to the infiltration capacity exponent (*sq*). Therefore these three parameters can be fixed to any value from the studied range (Table 1) without impacting on the model performance. It is suggested that the model is over-parameterised and therefore the dimensionality of the model could perhaps be reduced.

The rainfall threshold (*Thres*) parameter had a peculiar behaviour. It was suggested that any value higher than 1 was linked to the worst optimisation results and that a value between 0 and 1

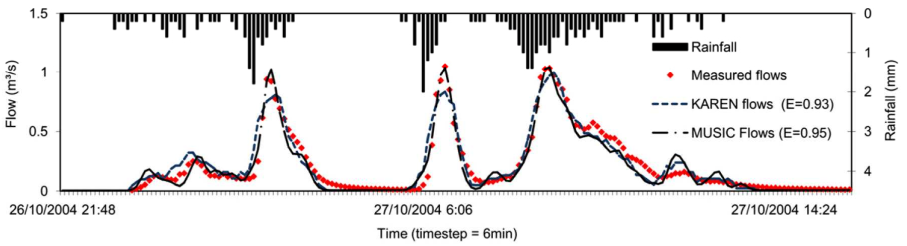


Fig. 4. Richmond example hydrograph with measured versus modelled flows for the single event calibration.

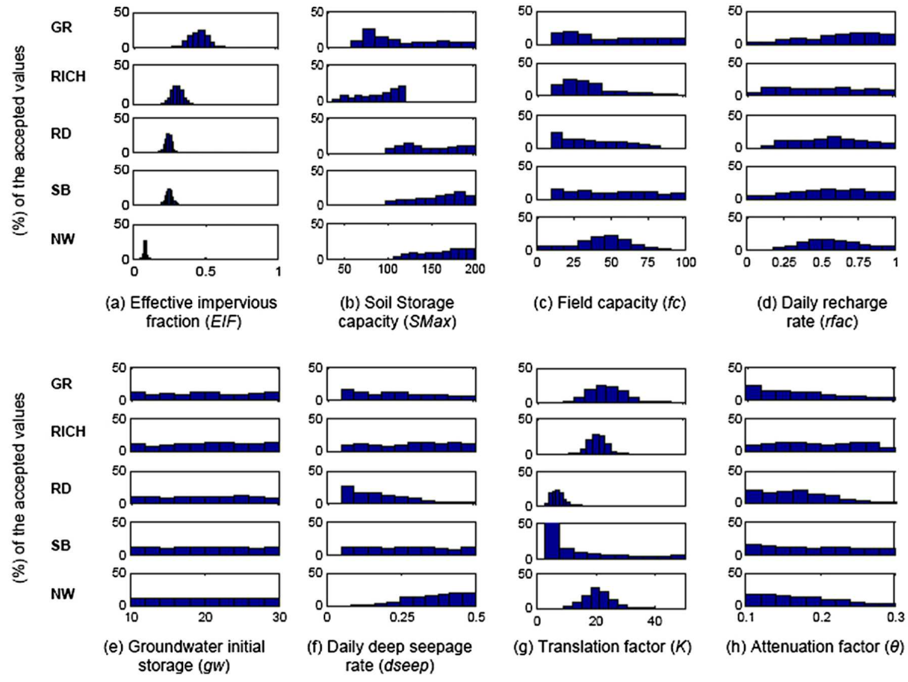


Fig. 5. Histograms for selected MUSIC parameters (units according to Table 1).

generally leads to similar reasonable results for all catchments. It is suggested that the default value of 1 is a practical option for the catchments and dataset used in this study.

The model seemed insensitive to the Muskingum Cunge attenuation factor (θ) for most catchments (Fig. 5 (h)). To check this, this parameter was varied while fixing all other parameters at their optimised values. For some catchments (GR, RD, NW), significantly lower E values were generated when different values of θ were replaced in the calibrated model (e.g. E decreased from 0.62 to 0.49 for RD), indicating possible parameter interaction. However, for other catchments (RICH, SB) changing θ values did not make major differences in the model outputs, suggesting that this parameter is actually insensitive and any value between 0.1 and 0.3 can be used with equal success.

It is apparent that the field capacity (*fc*) may be fixed at any valued between 10 and 40 mm or even set to its default value (20 mm) for the two most urbanised catchments (GR and RICH). The same parameter did not show a consistent behaviour for the remaining catchments. The model was sensitive to *fc* for RD and NW, but not for SB (Fig. 5(c)). From this sensitivity study, it is recommended that *fc* should be calibrated for catchments with *EIF* lower than 0.3 to predict reliable estimates. The same pattern and recommendation are valid for the daily recharge rate (*rfac*), daily baseflow rate (*bfac*) and for the daily deep seepage rate (*dseep*).

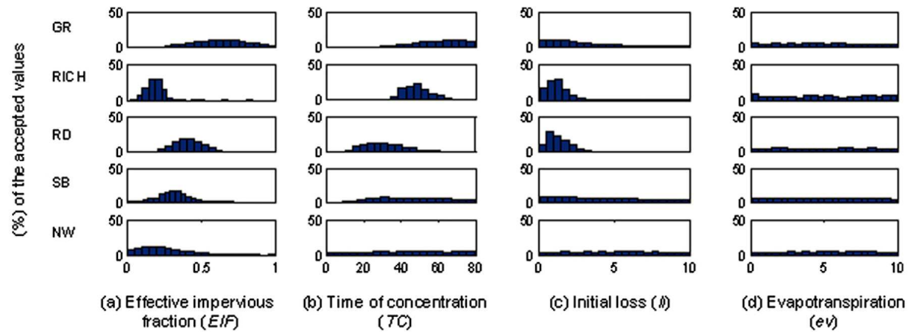


Fig. 6. Histograms for KAREN parameters (units according to Table 1).

According to Fig. 5(b) the soil storage capacity (*SMax*) optimum values increased as the catchments' imperviousness decreased. This could be caused by the fact that a catchment with a higher level of imperviousness typically has a higher population which could lead to more compact soils (foot traffic, etc) and thus lower soil storage capacity (McCarthy, 2008). On the other hand, a previous study developed by Dotto et al. (2009) suggested that this parameter is indeed a "calibration parameter" rather than a soil related property in the case of highly urbanised catchments as the ones used in this study and therefore, parameter compensation is expected at some extend, depending on their range of values.

3.1.2.2. KAREN. Analysis of parameter sensitivity (Fig. 6) of this model shows similar results as MUSIC. The most important calibration parameter is *EIF* followed by the time of concentration (*TC* representing similar processes as *K* in MUSIC) and the initial loss (*li*).

The model was insensitive to the evapotranspiration (*ev*) parameter for all catchments. This parameter only controls the drainage of the initial loss volume during dry weather periods. It follows that this parameter is only important for events which have a similar magnitude as the initial loss volume. For example, a 30 mm rainfall event with an initial loss of 1 mm would not be sensitive to the initial loss (because it is just 3% of the total rainfall) and therefore be insensitive to *ev*. Since the selected objective function emphasises peak flow rates (or events), it is reasonable that the *ev* parameter is not sensitive. Such insensitivity was also tested by varying *ev* within a range from 1 to 10 (mm/d) while keeping the other parameters fixed at their calibrated values. This had only a minor effect on *E* values. For instance, *E* for RICH varied from 0.75 to 0.70.

The time of concentration (*TC*) and the initial loss (*li*) parameters were only relevant in the case of the three most impervious catchments (Fig. 4). A relationship between the model sensitivity to these parameters and the catchment area was also apparent. The model was sensitive to them in the case of the largest catchments (e.g. RICH and RD) and insensitive for the smallest one (NW). This can be explained by the way in which runoff volume is calculated from rainfall according to the time-area method. The proportion of the catchment contributing to runoff continuously increases with each timestep until the whole catchment contributes to runoff once the *TC* is reached. Further discussion on the variability of *TC* and its estimation can be found in (McCarthy, 2008).

The insensitivity of initial loss (*li*) for mainly pervious catchments was not expected as usually losses are higher in pervious parts of a catchment. A possible reason for this effect could be that small values for initial loss represent losses on impervious parts and higher values represent losses on pervious parts and consequently a wider range of initial loss is experienced "on average" of the whole catchment.

No major correlations between calibration parameters could be identified for any of the catchments analysed. The large variation encountered in the histograms for several parameters indicates some difficulty in obtaining clear values for such parameters. It may again be speculated that input and calibration data errors are possible causes. It is suggested that these uncertainties need to be quantified in order to define more reliable outcomes.

3.1.3. Error structure

A few assumptions trigger the application of methods, in which the structure of the model residuals has to be known according to a chosen likelihood measure. According to the methods used in this study, the residuals were assumed to be independent and normally distributed. However, the residuals between untransformed measured and modelled data were found to be non-normal for all four models and for each of the five catchments. None of them passed the Kolmogorov–Smirnov test at 5% significance level. As an example, Fig. 7 (a) shows Richmond's normal probability plot of the residuals between the measured data and the outputs of the optimised KAREN model ((●) – untransformed data was used).

Similarly to that found in the literature (see Feyen et al., 2007; Yang et al., 2008), a Box–Cox transformation of both the measured and modelled flows ensured that the assumption of normality in the residuals was met. Fig. 7(a) shows Richmond's normal probability plot of the residuals between the transformed measured data and outputs of the optimised KAREN (○). Such figure indicates that the transformed residuals nearly match to a normal distribution, which was also verified by the Kolmogorov–Smirnov test at 5% significance level. However, this data transformation also effectively changes the objective function from one which emphasises peak flows (i.e. least squares) to one which considers all parts of the hydrograph similarly. This resulted in a decrease in model efficiency (from $E = 0.75$ to $E = 0.57$) which is logical since *E* is based on a least squares fit, and transforming the data means that the objective function is no longer of a least

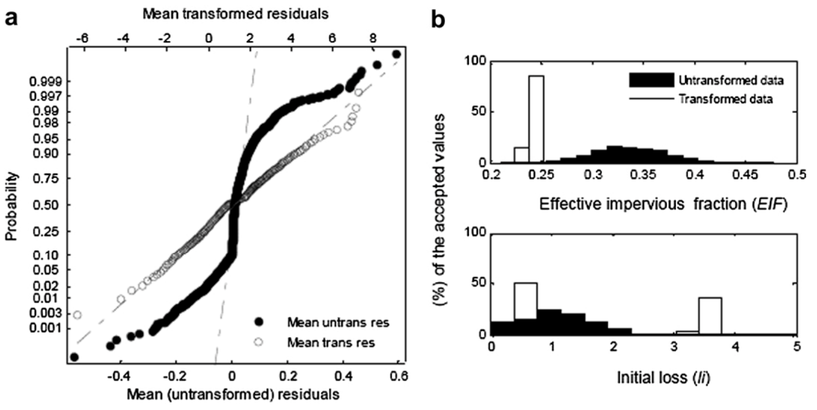


Fig. 7. (a) The normal probability plot for the mean untransformed (●) and transformed residuals (○) obtained from MICA runs for KAREN with RICH data, and (b) example of KAREN sensitivity when MICA was run with the Box–Cox transformed data for RICH catchment.

squares type. Model PDs also changed significantly (Fig. 7(b)), showing the importance of objective function on parameter sensitivity and optimised parameter values.

A change in objective function is not desirable, especially when it has been carefully chosen for the specified modelling purpose, as was done in this study. For the procedure outlined here, it is impossible to obtain normality in the residuals when trying to maintain the chosen objective function. To solve this, another sensitivity testing procedure, which does not have to meet the normality assumption, could be used (e.g. GLUE by Beven and Binley, 1992). However, alternate sensitivity methods also have other limitations, including subjective cut-off criteria and large simulation requirements (Freni et al., 2008; Doherty and Welter, 2010). In any case, the authors have compared the parameter distributions obtained from GLUE to those obtained in this paper and found no significant differences (data not shown).

3.1.4. *Model performance – validation*

The performance achieved with both models during the validation period is presented in Table 4 and the results indicate that the validation results were not very promising, except for RICH and possibly GR. A plausible explanation for these poor results is that the climatic patterns observed in the calibration period differed to that in the validation period (Table 3). As an extreme example, the mean event maximum runoff rate for the calibration period was 723 L/s at RD, while 165 L/s was registered for the validation period (Melbourne was experiencing a severe drought during this period). In addition, the fact that the models were calibrated with a least squares objective function, which places more emphasis on medium to high values, the models were not calibrated to represent the subsequent low runoff rates observed in the validation period. Different techniques are available to split the data for calibration and validation (Xu, 1999). Xu (1999) emphasised that models should be able to reasonably predict the system's responses from drier to wetter conditions and vice-versa. Therefore, an extreme validation period was chosen to truly test the models in terms of predictive capability in different climatic/hydrological periods.

3.1.5. *Practical outcomes for calibration of rainfall/runoff models*

Both models were highly sensitive to effective impervious area, *EIF*. It is therefore important to examine this parameter in more detail, and derive some suggestions for its reliable assessment (particularly given the difficulty and often subjectivity in its physical assessment).

All optimised values of *EIF* with MUSIC and KAREN were found to be significantly lower than the total impervious area, *TIF* as shown in Fig. 8. Such figure presents the mean and standard deviation values for each catchment, as obtained from the PDs (presented in Figs. 5 and 6). Optimised KAREN *EIF* values exceeded those obtained from MUSIC with an average ratio of 1.5 (Fig. 8 far right), probably because KAREN's runoff is restricted only to impervious areas, hence compensation occurs to represent the rare runoff from pervious areas. Moreover, *EIF* is an example of a calibration parameter which not only depends on catchment and rainfall characteristics, but if it is used as a model parameter (whether it may be input or a calibration parameter), its value will also depend on model structure. When comparing the standard deviations for *EIF*-MUSIC with *EIF*-KAREN a clear difference may be recognised. The standard deviation of *EIF* in KAREN simulations is considerably higher. The reason for this might also be the disregard of pervious catchment area in KAREN. This is an example of the contribution to uncertainty introduced by model structure. As in this study, different sources of uncertainties are not regarded explicitly, they are reflected by uncertainties in estimating calibration parameters. Therefore, different model structures can lead to different calibration parameters (although they represent the same physical background) as uncertainties are, to some extent, compensated during calibration.

In the present work, a consistent ratio between *EIF* and *TIF* was suggested for the two models (averages of 0.5 and 0.72 for MUSIC and KAREN, respectively). As seen in Fig. 8, the main outlier in this trend is RICH catchment, which has a *TIF* of 74% and *EIF* around 30%. However, this difference could be explained by the age of the catchment and its associated infrastructure; this catchment was developed primarily in the early 1920s and it is possible that a great portion of roof runoff is not connected to stormwater sewers, or that there are major leaks from the stormwater infrastructure into surrounding soils and groundwater.

Consistent with previous findings in the literature (see Zaghloul, 1983; Zoppou, 2001), it can be concluded, for both models, that the highly sensitive parameter *EIF* should be calibrated whenever possible. If data for model calibration is not available, the use of satellite images to determine the *TIF* and a GIS analysis of the drainage plan are required to assess *EIF*. For a city or region, the regressions such as those shown in Fig. 8 for Melbourne could be developed, for a specific rainfall/runoff model, and then used for ungauged catchments. However, further verification with a larger number of catchments is highly advised.

MUSIC was very sensitive to *K* (the translation factor in the Muskingum Cunge routing method) for all catchments, and KAREN was also sensitive to *TC* (the time of concentration parameter) for most of the catchments. It is recommended that these parameters should be calibrated whenever possible in order to maximize the performance of the flow routing method.

MUSIC was insensitive to some of the dry weather associated parameters for all catchments (*coeff*, *sq*, *SIni* and *gw*). Stormwater models mainly need impervious area related parameters because that is where most of runoff occurs in urban catchments. Similarly,

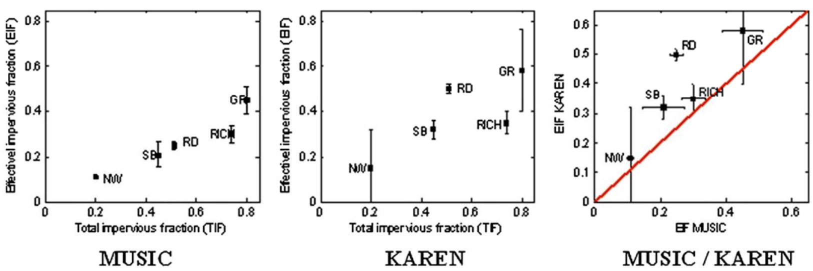


Fig. 8. Correlation between the total impervious fraction (*TIF*) and the calibrated effective impervious fraction (*EIF*).

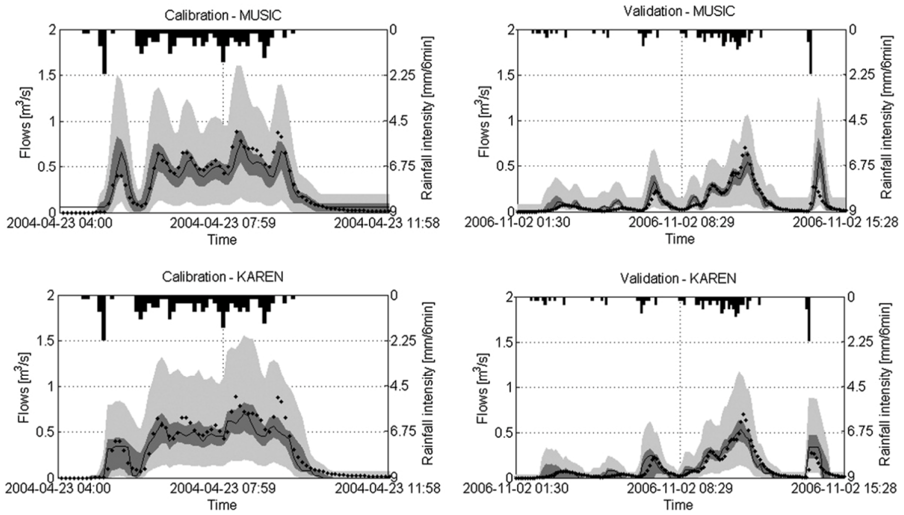


Fig. 9. Prediction uncertainty for sample hydrographs at Richmond site with MUSIC (top) and KAREN (bottom) for both calibration (left) and validation (right). The dark dots represent the observed data while the black line is the modelled data with the optimised parameter set; the dark shaded area shows the predictive uncertainty associated only with parameter uncertainty; and, the light shaded area shows the total predictive uncertainty associated with the total error related to the modelling residuals.

KAREN was insensitive to its evapotranspiration, *ev* parameter. Although this indicates that rainfall events, for these studied sites, could be regarded independently instead of continuous simulation, more complex systems will still demand continuous simulation.

3.1.6. Modelling uncertainty

The total uncertainty was estimated based on the error term, which was estimated in function of the modelled flows. Fig. 9 presents sample hydrographs during the calibration (left) and validation (right) period, at the RICH site with MUSIC (top) and KAREN (bottom). The events shown in these figures illustrate how the calibration period was characterised for higher volumes than the validation period (as Table 3 shows the calibration period included far larger events than the validation period).

It was identified that some observations were not covered in the parameter uncertainty bounds; which suggest that more accurate predictions might be obtained if the model structure and/or the measured data is improved. Moreover, the large total uncertainty bounds indicate that the uncertainty due to other sources than parameter uncertainty (e.g. measured input data including spatial rainfall distribution, model formulation and assumptions and selected objective function) are significant and cannot be neglected.

3.2. Water quality modelling

3.2.1. Model performance – calibration

Table 5 summarises all efficiencies of the two water quality models tested for both TSS and TN for all five catchments. Both models performed poorly, with the 4-parameter build-up/wash-off model performing slightly better than the 2-parameter regression model.

In general no trend between performance of water quality models and performance of the runoff modelling was found. While the best rainfall/runoff E was achieved for RICH ($E_{\text{MUSIC}} = 0.81$, $E_{\text{KAREN}} = 0.75$), E for TSS for build-up/wash-off model was only 0.12. For the NW catchment with the lowest E for runoff ($E_{\text{MUSIC}} = 0.49$, $E_{\text{KAREN}} = 0.60$), the E for TSS was the highest ($E_{\text{build-up/wash-off}} = 0.46$).

Nash Sutcliffe efficiencies for TN modelling were similar to TSS, with best values around 0 for the catchments GR and RICH and slightly better results for SB and NW and build-up/wash-off model (Table 5).

Fig. 10(a) shows examples of scatter plots of calibration parameters for GR as an example for this catchment but these parameters behave similarly for the other parameters and catchments. Calibration parameters for the build-up/wash-off model do not correlate, which is contrary to past findings by Kanso et al. (2006). They also used the Metropolis algorithm, however their results were based on just 11 events on a small street catchment (160 m^2).

Table 5
Efficiencies of water quality models: E_{best} is the best E obtained during the entire MCMC procedure; E_{mean} is the mean E derived from all the accepted parameter sets.

Catchment	Build-up/wash-off Model				Regression Model			
	TSS		TN		TSS		TN	
	E_{best}	E_{mean}	E_{best}	E_{mean}	E_{best}	E_{mean}	E_{best}	E_{mean}
Gilby Rd (GR)	0.07	−0.51	0.04	−2.45	0.093	0.010	−0.009	−0.204
Richmond (RICH)	0.12	−0.64	0.09	−1.06	0.043	0.015	−0.109	−0.265
Ruffeys Lake, Doncaster (RD)	0.22	−0.37	–	–	−0.003	−0.032	–	–
Shepherds Bush (SB)	0.06	−0.56	0.26	−1.53	0.252	0.235	−0.027	−0.257
Narre Warren (NW)	0.46	−1.72	0.36	−1.03	0.031	0.013	−0.001	−0.157

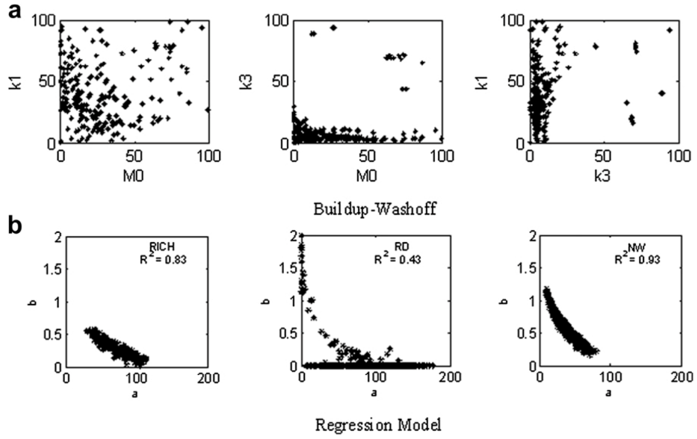


Fig. 10. Correlations between water quality model parameters: (a) Sample parameters for the build-up/wash-off model for TSS at GR, (b) Sample catchments for the simple regression model for TSS.

The high correlation between parameters of the regression model further complicated the calibration process (Fig. 10(b)), and illustrated the 'ill-posed' nature of this model. For example, R^2 between a and b ranged from 0.43 (RD) and 0.93 (RICH) for TSS indicating that different combinations of a and b can lead to the same results. Correlations were not very strong for TN, in which b values were generally close to zero.

3.2.2. Parameter sensitivity

3.2.2.1. Build-up/wash-off model. For TSS, the distributions obtained for the build-up/wash-off model did not indicate one clear peak while giving multiple local optima (Fig. 11-top). Several peaks in these distributions suggest that different combinations of parameters are possible, confirming the problems of equifinality (Beven and Freer, 2001) that may plague urban stormwater quality models.

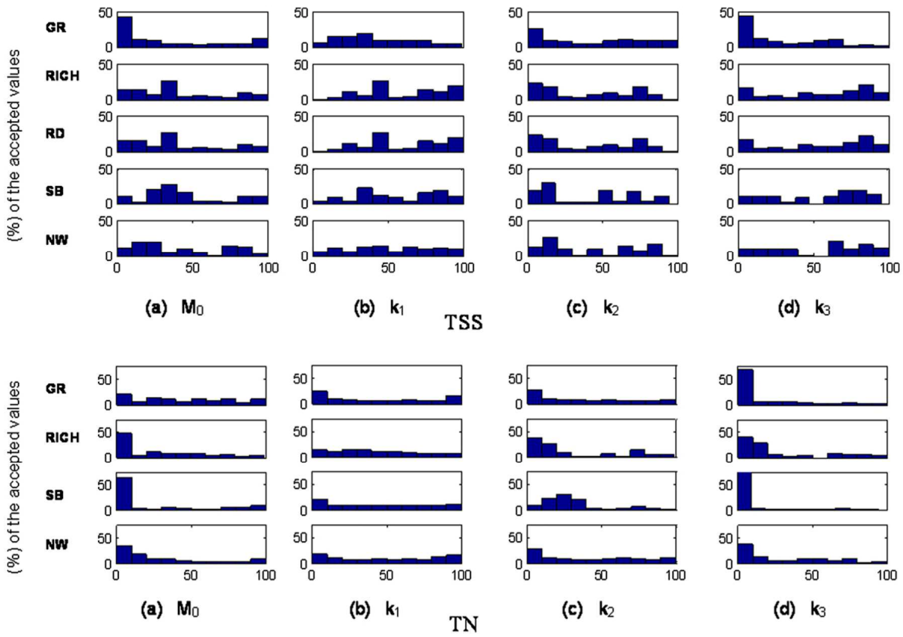


Fig. 11. Histograms for the build-up/wash-off model parameters, for TSS (top) and TN (bottom) (units according to Table 2).

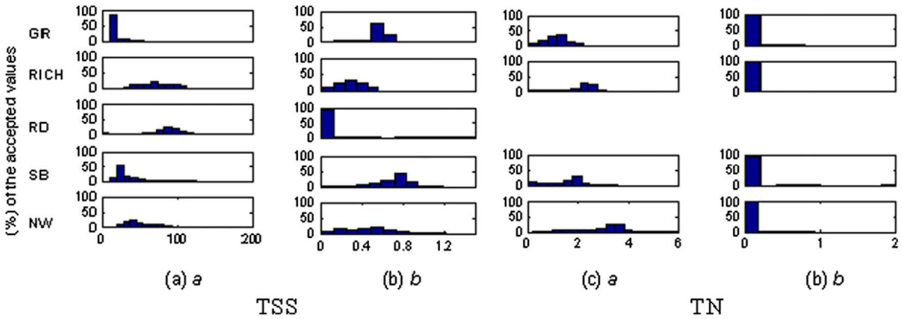


Fig. 12. Histograms for the regression water quality model parameters.

It is interesting to note that there is no link between parameter sensitivity and maximum calibration performance (see Table 5 and Fig. 11). Whilst NW's calibration parameters seemed to be the least sensitive of all the sites, it achieved the highest efficiency *E*. The effect of these less sensitive parameters is also reflected in the mean *E* of NW which is the lowest of all the catchments, although the best *E* at this site is the highest. This indicates problems of the MCMC simulation of "rejecting" model results with low likelihood and might be related to a complex response surface of the likelihood function of that nonlinear dynamic model.

Contrary to TSS, distributions of calibration parameters for mainly dissolved TN (Fig. 11 – bottom) show clearer peaks for three of four calibration parameters. Expected values for M_0 range mainly from 10 to 30 g/m^2, k_2 ranges from 5 to 50 (SB) and k_3 is close to zero with values lower than 1. Only k_1 is insensitive for all catchments. This is interesting as the results of the rainfall/runoff modelling are also insensitive to the parameter *ev*, which is also pertinent only during dry periods. This indicates that modelling of surface runoff can be done on independent events and a continuous simulation might not be necessary. This has also been suggested previously in the literature for stormwater quality models (Duncan, 1999; Vaze and Chiew, 2003).

3.2.2.2. Simple water quality regression model. Contrary to the build-up/wash-off model, the simple regression model was sensitive to both parameters for the two pollutants. The global minimum is clearly defined by the peaks in the histograms for both TSS and TN (Fig. 12). However, the models failed during calibration (Table 5), suggesting that the model structure requires improvement. The discussion below explores the improvements required.

The most probable values for *a* obtained during the TSS calibration ranged from 20 (GR) to 100 (RD) (Fig. 12 – left). The change in *a* illustrates the impact of TSS sources between sites on the model calibration. The minimum value achieved for GR may be explained by the fact that this is essentially a commercial area thus generating lower TSS concentrations than areas where residential activities are taken place (Duncan, 1999). The maximum values obtained for RD suggest the presence of wastewater cross connections and/or some wastewater leakage and infiltration into the stormwater system. For TN, the optimum values for *a* were between 1.3 (GR) and 3.5 (NW). There is a clear increase in *a* values with perviousness which is a rather interesting finding if compared to Tong and Chen (2002). They found that the total amount of nitrogen was higher in impervious urban areas than previous ones. The contrast between findings is possibly explained by the fact that Tong and Chen (2002) calibrated their water quality model against fluxes and therefore areas with higher runoff will produce higher loads.

Optimum values for *b*, for TSS modelling, varied from 0.10 (RD and NW) to 0.79 (SB), indicating the effect of the kinetic energy of the rainfall for detaching TSS from the surface. Very low value (around zero) was obtained for RD hence the power function would almost always result in value close to 1 and that the model is not using runoff as a predictive variable. From this result, it may be speculated that in less urbanised areas, the relatively denser vegetation may intercept the rainfall affecting the process more than kinetic energy of the rainfall on the surface. In addition, the outlier low value of *a* and high value of *b* for SB indicated some parameter compensation in the model. Most likely values of *b* were around 0.001 for all catchments during the model calibration for TN. Once again, values of *b* near zero indicate that TN modelling is mainly led by parameter *a*, suggesting that TN concentrations are independent of the kinetic energy of the rainfall and highly dependent on the sources of such pollutant. It also suggests that TN does not vary much at these catchments, as can be seen by comparing *b* to the coefficient of variation (CV) values found in Table 3.

The average CV value for TSS concentrations was 167%, while for TN it was 86%. The higher TSS CV values might explain the higher TSS *b* values and vice-versa for TN.

4. Conclusions

Two rainfall-runoff models were compared; MUSIC, which is a catchment model that simulates runoff from both impervious and pervious areas, as well as detention of flow in pipes, and the simple KAREN model that simulates only runoff from impervious surfaces using the time-area method. It was found that the effective impervious fraction, *EIF*, is the most important parameter in runoff prediction from both models. *EIF* should be calibrated whenever possible. It was clear that rainfall-runoff model MUSIC is not very sensitive to its pervious area parameters when applied to urbanised catchments. However, if the impervious area fraction is lower than 30%, the pervious area parameters (e.g. soil storage and field capacity) become important. In simple models where pervious areas of a catchment are not simulated (like KAREN), this is compensated by an increase in *EIF*. This is a good example of the impact of model structure uncertainties and exemplifies that calibrated parameters estimated for one model cannot be transferred to other models without a new model calibration, even if they represent the same physical background.

MUSIC was very sensitive to the translation factor in the Muskingum Cunge routing method (*K*) for all catchments, and KAREN was sensitive to the time of concentration parameter (*TC*) for most of the catchments; these parameters representing similar processes also require special consideration. It is recommended that K should be calibrated whenever possible in order to maximize the performance of the flow routing method. *TC* was an important parameter in the case of the three most impervious catchments.

Both models were insensitive to a few dry weather related parameters for all catchments. Although this indicates that rainfall events could be regarded independently instead of continuous simulation, more complex systems will still demand continuous simulation.

It was verified that the residuals between measured and modelled data, for all models and catchments, were not satisfying the normality assumption required by the likelihood measure used in the MCMC method. A Box-Cox transformation of both the measured and modelled flows ensured that the assumption of normality in the residuals was satisfied. However, this data transformation also significantly changed the objective function from one which emphasised peak flows (i.e. least squares) to another which considered all parts of the hydrograph similarly. This was not advantageous, especially when the objective function was chosen for the specified modelling purpose. It was concluded that, for the procedure outlined here, it is not possible to obtain normality in the residuals when trying to maintain the chosen objective function. However, we also advise that more work is needed to verify these findings and to develop a tool to assess uncertainties which are flexible, accurate and efficient.

In terms of total uncertainty, with the adopted approach, a number of observations were not covered in the parameter uncertainty bounds; which suggest that more accurate predictions might be obtained if the model and/or the measured data are improved. Moreover, the large total uncertainty bounds indicates that the uncertainty due to other sources than parameter uncertainty (e.g. input measured data, model formulation and assumptions and selected objective function) are significant and cannot be neglected.

The most widely used water quality models, the build-up/wash-off and a simple flow regression model, were tested. Even with the robust calibration and parameter sensitivity approach used, it is clear that these models poorly represent reality and have a high level of uncertainty. It was concluded that build-up wash-off model parameters are not very sensitive. However, it was surprising to notice that they are not cross-correlated, which conflicts with previous findings. The two parameters of the simple model predicting pollutant concentration in a timestep as a function of runoff (which totally failed to calibrate) were very sensitive and had high levels of cross-correlation. This latter observation clearly indicates that the model is 'ill-posed'. In general, the presently most often applied water quality models cannot represent complex reality of pollution generation and therefore, new models should be developed.

Acknowledgements

The authors would like to acknowledge eWater CRC and CAPES for the scholarship provided and *hydro*-IT GmbH for software support.

References

Achleitner, S., Moderl, M., Rauch, W., 2007. CITY DRAIN© – an open source approach for simulation of integrated urban drainage systems. Environmental Modelling & Software 22 (8), 1184–1195.
Afonso, P., da Conceição Cunha, M., 2002. Assessing parameter identifiability of activated sludge model number 1. Journal of Environmental Engineering 128 (8), 748–754.
Bates, B.C., Campbell, E.P., 2001. A Markov Chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resources Research 37 (4), 937–947.
Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. Water Resources Research 23 (8), 1393–1442.
Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6 (3), 279–298.
Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 246, 11–29.
Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? Journal of Hydrology 354 (1–4), 15–32.
Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jobes, T.H., Donigian, A.S., 2001. HSPF User's Manual. Aqua Terra Consultants, Mountain View, California.
Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society Series B 26 (2), 211–252.
Chakravarti, L., Roy, 1967. Handbook of Methods of Applied Statistics. John Wiley and Sons, New York.
Chiew, F.H.S., McMahon, T.A., 1997. Modelling daily runoff and pollutant load from urban catchments. Water (Journal of the Australian Water and Wastewater Association), 16–17.
Chiew, F.H.S., McMahon, T.A., 1999. Modelling runoff and diffuse pollution loads in urban areas. Water Science & Technology 39 (12), 214–248.
Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D., Summers, R.M., 1992. The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay. Water Resources Research 28 (9), 2353–2363.
CRCCH, 2005. MUSIC Version3-User Guide. Cooperative Research Centre for Catchment Hydrology.
Deletic, A., Dotto, C.B.S., Fletcher, T.D., McCarthy, D.T., Bertrand-Krajewski, J., Rauch, W., Kleidorfer, M., Freni, G., Mannina, G., Tait, S., 2009. Defining uncertainties in modelling of urban drainage systems. In: Proceedings of the 8th International Conference on Urban Drainage Modelling Jointly with the 2nd International Conference on Rainwater Harvesting and Management, Tokyo, Japan.
DHI, 2002. MOUSE Surface Runoff Models Reference Manual. D. Software, Horsolm, Denmark.
Doherty, J., 2003. MICA – Model-Independent Markov Chain Monte Carlo Analysis – User Manual, Watermark Numerical Computing. U.S. EPA.
Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resources Research 46, W05525. doi:10.1029/2009WR008377.
Dotto, C.B.S., Deletic, A., Fletcher, T.D., McCarthy, D.T., 2009. Parameter sensitivity analysis of stormwater models. In: The 6th International Water Sensitive Urban Design Conference and 3rd International Hydropolis. Perth, Western Australia.
Duncan, H.P., 1999. Urban Stormwater Quality: a Statistical Overview. Cooperative Research Centre for Catchment Hydrology, Melbourne, Australia, 80 pp.
Egodawatta, P., Thomas, E., Goonetilleke, A., 2007. Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall. Water Research 41 (13), 3025–3031.
Elliott, A.H., Trowsdale, S.A., 2007. A review of models for low impact urban stormwater drainage. Environmental Modelling & Software 22 (3), 394–405.
Feyen, L., Vrugt, J.A., Ó Nualláin, B., van der Knijff, J., De Roo, A., 2007. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. Journal of Hydrology 332 (3–4), 276–289.
Francey, M., Fletcher, T.D., Deletic, A., Duncan, H., 2010. New insights into water quality of urban stormwater in South Eastern Australia. Journal of Environmental Engineering 136 (4), 381–390.
Freni, G., Mannina, G., Viviani, G., 2008. Uncertainty in urban stormwater quality modelling: the effect of acceptability threshold in the GLUE methodology. Water Research 42 (8–9), 2061–2072.
Freni, G., Mannina, G., Viviani, G., 2009. Urban runoff modelling uncertainty: comparison among Bayesian and pseudo-Bayesian method. Environmental Modelling & Software 24 (9), 1100–1111.
Gallagher, M., Doherty, J., 2007a. Parameter estimation and uncertainty analysis for a watershed model. Environmental Modelling & Software 22 (7), 1000–1020.
Gallagher, M., Doherty, J., 2007b. Parameter interdependence and uncertainty induced by lumping in a hydrologic model. Water Resources Research 43. doi:10.1029/2006WR005347 W05421.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, New York.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, New York.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
INSA/SOGREAH, 1999. CANOE: Manuel D'utilisation. ALISON, INSA LYON, Villeurbanne, France.
Kanso, A., Chebbo, G., Tassin, B., 2005a. Stormwater quality modelling in combined sewers: calibration and uncertainties analysis. Water Science & Technology 52 (3), 63–71.
Kanso, A., Chebbo, G., Tassin, B., 2006. Application of MCMC-GSA model calibration method to urban runoff quality modeling. Reliability Engineering & System Safety 91 (10–11), 1398–1405.
Kanso, A., Tassin, B., Chebbo, G., 2005b. A benchmark methodology for managing uncertainties in urban runoff quality models. Water Science & Technology 51 (2), 163–170.
Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211 (1–4), 69–85.

Lindenschmidt, K.E., 2006. The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling. Ecological Modelling 190, 72–86.
Makowski, D., Wallach, D., Tremblay, M., 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. Agronomie 22 (S1), 191–203.
McCarthy, D.T., 2008. Modelling Microorganisms in Urban Stormwater. Civil Engineering. PhD Thesis: 488. Monash University, Melbourne, Australia.
McCarthy, D.T., Deletic, A., Mitchell, V.G., Fletcher, T.D., Diaper, C., 2008. Uncertainties in stormwater *E. coli* levels. Water Research 42 (6–7), 1812–1824.
Montanari, A., 2005. Large sample behaviours of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resources Research 41 (8), W08406.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. Journal of Hydrology 10 (3), 282–290.
Palmstrom, N., Walker, W.W.J., 1990. P8 Urban Catchment Model: User's Guide. Program Documentation, and Evaluation of Existing Models, Design Concepts, and Hunt-Potowomut Data Inventory, T. N. B. P. R. N. NBP-90-50.
Rauch, W., Kinzel, H., 2007. KAREN – User Manual, Hydro-IT GmbH Innsbruck. (in German).
Rossman, L.A., 2008. Storm Water Management Model – User's Manual Version 5.0. National Risk Management Research Laboratory – U.S. Environmental Protection Agency, Cincinnati, Ohio.
Sartor, J.D., Boyd, G.B., 1972. Water Pollution Aspects of Street Surface Contaminants. Rep. EPA-R2-72-081. U.S. Environ. Prot. Agency, Washington, DC.
Sartor, J.D., Boyd, G.B., Agardy, F.J., 1974. Water pollution aspects of street surface contaminants. Journal of the Water Pollution Control Federation 46 (3), 458–467.
Sherman, L.K., 1932. Stream-flow from rainfall by the unit-Graph method. Engineering News-Record 108, 501–505.
Thierfelder, T., 1999. Empirical/statistical modeling of water quality in dimictic glacial/boreal lakes. Journal of Hydrology 220 (3–4), 186–208.
Thyer, M.A., Renard, B., Kavetski, D.N., Kuczera, G.A., Srikanthan, S., 2007. Bayesian total error analysis for hydrological models: preliminary evaluation using multi-site catchment rainfall data. In: Proceedings of the MODSIM07. International Congress on Modelling and Simulation, Christchurch, New Zealand.
Tong, S.T.Y., Chen, W., 2002. Modeling the relationship between land use and surface water quality. Journal of Environmental Management 66 (4), 377–393.
Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. Environmental Modelling & Software 25 (3), 310–319.
Vaze, J., Chiew, F.H.S., 2003. Comparative evaluation of urban storm water quality models. Water Resources Research 39 (10), 1280.
Wallingford Software, 2009. Infoworks CS Help. Wallingford Software, United Kingdom.
XP-SOFTWARE, 1999. XP-AQUALM – Technical Description. XP Software, Australia.
Xu, C.-y., 1999. Operational testing of a water balance model for predicting climate change impacts. Agricultural and Forest Meteorology 98–99, 295–304.
Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology 358 (1–2), 1–23.
Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology 181 (1–4), 23–48.
Zaghloul, N.A., 1983. Sensitivity analysis of the SWMM runoff-transport parameters and the effects of catchment discretisation. Advances in Water Resources 6 (4), 214–223.
Zoppou, C., 2001. Review of urban storm water models. Environmental Modelling & Software 16 (3), 195–231.

5.5 Conclusions

This chapter presented three studies that explored parameter calibration, model sensitivity and the resulting predictive uncertainties in urban rainfall runoff and pollution generation models with different level of complexities.

It was found that the effective impervious fraction is the most important parameter in runoff prediction. This was followed by the parameters related to the time of concentration. Therefore, such parameters should be calibrated whenever possible. The results showed that MUSIC rainfall/runoff model was not very sensitive to its pervious area parameters when applied to highly urbanised catchments, in which pervious area runoff and baseflow are almost inexistent and that some soil related parameters could be fixed to any value between the obtained posterior distribution ranges. So it can be argued that the model can be simplified when applied for highly urbanised catchments.

In addition, results suggested that the pervious area parameters in MUSIC (e.g. soil storage and field capacity) are in fact "calibration parameters" and are not really related to physical characteristics of the catchment. Whilst some of these parameters are strongly related to the catchment's effective impervious fraction, further work is required in catchments with significant pervious flows to understand their soil profile characteristics. It was suggested that such results can be used for modelling catchments with similar land use, climatic characteristics and hydrological behaviour. It is advised however, that MUSIC should be calibrated against local flow whenever data is available.

The water quality models were shown to be 'ill-posed' and unable to reproduce the pollutant processes in the catchment. While the water quality models were sensitive to all wet weather related parameters, the build-up/wash-off model was not sensitive to the dry weather related parameters. In general, the water quality models presented a high level of uncertainty. However, the outcomes provided useful information for the improvement of existing models and also offered insights for the development of new model formulations. For example, it is recommended that future efforts be put into the development of models which use routed runoff or rainfall intensities, rather than the models which use 'unrouted' variables. Whilst routing essentially introduces extra model parameters, the temporal accuracy gained is likely to outweigh the calibration costs.

Results from the uncertainty analysis showed that some observations were not covered in the parameter uncertainty bounds; which suggest that more accurate predictions might be obtained if the model structure and/or the measured data were improved. Moreover, the large total uncertainty bounds indicates that the uncertainty due to other sources than parameter uncertainty (e.g. measured input data including spatial rainfall distribution, model formulation and assumptions and selected objective function) are significant and cannot be neglected. This topic is explored in Chapter 7.

Also as a result of this work, it was verified that the underlying assumption of the applied uncertainty analysis method (about distribution of the model errors) was not met, and that the method applied to verify the assumption significantly influenced the sensitivity of the model parameters. Further investigation about this is in Chapter 6.

5.6 References

Carrera, J. and Neuman, S. P. (1986). Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 2. Uniqueness, Stability, and Solution Algorithms. *Water Resources Research* **22**(2): 211-227.

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology* **331**(1-2): 161-177.

Chapter 6

Requirements for normally distributed residuals

DECLARATION FOR THESIS CHAPTER 6

Declaration by candidate

In the case of Section 6.2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up.	70

The following co-authors contributed to the work:

Name	Nature of contribution	Extent of contribution (%)
Ana Deletic	Initiation, ideas and reviewing	n/a
David T. McCarthy	Initiation, ideas and reviewing	n/a

Candidate's Signature	[redacted]	**Date** 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(13) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(14) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(15) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(16) there are no other authors of the publication according to these criteria;

(17) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(18) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus

Signature Ana Deletic	[redacted]	**Date** 29/10/2012
Signature David T. McCarthy		**Date** 29/10/2012

6.1 Introduction

Bayesian approaches require the use of likelihood functions which estimate the model's parameters given the measured and modelled datasets. This process often demands a number of assumptions be met, including normally distributed, non-correlated and heteroscedastic residuals (Schoups and Vrugt, 2010). However, most studies do not check these required assumptions (Larssen et al., 2007; Varella et al., 2010). In the cases where they are checked, it is commonly found that these assumptions are not met and are still presented 'as is'. In the literature, a transformation of measured and modelled data (e.g. log or Box-Cox transformation) is used by some modellers to ensure that the assumption of normally distributed residuals is met (Gallagher and Doherty, 2007; Yang et al., 2008). However, as presented in Section 5.4, all transformation methods change the content of the observations (Beven et al., 2008), which then influences the emphasis on various parts of the hydrograph (or pollutograph). This is sometimes not desired if the modelling purpose is to focus on specific parts of the dataset (e.g. flood prediction is linked with peak flows, which are deemphasised when using Box-Cox transformations) (Doherty and Welter, 2010). Furthermore, all observed data have uncertainty, and this should be taken into account in the likelihood function so that the parameters are estimated appropriately; indeed, it is important that the function places more emphasis on data which has lower uncertainty. Weighting strategies can be used to re-adjust how the likelihood function emphasises various parts of the dataset to (1) consider measured data uncertainty and (2) compensate for the Box-Cox transformation which may have adjusted the emphasis in an undesirable way.

This context of data transformation to verify the normality assumption of the model residuals and its consequence in the modelling exercise has not been explored in the urban drainage field. This chapter focuses on ***assessing the impacts of verifying the assumed structure of model errors (here the assumption that the model residuals follow a normal distribution) on model parameter sensitivity and associated predictive uncertainty of stormwater models, and it also explores an alternative strategy to mitigate such impacts.***

This study addresses the following key research questions and hypotheses:

- What are the implications with respect to model efficiency, parameter sensitivity and predictive uncertainty of verifying the assumption that the model residuals follow a normal distribution?
    - Verifying the underlying assumption of the sensitivity and uncertainty analysis method will result in the most comprehensive understanding of the model's uncertainty.
- To what extent can a weighting strategy used to account for the measured data uncertainty also compensate for the impacts caused by the data transformation methods used to ensure normally distributed residuals?

- Box-Cox transformation will reduce the emphasis on peak measured data, yet these peaks (i.e. peak flows for rainfall/runoff models and peak concentrations for water quality models) are considered important in urban drainage modelling because these pose the highest risks. However, at the same time, these peak values often have the lowest relative uncertainty (see McCarthy et al., 2008). As such, it is hypothesised that a weighting strategy used to account for measurement uncertainty in the likelihood function will simultaneously reduce the influence of the Box-Cox transformation process.

This investigation has been collated into one journal paper, which mainly investigated the impacts of verifying the assumption of normally distributed residuals on parameter sensitivity and its associated predictive uncertainty in two urban rainfall-runoff models. The paper was initially presented at the *9th International Conference on Urban Drainage* in Belgrade, Serbia, in 2012 and was subsequently selected for publication in *Water Science and Technology*. The manuscript, *Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?*, is currently in press and forms the body of text of Section 6.2.

6.2 Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?

Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?

C. B. S. Dotto, A. Deletic and D. T. McCarthy

ABSTRACT

This study presents results on the assessment of the application of a Bayesian approach to evaluate the sensitivity and uncertainty associated with urban rainfall runoff models. The software MICA was adopted, in which the prior information about the parameters is updated to generate the parameter Posterior Distribution. The likelihood function adopted in MICA assumes that the residuals between the measured and modelled values have a normal distribution. This is a trait of many uncertainty/sensitivity procedures. This study compares the results from three different scenarios: (i) when normality of the residuals was checked but if they were not normal then nothing was done (unverified); (ii) normality assumption was checked, verified (using data transformations) and a weighting strategy was used that gives more importance to high flows; and, (iii) normality assumption was checked and verified, but no weights were applied. The modelling implications of such scenarios were analysed in terms of model efficiency, sensitivity and uncertainty assessment. The overall results indicated that verifying the normality assumption required the models to fit a wider portion of the hydrograph, allowing a more detailed inspection of parameters and processes simulated in both models. Such outcome provided important information about the advantages and limitations of the models' structure.

Key words | Bayesian approach, normality assumption, uncertainty analysis, urban drainage models

C. B. S. Dotto (corresponding author)
A. Deletic
D. T. McCarthy
CRC for Water Sensitive Cities & Monash Water for Liveability,
Department of Civil Engineering,
Monash University,
Victoria 3800,
Australia
[redacted]

INTRODUCTION

Although all models have errors, they can still provide valuable information as long as these errors are understood. Uncertainty analysis techniques are powerful tools because they provide information about the model sensitivity while estimating the confidence intervals around the model's outputs. A number of methods have now been tested to evaluate urban drainage models, nevertheless they each have their inherent limitations and disadvantages (Dotto *et al.* 2012). For example, the GLUE methodology (Beven & Binley 1992) is easy to implement, however the method requires a large number of Monte Carlo simulations and uses a subjective acceptance threshold to distinguish behavioural from non-behavioural simulations. While Bayesian approaches are not limited by this acceptance threshold, the user needs to understand the distribution of the model errors and verify certain assumptions against this error structure. Indeed, it has been noted that a misspecification of the error structure can lead to an erroneous quantification of the model prediction uncertainty when using these methods (Beven *et al.* 2008). Yet, the robustness of the Bayesian approach compared to other methods is apparent from several studies (e.g. Bates & Campbell 2001).

Many Bayesian approaches assume that the model errors (or residuals between the measured and modelled values) are normally distributed. However, this assumption is often not checked; this is the case for both scientific literature (e.g. Kanso *et al.* 2006; Varella *et al.* 2010) and modelling practitioners, who are often not fully acquainted with uncertainty procedures. In the cases where these assumptions are checked, it is commonly found that the error does not follow any specific distribution and the results are still presented 'as is' (Dotto *et al.* 2011). In the literature, a transformation of measured and modelled data (e.g. log or Box–Cox transformation) is used by some modellers to ensure they meet the assumptions (Gallagher & Doherty 2007; Yang *et al.* 2008). However, it is noted that all transformation methods will intrinsically change the implied information content of the observations (Beven *et al.* 2008). For example, in an

doi: 10.2166/wst.2013.360

urban drainage model, using a log or Box–Cox transformation to meet normality of residuals will place more emphasis on different parts of the hydrograph (i.e. lower flow rates), which in turn significantly influences the sensitivity of the model parameters (Yang *et al.* 2008; Dotto *et al.* 2011).

In summary, previous studies have identified that most of modelling applications do not result in normally distributed residuals (e.g. Doherty & Welter 2010) and that the common methods used to adjust the simulation to match the distribution of the model errors significantly influences the sensitivity of the model parameters (Yang *et al.* 2008; Dotto *et al.* 2011). However, the impacts of verifying the normality assumption on the model sensitivity and associated parameter uncertainty are not understood. This study investigates such impacts with respect to urban rainfall runoff models.

METHODOLOGY

Models and data

MUSIC (eWater CRC 2012). MUSIC requires the catchment area and both rainfall and evapotranspiration time series as input data to continuously simulate flows from urban areas. The model has 13 calibration parameters related to impervious and pervious areas, and a groundwater store. Flows from impervious and pervious areas are modelled separately, with impervious area runoff being primarily as a function of the proportion of catchment imperviousness (*EIF*), with an initial loss term (*Thres*). Runoff from the pervious areas will only occur for large or intense storm events, when the pervious soil storage is either saturated or its infiltration rate is exceeded. The pervious area runoff related parameters are mainly described by: (1) soil storage capacities, the maximum soil storage (*SMax*) and the initial storage level (*SIni*), (2) field capacity (*fc*), and (3) infiltration factors (*coeff* and *sq*). Groundwater is modelled as a store which is recharged when the level in the pervious soil store exceeds the field capacity. The rate of this recharge (*rfac*) is a set percentage of the water in the soil store (*gw*). This store is emptied via baseflow, which is modelled as a percentage of the water within the groundwater store, also a model parameter, *bfac*. In a similar way, deep seepage (*dseep*) is set as a percentage of the groundwater store. The Muskingum Cunge routing method is applied for the routing of flows through the drainage system; the translation (K) and the attenuation (θ) factors require calibration.

KAREN (Rauch & Kinzel 2007). KAREN is a simple linear reservoir model, which requires the catchment area and a rainfall time series as inputs to generate a series of flows originating from impervious areas only. The effective impervious area of the catchment is calibrated as the *EIF* parameter. Runoff from impervious areas occurs after a rainfall threshold has been exceeded. This threshold is represented by the initial loss parameter (*li*) and it is modelled as a single reservoir, similar to that found in MUSIC. While MUSIC's impervious runoff threshold is calculated and reset on a daily basis, KAREN's initial loss is calculated continuously and fills during rainfall and is drained during dry weather by a permanent loss calibration parameter (*ev*). Surface runoff volume is calculated using the linear time-area method, which is similar to the unit hydrograph method. At the beginning of a rainfall event, the effective impervious area is increased according to the flow time on the catchment surface until the whole catchment contributes to runoff after time of concentration (another calibration parameter *TOC*).

Dataset. The data used in this study consist of 2 years of continuous flow and rainfall measurements (in 6 minutes timestep) from an urban catchment in the inner eastern suburbs of Melbourne, Australia. The catchment is drained by a separate stormwater system, with measurements taken within the outlet pipe. The site has a total area of 89 ha, the land use is high-density residential with a total imperviousness of 74% and an average slope of less than 0.1%. The event total rainfall ranges from 1.2 to 40.8 mm, and the mean maximum event runoff rate is 547 L/s.

Model sensitivity and uncertainty analysis

The parameter posterior distributions (PDs) of the model parameters were generated with the software MICA (Doherty 2003). MICA undertakes a Markov Chain Monte Carlo (MCMC) analysis with the Metropolis–Hastings algorithm (Hastings 1970) by sampling mainly in areas of high likelihood, but also allowing exploration of low likelihood areas. The likelihood function adopted in MICA is least square based and assumes that the residuals between the measured and modelled values have a normal distribution (e.g. Gilks *et al.* 1996). The performance of the model was evaluated using the Nash–Sutcliffe efficiency criterion (E) (Nash & Sutcliffe 1970) corresponding to the minimum least square value achieved with MICA. Both the least square likelihood function and E place emphasis on medium/large values, which are often the goals for stormwater management practices (i.e. high volumes – e.g. Chiew & McMahon 1999). Three scenarios were assessed to compare the implications of verifying normality assumption (or not) in the model sensitivity and its associated uncertainty: (i) normality of the residuals was checked but not verified ('*Unverified*'); (ii) normality assumption was verified and a weighting strategy that gives more importance to high flows in the likelihood function was applied ('*Verified1*'); and, (iii) normality assumption was verified, but no weights were applied to the data ('*Verified2*').

For the unverified scenario ('*Unverified*'), while normality was checked, both measured calibration and modelled flow datasets were used without modification (as in Dotto *et al.* (2011)). For the verified scenarios, the measured calibration and modelled data series were transformed using a Box–Cox transformation (Box & Cox 1964) to achieve homoscedascity and ensure the residuals are normality distributed (through Kolmogorov–Smirnov test). The parameters for the Box–Cox transformation (please refer to Box & Cox (1964) detailed description) were set to values determined in preliminary calibration stage. However, a by-product of this transformation is that the used likelihood function now provides more emphasis to lower values, which significantly influences the sensitivity of the model parameters (Yang *et al.* 2008; Dotto *et al.* 2011). To overcome this issue, a weighting strategy which gives higher flows more importance in the likelihood function (as suggested by Gallagher & Doherty 2007) was applied ('*Verified1*'). The weights were computed based on the inverse of the relative uncertainty in the measured data (i.e. the relative error in the measured flow rates calculated using the Law of Propagation of Uncertainties; see McCarthy *et al.* (2008) for more information). In order to analyse the impacts of data transformation alone (i.e. without a weighting strategy) we tested a third scenario in which no weights were applied to the data ('*Verified2*').

Probabilistic predictions of the hydrograph were obtained by estimating the prediction uncertainty originated only from parameter uncertainty and from parameter plus other sources. The percentage of observations covered within the bounds was also calculated.

Parameter uncertainties. For both scenarios, the predictive uncertainty resulting only from parameter uncertainty was obtained by running the models with 1,500 parameter sets sampled from the parameter PDs as adopted by Feyen *et al.* (2007) and Dotto *et al.* (2011). For the verified scenarios, the residuals between the transformed measured and modelled for each of the 1,500 datasets were computed and inspected to ensure that the data transformation was effective and that the parameters' PDs properly represent the parameter uncertainty. In addition, a modified version of the Average Relative Interval Length (ARIL as per Vezzaro & Mikkelsen (2012)) was used to compare the parameter uncertainty results and is the relative width of the uncertainty bounds:

$$\text{ARIL} = \text{Median} \sum_{i=1}^{N} \frac{\text{Limit}_{\text{Upper},i} - \text{Limit}_{\text{Lower},i}}{Q_i} \quad (1)$$

where, $\text{Limit}_{\text{Upper},i}$ and $\text{Limit}_{\text{Lower},i}$ are respectively, the upper and the lower boundary values of the 95% confidence interval; Q_i is the ith modelled value with the most likely parameter set; and, N is the number of timesteps.

Total uncertainties. The total uncertainty (i.e. parameter and all other sources) was estimated according to the method proposed by Dotto *et al.* (2011) for the *Unverified* scenario, while the methodology adopted by Feyen *et al.* (2007) was used for the *Verified1* and *Verified2* scenarios. A step-by-step application of both methods is provided in the referred papers, but their application is summarised below:

- **Unverified.** The total uncertainty term was estimated as a function of the modelled flows. After calculating the residuals, the simulated values and their corresponding residuals were grouped in different bins according to the magnitude of simulated values. The 2.5 and 97.5 percentiles of the residuals in each bin were grouped and plotted against their corresponding simulated values, from which two linear relationships were estimated (one for each of the used percentiles). The error terms were calculated as functions of the simulated values and were added to the simulated values at each time step. The process was repeated for all parameter sets randomly sampled from the generated parameter PDs. The 2.5 and the 97.5 percentiles of all simulations with the added error terms were obtained and used to represent the total uncertainty bounds.
- **Verified.** In the data transformed space, the standard deviation of the error model is assumed constant and is obtained from the RMSE between the transformed observed and simulated values using the most likely parameter set. The total uncertainty was estimated by adding this constant Gaussian error (equal to $\pm 1.96 \times \text{RMSE}$) to the transformed predictions at each timestep. The obtained prediction limits in the transformed space were then back-transformed to the original data space.

RESULTS

Model performance

The overall efficiency of both rainfall runoff models is represented by the Nash–Sutcliffe efficiency criterion (*E*). Figure 1 presents *E* values and shows plots of modelled versus measured flows for both MUSIC and KAREN when the most likely parameter set from each of the scenarios was used (*Unverified* on the left, *Verified1* in the middle and *Verified2* on the right); the bottom plots are in logarithmic scale to provide a better representation of the lower values.

In terms of model efficiency (i.e. E), MUSIC and KAREN performed very similarly under the three scenarios; both models showed high efficiency in the *Unverified* scenario, with this decreasing in *Verified1* and more so in *Verified2* scenario. This decrease in performance was expected; indeed, *E* is biased toward peak flows, yet the likelihood functions used to generate the optimised parameter sets in both *Verified* scenarios are geared to reduce the emphasis placed on these peaks (i.e. least-squares using transformed data). This is clearly shown in the bottom graphs of Figure 1 where the lower flows were more accurately represented using *Verified* scenarios (hence lower *E*), while different parts of the hydrograph (medium/higher flows) were better calibrated in the *Unverified* scenarios (hence higher *E*). The slightly higher efficiency found in *Verified1* (as compared to *Verified2* scenario) reflects the weighting strategy applied, which slightly re-emphasises the peaks after transformation (see methods). The difference between the *Unverified* and the *Verified1* scenarios suggests that the weighting strategy applied in *Verified1* scenario was not sufficient to counteract the effect that the transformation had on emphasising different parts of the hydrograph.

Model sensitivity

The changes in parameter PDs are significant between the three scenarios (Figure 2). When the residuals are normally distributed (i.e. *Verified* scenarios), three of the four parameters in KAREN exhibited much narrower PDs when compared to the *Unverified* scenario. Furthermore, the shift in the most likely parameter values between the scenarios also shows that the data transformation is changing the emphasis placed on certain parts of the hydrograph, even though a weighting strategy was used to solve this issue in *Verified1* scenario.

For both KAREN and MUSIC, the most probable *EIF* values were reduced in both *Verified* scenarios when compared with the *Unverified* scenario. It was proposed that this shift was caused by the higher runoff-coefficients found in larger events; for example, using the measured data, the largest 5% of events (ranked according to maximum flow) had runoff-coefficients which were around 50% larger than the runoff-coefficients of the remaining 95% of events. As such, the *Unverified* scenario places most emphasis on higher flows/events, which in turn have higher runoff coefficients and therefore cause the most-probable value of *EIF* to increase. The opposite is true for the *Verified*

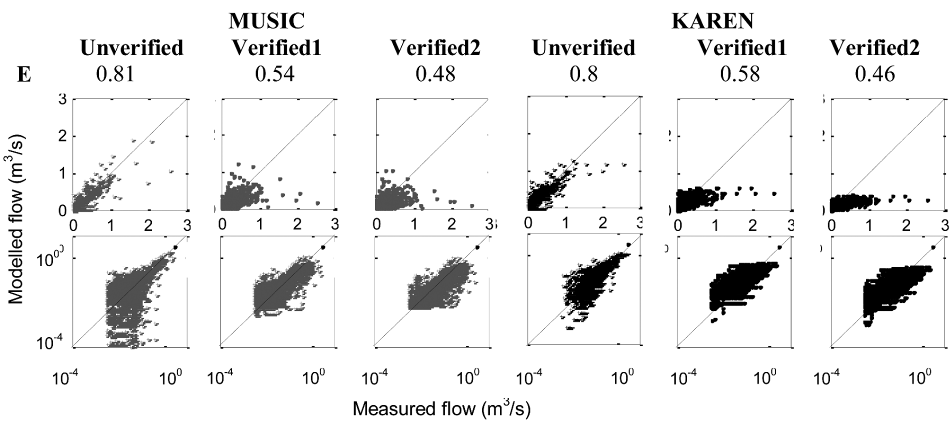


Figure 1 | Model efficiency (E) and measured versus modelled flows. The first three pair of plots (in grey) represent MUSIC: *Unverified* (left) and *Verified1* (middle) and *Verified2* (right) scenarios; and the last three pair of plots (in black) represent KAREN: *Unverified* (left) and *Verified1* (middle) and *Verified2* (right) scenarios. The bottom plots are in logarithmic scale. Q3

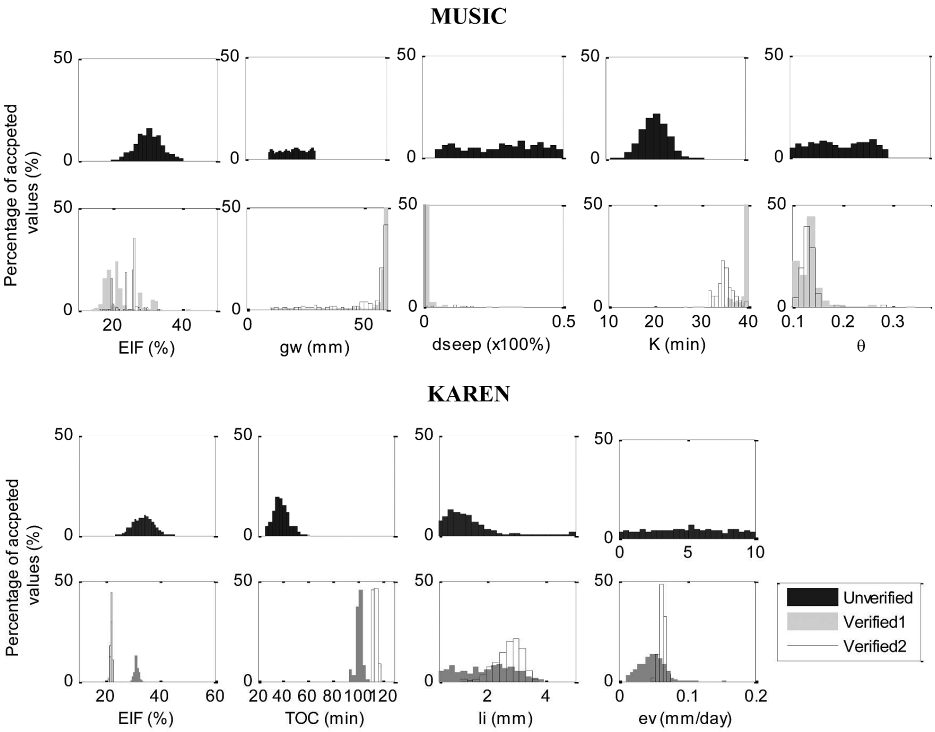


Figure 2 | Sample histograms for MUSIC (top) and KAREN (bottom) parameter PDs for the three scenarios.

scenarios. A similar explanation was found for the shift in the *TOC* and *K* parameters for the two models between the three scenarios; indeed, the measured data showed that the lag-time between the rainfall hyetograph and runoff hydrograph was smaller for larger events (hence emphasising large flows/events would produce lower *TOC* and *K* values).

The changes in parameter PDs for KAREN's *li* and *ev* parameters can also be explained by the change in emphasis of the different portions of the hydrograph between the three scenarios. The *ev* is the constant rate that the initial loss (*li*) volume drains during dry weather periods and it is only important for events that have a similar magnitude as the initial loss volume. For the *Verified* scenarios, *li* values expanded towards higher values than the ones compared to the *Unverified* scenario. The model showed to be sensitive to *ev* for the *Verified* scenarios, in which the focus on low flows could represent events with smaller magnitudes similar to these higher *li* values. This is even more evident in the *Verified2* scenario, in which no weighting strategy was applied to compensate the data transformation. Interestingly, MUSIC's *Thres* (the initial loss related parameter) had similar behaviour in all scenarios as it was found that it can be fixed at any value between 0 and 1 (values above 1 impact on the model performance for all the scenarios – data not shown). Figure 2 also illustrates that MUSIC was not sensitive to the baseflow related parameters (*gw* and *dseep*) in the *Unverified* scenario, but sensitive to the same parameters when the model is required to represent a wider portion of the hydrograph in *Verified* scenarios.

Modelling uncertainty

It was found that the PDs of parameters, obtained with the *Verified* scenarios, adequately described the parameter uncertainty. This is illustrated by the normality plots presented in Figure 3, in which the mean transformed residuals (mean transformed residuals of the 1,500 simulations) from both models closely meet a normal distribution. The figure represents the *Verified1,* while *Verified2* looked very similar – results not shown).

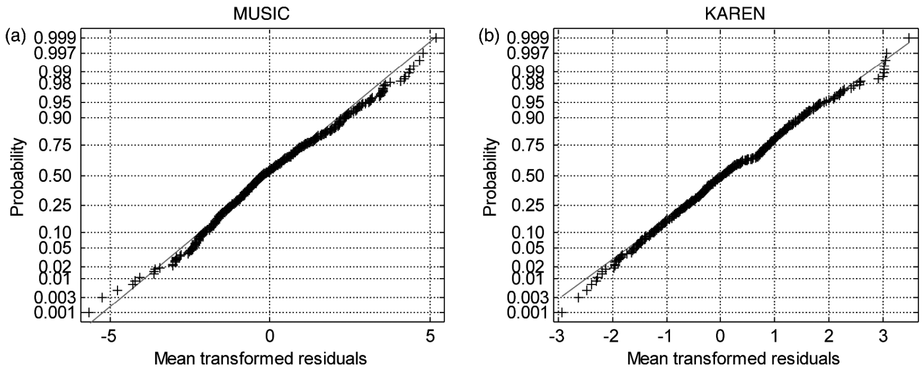


Figure 3 | Normal probability plots of the mean transformed residuals in the *Verified1* scenario.

Table 1 reports the percentage of observations covered by the parameter and total uncertainty bounds for the three scenarios for both models, while Figure 4 presents sample hydrographs with the prediction uncertainty bounds. Coverage from parameter uncertainties was always lower than the percentage coverage for the total uncertainty; similar results were found by Li *et al.* (2010) who showed that coverage obtain for parameter uncertainty was around 20% while for total uncertainty this was around 80%. This significant increase in coverage reflects the fact that there are possibly more significant uncertainties prevalent in the model than parameter uncertainties (e.g. input measured data, observed data, model structure, formulation and assumptions).

The coverage from total uncertainties was significantly different for the *Verified* and *Unverified* scenarios. The coverage from parameter uncertainties varied significantly between models and between each scenario. For MUSIC, the number of observations covered by the parameter uncertainty for the *Unverified* scenario was smaller than the number covered by *Verified* scenarios; the opposite trend was observed for the KAREN model. The coverage from parameter uncertainties is directly linked to the shape and form of the parameter PDs shown in Figure 2; as such, it is logical that since we see clear differences in parameter PDs between models and scenarios, the percentage coverage also varies both with model type and between *Unverified* and *Verified* scenarios. For example, KAREN's parameter PDs in the *Unverified* scenario are mostly wider/flatter which produce wider uncertainty bounds and hence higher coverage; this is compared with the *Verified* scenarios which have narrower/sharper parameter distributions which produce narrower bounds and therefore lower coverage.

DISCUSSION

In the *Unverified* scenario, the least square likelihood function placed emphasis on medium/large flows. The results above indicated that in the *Unverified* scenario both MUSIC and KAREN were driven primarily by parameters which describe the effective imperviousness and time of concentration effects (MUSIC K and KAREN TOC), while the other parameters played a secondary role. This suggests that only some of the processes represented by each model are being utilised, while others lay dormant. In fact, it is hypothesised that in the *Unverified* scenario, both models are behaving as 'black-box' models, and simply attempt to represent peak flows without trying to accurately represent

Table 1 | Summary of the observations within the uncertainty bounds

	MUSIC			KAREN		
	Unverified	***Verified1***	***Verified2***	***Unverified***	***Verified1***	***Verified2***
Observations within the *parameter* uncertainty bound (%)	32	55	61	45	9	5
Observations within the *total* uncertainty bound (%)	98	73	80	99	63	71
ARIL *parameter* uncertainty	0.91	1.45	2.14	1.23	0.12	0.07

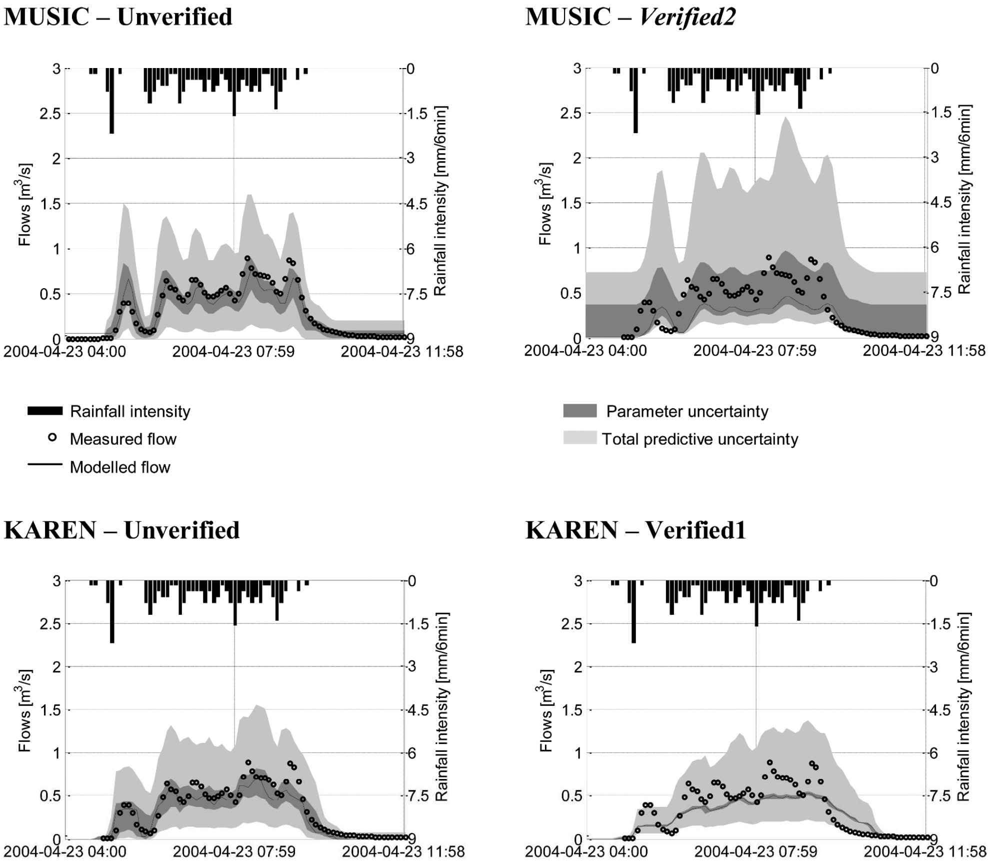


Figure 4 | Sample hydrographs with the prediction uncertainty (23rd April 2004). The dark dots represent the observed data while the black line is the modelled data with the optimised parameter set (according to least squares); the shaded area shows the total predictive uncertainty associated with the total error related to the modelling residuals.

reality. However, the results suggest that, even in this state, the models can represent these peak flows, with very similar performances obtained from both models; it is noted though that verification (a true test of a model's performance) has not been performed here as this was conducted elsewhere (Dotto *et al.* 2011).

When a Box–Cox transformation was applied to the measured and modelled data (*Verified1* and *Verified2* scenarios), the likelihood function placed more emphasis on different parts of the hydrograph and the low flows started playing a bigger role. To account and represent such lower values, parameters which were deemed insensitive in the *Unverified* scenario began to influence the models' outputs. For example, KAREN's *li* and *ev* parameters became sensitive in the *Verified* scenarios suggesting that initial/depression loss processes were active, while MUSIC's pervious area related parameters (and hence processes) also started responding and influencing the model outcomes.

It is hypothesised that using transformations can more adequately calibrate the models as more of the fundamental processes are activated during this calibration procedure. As a result, in the *Verified* scenarios, MUSIC was capable of predicting the lower flows with more accuracy (as the base-flow processes were activated); this is reflected in Figure 1 with the predicted and measured low flows being more closely aligned. Further proof might be gauged from Table 1, which shows MUSIC's coverage (from parameter uncertainty) increased to over 50% in the *Verified* scenarios. This increase in coverage is directly related to an increase in the width of the uncertainty bands (see ARIL, Table 1);

ARIL in the *Verified* scenarios is at least 50% greater than that of the *Unverified* scenario. These wider bands are simply a by-product of the fact that the model is being asked in the *Verified* scenarios to represent a wider portion of the hydrograph, not just peak flows.

CONCLUSIONS

This paper investigated the application of a Bayesian approach, MICA, to evaluate the sensitivity and uncertainty associated with urban rainfall runoff models when the assumption about the normality of the residuals is verified or not. The study compared the results from three different scenarios: (i) normality of the residuals was checked but not verified; (ii) normality assumption was verified and a weighting strategy that gives more importance to high flows in the likelihood function was applied; and, (iii) normality assumption was verified, but weights were not applied to the data. The modelling implications of such scenarios were analysed in terms of model efficiency, sensitivity and uncertainty assessment.

The overall efficiency of both rainfall runoff models was similar under the tested scenarios. Both models showed high efficiency in the scenario in which the normality assumption was not verified, with this decreasing in the scenarios that the assumption was verified. This decrease confirmed that the data transformation modified their implied information content, which also reflects a change in the likelihood function. In addition, the changes in parameter PDs were significant between the scenarios. The evident shift in the most likely parameter values between the scenarios, also showed that the data transformation is altering the emphasis placed on certain parts of the hydrograph, not only when no weights are applied but also when a weighting strategy was used.

It is hypothesised that verifying the normality assumption can more adequately calibrate the models as more of the fundamental processes are activated during this calibration procedure. This is explained by the fact that the verified scenarios activated most of the parameters processes (e.g. dry weather related) in the models, while the unverified scenario was driven by one main parameter (representing the impervious runoff in the catchment). As such, MUSIC, a model that represents impervious, pervious and base-flows was probably better calibrated in the verified scenarios. The inspection of measured and modelled data points and the larger parameter uncertainty coverage revealed that the model better simulated the lower flows, which are the most frequent in the dataset. As opposite, in the unverified scenario, the low flows were not well predicted while the peak flows (which are the minority in the dataset) were more accurately modelled. This being a consequence of the least square likelihood function that favours peaks, but scarifies the remaining parts of the hydrograph. In addition, verifying the normality helped to identify some structure errors. KAREN was sensitive to most of its parameters in the verified scenario, however it seems that model structure does not incorporate sufficient processes to represent reality when lower flows were favoured in the likelihood function. In summary, the results indicated that verifying the normality assumption required the models to fit a wider portion of the hydrograph, allowing a more detailed inspection of parameters and processes simulated in both models. Such outcome provided important information about the advantages and limitations of the models' structure.

REFERENCES

Bates, B. C. & Campbell, E. P. 2001 A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research* **37** (4), 937–947.

Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6** (3), 279–298.

Beven, K. J., Smith, P. J. & Freer, J. E. 2008 So just why would a modeller choose to be incoherent? *Journal of Hydrology* **354** (1–4), 15–32.

Box, G. E. P. & Cox, D. R. 1964 An analysis of transformations. *Journal of the Royal Statistical Society. Series B* **26** (2), 211–252.

Chiew, F. H. S. & McMahon, T. A. 1999 Modelling runoff and diffuse pollution loads in urban areas. *Water Science and Technology* **39** (12), 214–248.

Doherty, J. 2003 *Model Independent Markov Chain Monte Carlo Analysis – MICA*. Watermark Numerical Computing and US EPA, US.

Doherty, J. & Welter, D. 2010 A short exploration of structural noise. *Water Resources Research* **46**, W05525. **Q1**

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

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W. & Deletic, A. 2012 Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research*(0). **Q2** eWater CRC 2012 *Model for Urban Stormwater Improvement Conceptualisation (MUSIC) (Version 5.0)*. eWater Cooperative Research Centre, Canberra.

Feyen, L., Vrugt, J. A., Nualláin, B. Ó., van der Knijff, J. & De Roo, A. 2007 Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology* **332** (3–4), 276–289.

Gallagher, M. & Doherty, J. 2007 Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling and Software* **22** (7), 1000–1020.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. 1996 *Markov Chain Monte Carlo in Practice*. Chapman and Hall, New York.

Hastings, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1), 97–109.

Kanso, A., Chebbo, G. & Tassin, B. 2006 Application of MCMC-GSA model calibration method to urban runoff quality modeling. *Reliability Engineering and System Safety* **91** (10–11), 1398–1405.

Li, L., Xia, J., Xu, C.-Y. & Singh, V. P. 2010 Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *Journal of Hydrology* **390** (3–4), 210–221.

McCarthy, D. T., Deletic, A., Mitchell, V. G., Fletcher, T. D. & Diaper, C. 2008 Uncertainties in stormwater *E. coli* levels. *Water Research* **42** (6–7), 1812–1824.

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

Rauch, W. & Kinzel, H. 2007 *KAREN – User Manual*. hydro-IT GmbH, Innsbruck, (in German).

Varella, H., Guérif, M. & Buis, S. 2010 Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. *Environmental Modelling and Software* **25** (3), 310–319.

Vezzaro, L. & Mikkelsen, P. S. 2012 Application of global sensitivity analysis and uncertainty quantification in dynamic modelling of micropollutants in stormwater runoff. *Environmental Modelling and Software* **27–28** (0), 40–51.

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J. & Yang, H. 2008 Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology* **358** (1–2), 1–23.

First received 14 November 2012; accepted in revised form 29 April 2013

6.3 Conclusions

This chapter studied the impacts of verifying the assumption around the distribution of the model errors on the parameter sensitivity and its associated predictive uncertainty of two rainfall runoff models.

Ensuring the residuals were normally distributed produced different model efficiencies, posterior parameter distributions and prediction bounds than when the residuals were not correctly distributed. The main reason for this was that the data transformation used to meet the normality assumption altered how the likelihood function emphasised various parts of the measured dataset and the fact that the weighting strategy based on measurement uncertainties could not entirely compensate for this alteration. Results also indicated that verifying the normality assumption could more adequately calibrate the models; indeed, when the normality assumption was verified, most of the model's processes were activated (resulting in more influential parameters), while only few parameters drove the outputs when the normality assumption was not verified (i.e. only some parameters were influential). As such, the data transformation approach coupled with the weighting strategy was chosen for further application of the method in Chapter 7.

6.4 References

Beven, K. J., Smith, P. J. and Freer, J. E. (2008). So just why would a modeller choose to be incoherent? *Journal of Hydrology* **354**(1-4): 15-32.

Doherty, J. and Welter, D. (2010). A short exploration of structural noise. *Water Resources Research* **46**(W05525, doi:10.1029/2009WR008377).

Gallagher, M. and Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling and Software* **22**(7): 1000-1020.

Larssen, T., Høgåsen, T. and Cosby, B. J. (2007). Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Ecological Modelling* **207**(1): 22-33.

Varella, H., Guérif, M. and Buis, S. (2010). Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environmental Modelling and Software* **25**(3): 310-319.

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J. and Yang, H. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology* **358**(1-2): 1-23

Chapter 7

Impact of input and calibration data uncertainties on the sensitivity and uncertainty of stormwater models

DECLARATION FOR THESIS CHAPTER 7

Declaration by candidate

In the case of Section 7.6, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Initiation, ideas, methodology set up, data preparation and model runs, analysis of the results, and leading write-up.	65

The following co-authors contributed to the work:

Name	Nature of contribution	Extent of contribution (%)
Manfred Kleidorfer	Initiation, ideas, methodology set up, data preparation and interpretation, analysis of the results, and contribution to write-up.	15
Ana Deletic	Initiation, ideas and reviewing	n/a
Wolfgang Rauch	Initiation, ideas and reviewing	n/a
David T. McCarthy	Initiation, ideas and reviewing	n/a

Candidate's Signature	[redacted]	**Date** 29/10/2012

Declaration by co-authors

The undersigned hereby certify that:

(19) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(20) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(21) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility

(22) there are no other authors of the publication according to these criteria;

(23) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(24) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location	Unit for Environmental Engineering, University of Innsbruck, Austria	
Signature Manfred Kleidorfer	[redacted]	**Date** 29/10/2012
Signature Wolfgang Rauch		**Date** 29/10/2012

Location	CRC for Water Sensitive Cities & Monash Water for Liveability. Department of Civil Engineering, Monash University Clayton Campus	
Signature Ana Deletic	[redacted]	**Date** 29/10/2012
Signature David T. McCarthy		**Date** 29/10/2012

7.1 Introduction

Measured data is imperative for the application of any model. While rainfall data is the main input for most rainfall runoff models, flow data is used to calibrate and validate these models (Achleitner et al., 2007; McCarthy, 2008; eWater CRC, 2012). In water quality modelling, rainfall intensities or runoff rates are usually the input for the models and pollutant measured concentrations are used for model calibration and validation (Beck, 1987; WP Software, 1995). Uncertainties are inherent to any data monitoring and the predictive performance of stormwater models is limited by the uncertainty in measured data. Therefore these uncertainties and their impacts on the models should be explored. Thus, the main aim of this study was to ***explore the impact of measured input and calibration data uncertainty on the performance, sensitivity and predictive uncertainty of stormwater quantity and quality models.***

This chapter focuses on addressing the following key research question and hypotheses:

- What are the impacts of input and calibration data uncertainties on the sensitivity and predictive uncertainty of stormwater models?
    - the model parameters can entirely compensate for the uncertainty in input and calibration data; and,
    - systematic errors in measured data will have more impact on the model sensitivity and uncertainty than random errors because they are time-dependent, and therefore they will be continuously propagated through the model.

In order to explore the impact of input and calibration data uncertainty on the sensitivity and predictive uncertainty of stormwater models, it is important to first understand the sources of uncertainties in this data. An introduction of the main sources of uncertainties in the measured data was presented in Subsection 2.3.3. A summary of the main sources of uncertainties in the measured variables of interest and a review of their quantification are presented in Sections 7.2 to 7.4. Based on the values compiled in these sections, error models were developed to represent measured data uncertainty in the modelling exercise (Section 7.5). The investigation of the impact of the input and calibration measured data uncertainty (here, estimated through the error models) on the model sensitivity and predictive uncertainty is presented in a journal paper, *Impacts of measured data uncertainty on urban stormwater models* submitted to *Journal of Hydrology*. This paper forms the body of text of Section 7.6.

7.2 Rainfall data uncertainty

Tipping bucket rainfall gauges are the standard and most used device for measuring rainfall data (Sevruk, 2002). The main sources of uncertainties in the data measured with these gauges are related to both catching and counting errors (Molini et al., 2005b). The total amount of daily, monthly or longer period rainfall tend to be underestimated due to the effect of evaporation, wetting, splashing and wind occurring within or on the top the bucket (Molini et al., 2005b).While splashing losses were found to be only up to 2% and evaporation losses were up to 4%, the wind losses were found to be inversely proportional to the rain intensity and were up to 30% for rainfall intensity around 0.25 mm/h (Sevruk, 1982).

Counting errors are related to the inherent mechanical errors of the tipping bucket. As opposite to the catching errors, the counting errors have stronger impact on rainfall intensities than on total rainfall amount. For example, the gauge delays to respond to quick changes in the rainfall intensities because of the time required for the bucket to fill up and dispense. In addition, the rainfall intensities tend to be underestimated during extreme intense events because the bucket cannot tip fast enough (some rainwater is lost during the tipping movement of the bucket) (Molini et al., 2005b; Wang et al., 2008). Some manufactures claim that the maximum error range for their tipping bucket device for rainfall intensities between 2 to 400 mm/h is between -5% and 5%. However, this type of errors was reported to induce an error of -10 to -15% for rainfall intensities higher than 200 mm/h (Molini et al., 2005b).

It is often assumed that the rainfall intensity I is simply calculated by assuming a linear gauge response (Maksimovic et al., 1991). However, the linear relationship can be quickly revoked as the volume of water that tips is not constant, but a function of the rainfall intensity, and at higher intensities tipping buckets usually underestimate rainfall because water is lost during the tipping movement. It follows that the relationship between rainfall intensity and tipping rate is not always linear and Molini et al. (2005a) reported that neglecting these systematic mechanical errors impacted the assessment of the design rainfall for urban scale applications.

Dynamic calibration of the tipping bucket was carried out by many to determine the form, shape and parameters of such a relationship (Niemczynowicz, 1986; Maksimovic et al., 1991; Simic and Maksimovic, 1994; Molini et al., 2005a; Molini et al., 2005b; Pavlyukov, 2007). Results of these many studies confirmed that a simple power relationship works well:

$$I = \alpha N^{\beta} \qquad \text{Equation 7.1}$$

where I is the rainfall intensity (mm/min), N the number of tips per time (tips/min); and, α and β are parameters depending on the tipping bucket. In this context, it seems that the main source of errors comes from:

- relying on the nominal value for water volume (e.g. 0.2 mm), which influences the N term in Equation 7.1; and,
- assuming a linear relationship between I and N .

Measured rainfall data can also be influenced by the fact that the tipping bucket device cannot always grasp some of the rainfall temporal distribution. For example, it is possible that the rainfall may stop before the collector has tipped. Moreover the relationship between the real rain drop rate and tipping rate is not necessary linear. For this reason different approaches have been proposed and tested for the estimation of rainfall intensities from rainfall recorded tips (McCarthy, 2008; Wang et al., 2008).

Battery, logger and computer clock failures are significant source of errors in rainfall measurements. Time drifts are inherent to any battery controlling logging devices and values around 0.07 min/day were reported by McCarthy (2008).

In summary, except for the catching errors that are physics-based, most of the errors in rainfall measurements can be detected or fixed though calibration. On the other hand, the inadequate or lack of calibration can cause systematic errors due to the same sources. Stransky et al. (2006) demonstrated that the flows modelled with a rainfall runoff model were impacted by inadequate and/or lack of static and dynamic calibration of tipping bucket rain gauges.

The spatial variability of rainfall is another issue. It is common that the point rainfall measured with the tipping bucket is different from the average rainfall calculated if several gauges were installed along the catchment. Haydon and Deletic (2009) reported variations of up to 30% for the rainfall from three rain gauges in a rural catchment.

All these errors associated to the rainfall data eventually propagate through the models and most of the time the modeller is not even aware of them. Some studies (most in the hydrologic field - Vrugt et al. (2008) and Thyer et al. (2009b)) introduced an error model with calibration parameters to correct the rainfall to achieve better model performance. While this is very important, the impact of erroneous rainfall data on the model performance and sensitivity was less addressed. For the purpose of this research, error models were developed to replicate the random and systematic errors in measuring rainfall; they are presented in subsection 7.5.1.

7.3 Flow data uncertainty

Uncertainties in flow measurements are very much related to the measurement equipment. Harmel et al. (2006a) reviewed and compiled uncertainties in individual stream flow measurements with a range of different methods and reported that the velocity-area method was the most accurate among other available methods. The flow rates are calculated by the product between the wetted cross sectional area and the velocity as follows:

$$Q = v\left[Rad^2 Arc\cos\left(\frac{Rad-h}{Rad}\right) - (Rad-h)\sqrt{2Rad\,h - h^2}\right] \quad \text{Equation 7.2}$$

where v is the measured stormwater velocity (m/s); Rad is the measured radius of the pipe (m); and, Q is the calculated flow (m3/s).

Flow measurement uncertainties for the velocity-area method ranged from 2% to 20% depending on efforts spent for the measurements (e.g. financial and personnel resources) and the hydrologic conditions. The uncertainty sources of this type of measurement are in the estimation of the: channel's cross section (the radius, Rad for circular pipes, depth (h) and velocity (v)).The errors from these three sources can be estimated/calculated using multiple measurements or can be based on scientific literature values. The variables h, v and Rad, measured with different instruments, are assumed independent and not correlated. Under these conditions, the law of propagation of uncertainty can be used to estimate the combined standard uncertainty (Bertrand-Krajewski and Muste, 2007). The Law of Propagation of Uncertainty (LPU - Taylor and Kuyatt, 1994) propagates these sources of uncertainty through Equation 7.1 to estimate the uncertainty in the flow measurements. The LPU is only outlined, but is fully described in Taylor and Kuyatt (1994).

Often a measurand Y is obtained as a function of n other quantities that can be directly measured $X_1, X_2, \dots X_n$ such that $Y = f(X_1, X_2, \dots X_n)$. As X_i and $Y$are not really known, they can be estimated as x_i and y (with $y = f(x_1, x_2, \dots x_n)$. Usually x_i is the mean of the n repeated measure of X_i. The uncertainty, $u(x_i)$ associated to x_i is the standard deviation of the mean. The true, but yet unknown x_i has about 95% chance of being within the interval $[x_i - 2u(x_i), x_i + 2u(x_i)]$. For independent uncorrelated variables (as the case of h, v and Rad) the uncertainty $u(y)$ is calculate as a first order Taylor series approximation of $Y = f(X_1, X_2, \dots X_n)$:

$$u(y)^2 = \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2\,(x_i) \quad \text{Equation 7.3}$$

While the random errors can be propagated using the LPU, very little knowledge on the actual systematic measurement error of flow is available. Nevertheless, three main potential errors are known:

- height measurement 'zero-point' drift - drift is common in many flow measurement devices, and is usually avoided using regular calibration; In general, pressure probes are more susceptible to zero drifts than ultrasound;
- inaccurate (re-)calibration of height measurement - when the probe requires recalibration (or even when the initial calibration of the probe is conducted) the calibration might be biased (i.e. include systematic error). For example, the crew may always over-estimate or underestimate the actual depth of water; and,
- inaccurate velocity calibration or incorrect probe set-up - the probe may always over or under-estimate due to factory default errors or by improper positioning within the pipe.

Prodanovic (2009) explained that different Doppler velocity probes use different water level measurements. Mostly, they use either pressure type sensors or ultrasonic, or they have the option to use an external level sensor. In general, pressure probes are susceptible to zero drifts more than ultrasonic probes. Also, the reference air pressure measurement is required, and as a result of added deposit on the sensor, the frequency response of the probe is changed over time, so it will become ‘slower’. Furthermore, the ultrasound level measurement depends on the water temperature and concentration of suspended solids as the velocity is not the same for clean and dirty water. Considering the velocity measurement, the equipment manufacturers usually provide the accuracy of the probes, although they are usually not realistic values. For instance, an error of ±2% is suggested by the manufacturer of the Sigma 950 flow meter. However, Prodanovic (2009) advised that the uncertainty associated to the velocity measurement is in reality dependent on how the software in the probe handles the dropdowns in signal (situation when there is no echo from measured volume). As such, the height measurement ‘zero-point’ error can be detected and fixed through regular calibration, but the two other errors are unlikely to be really eliminated.

As with the rainfall data errors, the errors associated with the measured flow data also propagate through the models. Some studies on hydrological modelling developed error models to correct the flow measurements (Vrugt et al., 2008; Thyer et al., 2009b); however, the impact of erroneous flow calibration data on the model performance and sensitivity has not been addressed in the urban drainage field. Similarly to the previous subsection, error models were developed to account for the random and systematic errors in flow measurements and they are described in subsection 7.5.2.

7.4 Uncertainty in pollutant discrete samples (TSS)

The uncertainty associated with discrete stormwater quality parameters (pollutant concentrations) originate from a wide range of sources (McCarthy et al., 2008):

1. sampling methods;
2. storage methods; and,
3. analytical/laboratory methods.

Substantial research has been developed to characterise the uncertainties associated to TSS discrete samples, e.g. Ahyerre et al. (1998), Harmel et al. (2006), Rode and Suhr (2007), and McCarthy (2008). In these studies the authors presented the uncertainty associated with each source for different pollutants. Some figures resultanting from their work are summarised in Table 7.1.

Sampling uncertainties are related to the fact that a sample is often taken from just one position within the water cross-section and is usually assumed to represent the entire water column. It is most common to collect samples from a point near to the bottom of the stormwater pipe. The position of the intake tubing in the water cross section influences particulate pollutants more so than soluble pollutants as the particulates tend to settle and the dissolved ones have a more even distribution along the water column. In addition, sampling uncertainty can be also caused by a poor setup of auto-samplers (e.g. alignment of the suction tube) or other sampling issues. Sampling uncertainty associated with TSS concentrations was reported to range between 2 to 33% (Harmel et al., 2006a).

Storage uncertainties are related to the time period between the sampling time and when the samples are analysed in the laboratory. The storage environment (i.e. if the samples are well preserved and/or refrigerated) can help reduce storage uncertainty for some pollutants (e.g. using refrigerated autosamplers for the collection of samples for microorganisms is often recommended). However, these storage requirements vary for different pollutants, and are mainly driven by the physical and chemical properties of the pollutant. The uncertainty due to storage and transport of samples to the laboratory has been reported as minimal for TSS, but significantly larger for dissolved pollutants (e.g. TN) (Kotlash and Chessman, 1998). For example, the range of -16 to 49% was reported for TN samples, even when they were kept in ice and analysed within 6 hours (Kotlash and Chessman, 1998).

Analytical uncertainty is associated with all the processes related to the laboratory analysis. Sample handling, preparation, staff expertise, analytical method and equipment are some examples of sources of error in the laboratory. Potentially, the uncertainties associated to TSS are lower than other pollutants because of the low complexity of the analytical technique used for TSS (simple filtration and weighing).

The magnitude of the systematic errors in pollutant discrete samples have also been computed and reported in previous studies (Gordon et al., 2000; Harmel et al., 2006a; Harmel et al., 2009; McCarthy et al., 2009). For example, Ahyerre et al. (1998) reported a difference of 15% between TSS concentrations sampled with two different samplers working at the same time (Table 7.1).

Table 7.1 Summary of literature with the different TSS discrete sample uncertainties.

Systematic errors				
	Uncertainty (%)			Source of literature
(1) Sampling	**Range**		**Median**	
Position of sampling point in the pipe	14	33	20	Martin et al. (1992)
	4.7	23.8	14.0	McCarthy et al. (2009)
	2	12		Rode and Suhr (2007)
Total sampling uncertainty	**Range of unc for diff events**			
	Min	**Max**	**Median**	
	21.61	40.06	31.63	Huang et al. (2010)
	15	20	17.5	Martin et al. (1992)
	12	26	19	Harmel et al. (2009)
	Range of unc for diff events			
(2) Storage	**Min**	**Max**	**Median**	
Total storage uncertainty			10	Bertrand-Krajewski and Bardin (2002)
	Uncertainty (%)			
(3) Analytical uncertainties	**Range**		**Median**	
	-9.8	5.1	-4.9 to -2.5	Gordon et al. (2000)
			40	Ahyerre et al. (1998)
Total analytical uncertainty	**Range of unc for diff events**			
	Min	**Max**	**Median**	
	<8			Harmel et al. (2009)
Cumulative - Combining sources	**Range of unc for diff events**			
	Min	**Max**	**Median**	
	14	104	23	Harmel et al. (2009)
Random errors	**Range of unc**			
	± 25 to ± 30			Bertrand-Krajewski and Bardin (2002)
	± 12% to ±26%			Harmel et al. (2009)

Random errors associated with the discrete stormwater quality parameters have often been reported in the literature. While Bertrand-Krajewski and Bardin (2002) and Bertrand-Krajewski et al. (2003) reported values between 25 and 30% for random errors in TSS, Harmel et al. (2009) presented values ranging from 12 to 26% (with median of 18%) for the same pollutant.

The impact of erroneous TSS calibration data on the model performance and sensitivity has not been explored. Similarly to the previous subsections, error models were developed to account for the random and systematic errors in TSS discrete samples and they are described in subsection 7.5.3.

7.5 Error models

Error models were developed to disturb measured data with errors to evaluate the impact of errors in input and calibration model sensitivity and uncertainty. The error models were developed based on the information about the uncertainty in rainfall, flows and pollutants data as provided in the previous section. Different error models were created for random and systematic errors in each of the variables.

7.5.1 Rainfall error model

Rainfall random errors

The random error for rainfall data due to wetting, splashing, evaporation and wind effects was sampled from a uniform distribution in the range of [-0.5, 0.5] as previously adopted in the literature (Rauch et al., 1998; Haydon and Deletic, 2009). Additionally, the random effect of the spatial variation of the rainfall was also evaluated; for this each event was disturbed by a different factor sampled from a uniform distribution in the range of [-0.7, 1.3]; this is a rather simplistic approach and was designed to reflect a local spatial variation (around the gauge) and not spatial variation throughout the whole catchment.

Rainfall systematic errors

For the systematic errors, two main sources of errors were considered: time drifting and mechanical errors. The development of the error model is explained next.

Time drift - the first step was to account for the time drifting effect in the rainfall loggers. It was assumed that the rain gauge is calibrated regularly, and that the logger time drift is linear with a rate of t_{drift} (mm/day). The time between successive calibrations is called t_{reset} (month). After whic, the logger time becomes equal to the real time and the time drift re-start from zero. In other words, the time logger will drift according to $t_{drift} = t_{true} - t_{logger}$, here assumed as a constant rate, every day until the rain gauge is re-calibrated in every t_{reset} month(s) and the logger time is adjusted to the real time $t_{true} = t_{logger}$. The rate of ±0.07 min/day was assumed for this study. The re-calibration time t_{reset} was assumed to be 1 and 6 months for best and worst case scenarios, respectively. Finally, a time drifting effect of ±0.14 minutes/day was assumed in the rainfall and flow loggers.

Mechanical errors - an offset of ±30% was assumed to represent the error in the 0.2 mm nominal volume of the tipping bucked used in this study. This offset is realistic when compared to the values reported in the literature; previous studies showed that values between 0.17 and 0.25 mm are very common (e.g. Niemczynowicz, 1986). In addition, the rainfall error model was also formulated according to the power based relationship in Equation 7.1:

$$I^* = \pm 30\%\ \alpha N^{\beta} \qquad \text{Equation 7.4}$$

where: I^* is the disturbed rainfall; $\pm 30\%$ represents the error in the nominal volume of a 0.2 mm tipping bucket; and, α and β are parameters depending on the tipping bucket. α and β are values of 0.1848 and 1.047, respectively, adopted from literature (Niemczynowicz, 1986; Molini et al., 2005a). Within this context and assumptions, the error model to account for systematic errors in rainfall data is a function of $t_{drift}, t_{reset}, Vol, \alpha\ and\ \beta$.

In addition, a single offset ±30% was applied to the rainfall data. This constant error was previously adopted by Rauch et al. (1998) and Kleidorfer et al. (2009).

The different rainfall scenarios generated a set, in which the coefficient of variance of the mean annual rainfall was 29.4%, and 32.2% for the mean event maximum rainfall intensity.

7.5.2 Flow error model

Flow random errors and the LPU

The flow measurements used in this research were collected in the pipes located at the outlets of each catchment. As described in Subsection 7.3, the random errors from the radius of the pipe, (Rad) and uncertainty in the water depth (h) and velocity (v) estimates can be calculated based on accuracy of the measuring equipment. For this work, it was assumed that the variables Rad, h and v that are measured with different instruments, are independent and not correlated. Under these conditions, the law of propagation of uncertainty can be used to estimate the combined standard uncertainty $u(Q)$:

$$u(Q)^2 = u(Rad)^2\left(\frac{\partial Q}{\partial Rad}\right)^2 + u(h)^2\left(\frac{\partial Q}{\partial h}\right)^2 + u(v)^2\left(\frac{\partial Q}{\partial v}\right)^2 \qquad \text{Equation 7.5}$$

where

$$\left(\frac{\partial Q}{\partial Rad}\right) = 2\ v\ Rad\ Arccos\left(1 - \frac{h}{Rad}\right) - 2\ v\ Rad\ \sqrt{2Rad\ h - h^2} \qquad \text{Equation 7.6}$$

$$\left(\frac{\partial Q}{\partial h}\right) = -2\ v\ Rad\ \sqrt{2Rad\ h - h^2} \qquad \text{Equation 7.7}$$

$$\left(\frac{\partial Q}{\partial v}\right) = Rad^2\ Arccos\left(1 - \frac{h}{Rad}\right) - (Rad - h)\ \sqrt{2Rad\ h - h^2} \qquad \text{Equation 7.8}$$

The true discharge Q_t has approximately 95% probability of being within the range of $Q \pm 2u(Q)$.

The standard uncertainties in the three variables Rad, h and v are used as previously proposed in the literature by Bertrand-Krajewski and Muste (2007): u($Rad) = 0.002$ m u($h) = 0.003$ m and u($v) = 0.1$ m/s.

To account for random errors in flow data, the disturbed flow values for each timestep are sampled from a normal distribution, in which the mean is the measured Q value with $u(Q)$ as the standard deviation. The random error scenario is made up of 10 different samples.

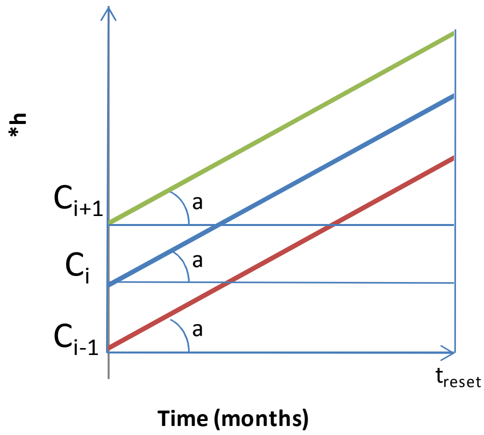
Flow systematic errors

Very little knowledge on the actual systematic measurement error of flow is available. However, it is assumed that the incorporation of a fitted flow error model of calibration data errors is crucial in order to provide reasonable model estimations uncertainty. For the development of a flow error model, the main sources of systematic error in flows data were considered:

As follows, the flow disturbed by systematic errors Q_S^* is calculated using the disturbed height measurement $h^*$and disturbed velocity measurement v^*, which can be written as:

$$Q_S^* = v^* \left[Rad^{*2} Arc\cos\left(\frac{Rad^* - h^*}{Rad^*}\right) - (Rad^* - h^*)\sqrt{2Rad^*\ h^* - {h^*}^2}\right] \qquad \text{Equation 7.9}$$

Estimating h^* - the ‘zero’ point in the height measurement may drift with time. We assume that the equipment is calibrated regularly, and that the probe drift is linear with a rate of a (m/month). The time between successive calibrations is called t_{reset} (months). Although it might be assumed that the readings will re-start from zero error, this zero point cannot be measured with complete accuracy, so a systematic shift C might also result for the period between calibrations.



While $t < t_{reset}$ (t_{reset} = time is reset according to regular calibration)

$h^* = h + (C_i + a.t)$ $\quad$ *i=1, 2,… to* nt_{reset}

Then $t_{reset} = 0$

Figure 7.1 Schematic of the flow error model.

Figure 7.1 presents a schematic of the flow error model, in which h is the measured depth; $h^*$is the perturbed h; C_i is a random number sampled from a uniform distribution between pre-established values every time $t\ t_{reset} = 0$; and, a is a constant representing the slope (i.e. the drift rate, e.g. 0.5 m/month). From practical experience in managing flow gauges, it is known that at a good site a drift rate of 2 mm/month or less might be expected (Prodanovic, 2009); whereas at an 'average' site, up to 10 mm/month is possible and only a 'flawed' site would have worse than this (Fletcher, 2008). As such, these values were adopted for the best and worst case scenarios, respectively.

Velocity - It is not possible to effectively and fully calibrate a Doppler sensor. Therefore a constant and linear noise was assumed. Values of 10% and 30% were used for best and worst case scenarios, respectively.

$$v^* = v + e\,v \qquad \text{Equation 7.10}$$

where e is the constant and linear noise. The flow disturbed by systematic errors Q_S^* is then calculated as a function of C, a, t_{reset}, e.

In addition, a single offset of ±30% was applied to the whole flow time series.

Discussion

A flow threshold was set at 3 L/s; the flow below this value is regarded as non reliable. The mean relative uncertainty in measured flows was 78% (Coefficient of variance of 60%). Although this number is high when compared to the ones in the literature for similar measuring devices (e.g. 5-25% in Ahyerre et al. (1998) and 2-20% in Harmel et al. (2006)) by the low flows which have a large relative uncertainty, mainly due to the high uncertainty in the low velocities measurements. Even when we discard flows lower then 3 L/s, some higher flows had larger uncertainty because of the velocities (it is probably because the uncertainty associated with low velocities is so significant that a large number of velocities were incorrect). In addition, velocities under 0.1 m/s generated relative uncertainties of over 180% independently of the depth, (h). Figure 7.2 illustrates this discussion.

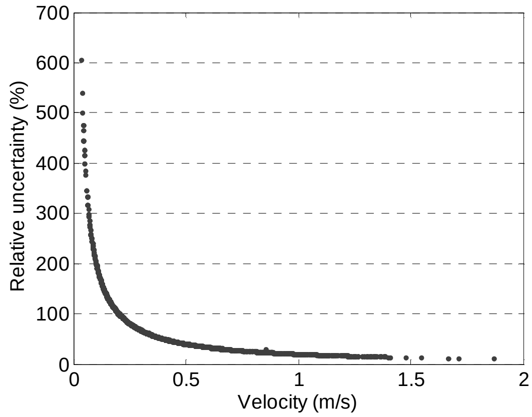


Figure 7.2 Measured velocities in m/s versus the flow relative uncertainty.

The mean event maximum runoff rate generated with the different rainfall scenarios ranged from 0.3 to a little over 3000 L/s and the coefficient of variation for the mean event maximum runoff rate was 42%. Figure 7.3 presents the flow duration curves for the different flow scenarios (mm/min).

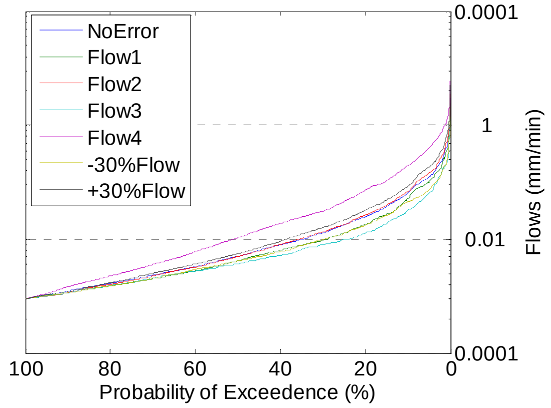


Figure 7.3 Flow duration curves with the different flow scenarios (mm/min). Section 7.6 presents a detailed description of the scenarios.

7.5.3 Discrete samples error model

Based on the values presented in Table 7.1 the error model was developed to account for all sources of uncertainties in TSS concentrations.

Discrete samples random errors

TSS concentrations were disturbed with a value sampled from a uniform distribution in the range of [-0.28, 0.28].

Discrete samples systematic errors

The systematic errors in the discrete samples were accounted for by combining the systematic source values presented in the literature (Table 7.1). The best case scenarios were generated by picking the median uncertainty values and the worst case scenarios were generated by selecting the extreme uncertainty values.

Final values of -9% and +26% were obtained for the best case scenarios and were applied to the entire concentration dataset. Final values of -28% and +50% were the extreme values reported and were used for the worse case scenarios by means of applying these values to the entire concentration dataset. In addition, a single offset of ±20% was applied to the TSS concentrations, in which 20% was the median of the means reported by reviewed studies.

Figure 7.4 presents the Probability of exceedence plots during wet weather at RICH with the different TSS scenarios.

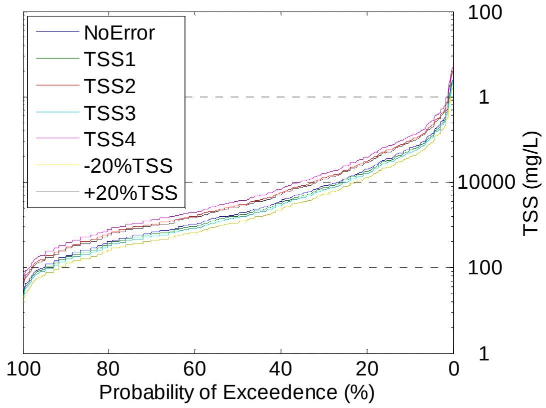


Figure 7.4 Probability of exceedence plots during wet weather at RICH with the different TSS scenarios (mg/L). Section 7.6 presents a detailed description of the scenarios.

7.6 Impacts of measured data uncertainty on urban stormwater models

Impacts of measured data uncertainty on urban stormwater models

C. B. S. Dotto[1]*, M. Kleidorfer[2], A. Deletic[1], W. Rauch[2] and D. T. McCarthy[1]

[1] CRC for Water Sensitive Cities & Monash Water for Liveability, Monash University, Victoria, Australia, 3800

[2] Unit of Environmental Engineering, Faculty of Civil Engineering University of Innsbruck, Technikerstrasse 13, A-6020 Innsbruck, Austria

**Corresponding author. Department of civil engineering, Building 60. Monash University,* [redacted]

ABSTRACT

Assessing uncertainties in models due to different sources of errors is crucial for advancing urban drainage modelling practice. This paper explores the impact of input and calibration data errors on the parameter sensitivity and predictive uncertainty by propagating these errors through an urban stormwater model (rainfall runoff model KAREN coupled with a build-up/wash-off water quality model). Error models were developed to disturb the measured input and calibration data to reflect common systematic and random uncertainties found in these types of datasets. A Bayesian approach was used for model sensitivity and uncertainty analysis. It was found that random errors in measured data had minor impact on the model performance and sensitivity. In general, systematic errors in input and calibration data impacted the parameter distributions (e.g. changed their shapes and location of peaks). In most of the systematic error scenarios (especially those where uncertainty in input and calibration data was represented using 'best-case' assumptions), the errors in measured data were fully compensated by the parameters. Parameters were unable to compensate in some of the scenarios where the systematic uncertainty in the input and calibration data were represented using extreme worst-case scenarios. As such, in these few worst case scenarios, the model's performance was reduced considerably.

KEYWORDS. Input and calibration data; Urban drainage; Modelling measurement errors; Sensitivity analysis; Bayesian inference, Parameter probability distributions; Uncertainties.

1 Introduction and background

Stormwater models underpin the decision making process in urban water management, policies and regulations. Moreover, they are key tools for the quantification of urban discharges and also for the design of stormwater treatment technologies. Uncertainties, however, are intrinsic to all models and it is hypothesised that the level of accuracy of any model's output is often compromised if the different sources of errors are not considered during the modelling exercise. Therefore, assessing uncertainties in models due to different sources of errors is crucial for advancing urban drainage modelling practice. Typically, three sources of random and systematic uncertainties are identified: errors in the measured input and calibration data, and errors due to incomplete or biased model structure (Butts et al., 2004). While the uncertainty in the calibrated parameter values combines the different sources, the impact of calibration and uncertainty analysis methods, different objective functions and calibration data availability on the model sensitivity are also recognised (Mourad et al., 2005; Dotto et al., 2012; Kleidorfer et al., 2012).

As with most models, the calibration of urban drainage models rarely results in one unique parameter set, and instead many equally plausible parameter sets are obtained, which reduces the confidence in the models when they are used for prediction (Kuczera and Parent, 1998). The uncertainty related to the model calibration parameters and its impact on the model outputs has been extensively studied (e.g. Kanso et al., 2003; Feyen et al., 2007). Global sensitivity analysis methods have been applied to estimate the confidence intervals around the model's prediction while revealing the sensitivity of the model outputs to each parameter (e.g. Feyen et al., 2007; Yang et al., 2008). Many methodologies are available to conduct these uncertainty/sensitivity analyses, including informal Bayesian methods (e.g. GLUE by Beven and Binley (1992)) and formal Bayesian approaches (e.g. MICA by Doherty (2003) and DREAM by Vrugt et al. (2009)). Comparisons have been made between these methods in various research areas (e.g. Yang et al., 2008; Matott et al., 2009), including urban drainage modelling (Dotto et al., 2012). These comparisons suggest that modellers should choose the method which is most suitable for the system they are modelling (e.g. complexity of the model's structure including the number of parameters), their skill and knowledge level, the available information, and the purpose of their study.

Measured data such as rainfall, flow rates and pollutant concentrations are needed for the application of urban drainage models. While rainfall data is the main input for most urban drainage models, flow rates and pollution concentration data are required for model calibration and validation. These measured datasets have inherent uncertainty and it has been shown that this uncertainty increases the data requirements for model calibration (Mourad et al., 2005). The input data used in stormwater modelling could be highly uncertain. For example, the main sources of uncertainties in rainfall intensities, commonly measured using tipping bucket rain gauges, are related to both rainfall catching and counting errors (Molini et al., 2005b). While splashing losses were found to be only up to 2% and evaporation losses were up to 4%, the wind losses were found to be inversely proportional to the rain intensity and were up to 30% for rainfall intensities around 0.25 mm/h (Sevruk, 1982; Rauch et al., 1998). Battery, logger and computer clock failures are also significant source of errors in rainfall measurements. For example, time drifts are inherent to any battery controlling logging device and values around 0.07 min/day were reported by McCarthy (2008). The spatial variability of rainfall often is a large source of errors when point source measurement methods are used (such as tipping bucket gauges). To address this issue radar rainfall data can be used to estimate precipitation, but radar data is also subject of several assumptions that introduce a number of errors (Einfalt et al., 2004).

While addressed in related fields (e.g. hydrologic models: Krzysztofowicz and Kelly, 2000; Haydon and Deletic, 2009), the impacts of input data uncertainties on urban drainage models are largely unknown. Only a few studies evaluated the propagation of input data uncertainties through urban drainage models (Rauch et al., 1998; Bertrand-Krajewski et al., 2003) and in all of them, the models were first calibrated assuming that measured inputs and outputs are without error, and the impacts of input data uncertainties were then propagated through the models, while keeping the model parameters fixed. Kleidorfer et al. (2009) developed this further by assessing the impact of input data uncertainties on model parameters and found that the parameters of both flow and pollution models were influenced by systematic errors in input data.

In addition, the techniques used to measure urban discharges and associated water quality parameters, that are needed for calibration of stormwater models, also contain error (Bertrand-Krajewski et al., 2003; Harmel et al., 2006; McCarthy et al., 2008). For example, uncertainties in stormwater flow data, commonly measured using velocity-area measurement method, range from 2% to 20% (Harmel et al, 2006). While these random errors can be estimated, uncertainties in flow measurements due to systematic errors (often related to the height measurement and inaccurate velocity calibration or incorrect probe set-up) were not explored (Harmel et al, 2006).

Errors in water quality data are far larger than for flows or rainfall. Sampling, storage and analytical/laboratory methods all have inherent errors which contribute to the uncertainty in the final sample's pollutant concentration(Harmel et al., 2006). While sampling errors, related to the position of the probe, are significant in TSS measurements, with values up to 33%, they are not significant for dissolved pollutants that do not settle (Harmel et al., 2006). Some dissolved pollutants are more impacted by storage uncertainties; values up to 49% were reported for TN even for samples which are kept iced and are analysed within 6 hours (Kotlash and Chessman, 1998). Uncertainty related to the laboratorial analysis was less explored, but values from -9.8 % to 5.1 % have been reported for TSS (Harmel et al., 2006). Although these uncertainties are acknowledged in the urban drainage field, the impact of them on stormwater models has not been explored.

In addition, the combined impact of input and calibration data on urban stormwater models is unknown. However, valuable information can be obtained from related studies on modelling of large natural catchments. For example, Renard et al. (2008) and Thyer et al. (2009) applied the Bayesian Total Error Analysis methodology (BATEA proposed by Kuczera et al., 2006) to evaluate the uncertainties in hydrological models arising from model input, output and structural errors. The BATEA framework is based on hierarchical Bayesian models and is very comprehensive and transferable (Renard et al., 2008). However, it is rather difficult for application, since it requires a large number of extra calibration parameters (that are associated with modelling the errors), is computationally demanding, and requires a significant level of understanding of the tested model structure (Renard et al., 2008).

In summary, the combined effect of input and calibration data uncertainty on the parameters and outputs of urban drainage models has not been explored. Recently, the International Working Group on Data and Models of the Joint Committee on Urban Drainage that works under IWA and IAHR proposed an overarching framework that could address this issue (Deletic et al., 2012). However, the framework has never been tested, lacking practical details on the methodology. This paper is the first attempt to test the proposed framework for assessing the impact of both input and calibration data errors on the parameter sensitivity and predictive uncertainty of an urban rainfall runoff and water quality model using a rich Melbourne dataset.

2 Methods

2.1 Adopted stormwater models

Rainfall runoff model. KAREN (Rauch and Kinzel, 2007) was selected for the study because of its simplicity and proven performance for urbanised catchments (Kleidorfer et al., 2009). KAREN is a linear reservoir model, which only requires the catchment area and a rainfall time series as inputs to generate a series of flows originating from impervious areas only. The effective impervious area of the catchment is calibrated as the *EIF* parameter. Runoff from impervious areas occurs after a rainfall threshold has been exceeded (calibration parameter *li*). The initial loss is calculated continuously and fills during rainfall and is drained during dry weather by a permanent loss calibration parameter (*ev*). Surface runoff volume is calculated using the linear time-area method, which is related to the unit hydrograph method (Sherman, 1932). At the beginning of a rainfall event, the effective impervious area is increased according to the flow time on the catchment surface until the whole catchment contributes to runoff after the catchment's time of concentration (calibration parameter *TOC*).

Water quality model. A very well researched and widely adopted build-up and wash-off model (initially proposed by Sartor and Boyd, (1972) was used to model TSS concentrations in catchments discharges. It was selected because of its widespread use in practice; e.g. it is used in SWMM (USEPA, 2007). The original model was slightly modified and hence the key equations are presented in Table 1 (formatted for a 6 min timestep).

Table 1 The governing equations of the build-up/wash-off model.

The main modification from the original is in the wash-off stage. The concentration of pollutants in the runoff within a timestep (C in mg/L) is a power function of the catchment runoff modelled with KAREN (q in mm/h) divided by the catchment runoff coefficient (RC - here assumed as the *EIF* calibrated with KAREN). RC was included to represent wash-off only from impervious surfaces, which is a safe assumption because majority of runoff from urban catchments are originated from impervious surfaces (Chiew and McMahon, 1999). If instead of q rainfall intensities were used, the model would have to include a routing algorithm (e.g. linear reservoir routing) resulting in an additional parameter(s). A transport related parameter (r) was used to represent the small lag time which is often noted between the hydrographs and the pollutographs (Vaze and Chiew, 2003). The amount of pollutants washed from the surface (W in kg) is then calculated in function of the predicted concentration and the volume (Vol in L). In total, there are 5 calibration parameters: M_0, (kg), k_1 (day^{-1}); k_2; k_3; and, r(timesteps).

2.1.1 Catchment and Dataset

An urban catchment, located in Richmond, an eastern suburb of Melbourne, Australia, was used. The site has a total area of 89 ha, the land use is high-density residential with a total imperviousness of 74% and an average slope of less than 0.1%. The catchment is drained by a separate stormwater system.

Rainfall is measured using a standard 0.2 mm/tip tipping gauge located 600 m from the catchment centroid. Flows were measured with the American Sigma/HACH area-velocity 950 sensor (HACH, 2008) installed in the outlet pipe. The water quality samples (TSS) were also collected at the outlet of the catchments by autosamplers using flow-based intervals and each sample being analysed for TSS. Details on the catchment, monitoring program and the datasets are available in Francey et al. (2010). Rainfall and flow data collected between 2004 and 2005, and 44 TSS pollutographs (approximately 250 samples) were used for calibration. The event total rainfall ranged from 2 to 60 mm, the mean maximum event runoff rate was 547 L/s and the average of the TSS Event Mean Concentration (EMC) was 125 mg/L (The EMCs were calculated using the discrete TSS concentrations with their associated volumes as per Leecaster et al., 2002). Rainfall and flow data collected from 2006 to 2007 were used for model validation; the event total rainfall ranged from 2 to 44 mm and the mean maximum event runoff rate was 212 L/s. The calibration of the water quality model resulted in very low performance, and therefore validation of this model would not succeed and was not carried out.

2.2 Assessing global uncertainties

The proposed framework (Figure 1) is a further development of the general framework proposed by the International Working Group on Data and Models (see Deletic et al, 2012). Firstly, the model is fed with a certain set of input data (X). The model then generates its outputs (Y) as a function of X and a set of calibration parameters (θ). By means of an appropriate uncertainty analysis method and a certain objective function (OF), the model is run repeated times until the misfit (ε) between the measured data (O) and the modelled data (Y) is reduced. Through this process, the parameter probability distributions (PDs) are generated. Finally, the model predictive uncertainty bands are obtained. To test the influence of input and calibration data uncertainties on the modelling procedure, both datasets were disturbed using error models. The classical approach is again employed, but this time using these disturbed datasets (disturbed input X* and calibration O* data). The parameter distributions and the predictive uncertainty bands produced using these disturbed datasets are then compared with those produced when using the undisturbed data; the differences are used to assess the impacts of these errors. To become operational, this untested/unapplied framework, had to be developed further as explained in the consequent sections.

Figure 1 Propagation of input and calibration data errors through models (after Deletic et al, 2012).

2.2.1 Method used for uncertainty analysis

A Bayesian approach was selected for evaluation of the model parameter sensitivity, since it has some important advantages when used for stormwater modelling (Dotto et al, 2012). The PDs of model parameters were generated using the outputs of the software package MICA (Doherty, 2003), which uses a Markov Chain Monte Carlo (MCMC) method with the Metropolis-Hastings algorithm sampler. The likelihood function adopted in MICA is least square based and assumes that the residuals between the measured and modelled values have a normal distribution (e.g. Feyen et al., 2007); as such, the measured calibration and modelled data series were transformed using a Box-Cox transformation (Box and Cox, 1964) to achieve homoscedasticity and normally distributed residuals. However, all transformation methods change the content of the observations (Beven et al., 2008), which then influences the emphasis on various parts of the hydrograph (or pollutograph). This is sometimes not desired if the modelling purpose is to focus on specific parts of the dataset (e.g. flood prediction is linked with peak flows, which are deemphasised when using Box-Cox transformations) (Doherty and Welter, 2010). Furthermore, all observed data have uncertainty, and this should be taken into account in the likelihood function so that the parameters are estimated appropriately; indeed, it is important that the function places more emphasis on data which has lower uncertainty. Weighting strategies can be used to re-adjust how the likelihood function emphasises various parts of the dataset to (1) consider measured data uncertainty and (2) compensate for the Box-Cox transformation which may have adjusted the emphasis in an undesirable way. Therefore, the weights were computed based on the inverse of the relative uncertainty in the measured (untransformed) data (i.e. the relative error in the measured flow rates calculated using the Law of Propagation of Uncertainties; see McCarthy et al., 2008 for more information). The serial correlation between the data points was not considered because the current used methods to account for autocorrelation (e.g. first order model) are not effective for small timesteps as used in this study (e.g. Yang et al., 2008; Métadier and Bertrand-Krajewski, 2012).

The performance of the model was evaluated using the Nash-Sutcliffe efficiency criterion (E) (Nash and Sutcliffe, 1970) corresponding to the minimum least square value achieved with MICA; it is noted that E was calculated using data which had been back-transformed to the original data-space and ignoring the weighting strategy.

The predictive bounds resulting from parameter uncertainty were obtained by running the models with parameter sets generated with MICA for each of the specific scenarios. The total predictive uncertainty was computed according to Feyen et al. (2007). In the data transformed space, the standard deviation of the model errors is assumed constant and is obtained from the root mean square error (RMSE) between the transformed observed and simulated values using the most likely parameter set. The total uncertainty was estimated by adding this constant Gaussian error (equal to ±1.96×RMSE) to the transformed predictions at each timestep. The obtained prediction limits in the transformed space were then back-transformed to the original data-space.

In order to provide some information about the potentiality of the rainfall runoff model to predict data outside the calibration period, some validation was performed. Specifically, to investigate how the model, calibrated for unbiased and biased data, perform in predicting unbiased data. As such, the model was run for specific scenarios with the parameter sets generated for each of the specific scenarios. The data in this simulation was used without introducing any error and, the Nash-Sutcliffe coefficients, corresponding to validation period, were computed by comparing the modelled data using the validation data without introducing any errors. Similarly, the predictive uncertainties were computed as described above, except that the validation dataset was used without adding any bias. This was done to provide some insights on how the observed validation data can be covered (or not) by the predictive uncertainty bands (generated with the parameter distributions) obtained for the calibrated period under the different scenarios.

2.2.2 Error models of input and calibration data uncertainty

Error models were developed to disturb the measured input and calibration data (i.e. to generate X* and O* in Figure 1) by reflecting common systematic and random uncertainties found in these types of datasets. Multiple scenarios were developed using these error models, ranging from best-case scenarios (i.e. low errors, corresponding to adequate instrumental calibration) to worst-case scenarios (i.e. high errors, when instruments are rarely checked or calibrated). These scenarios, shown in Tables 2 and 3, are then used to assess the impact of input and calibration data uncertainties on model performance, parameter sensitivity and predictive uncertainty.

Error models for the rainfall runoff modelling

Input (rainfall) data error models. The following main sources of errors (as reported in literature) have been taken into accounts: (i) the stochastic nature of the measurement instrument (Rauch et al., 1998; Sevruk, 2002; Molini et al., 2005a; Haydon and Deletic, 2009); (ii) the systematic mechanical errors associated with the measurement device (e.g. Simic and Maksimovic, 1994), which was reported to induce an error of -10 to -15% for rainfall intensities higher than 200 mm/h (Molini et al., 2005a); and (iii) the randomness caused by the variability of rainfall over a catchment area and using a single gauge to represent this variability (Chaubey et al., 1999).

Combining these sources, eight scenarios were developed as reported in Table 2. The '*Random error only scenario*' was developed to simulate the effects of only random errors on the modelling process. To apply this scenario, each data point in the measured rainfall dataset was disturbed by adding a random number between -0.5 mm and 0.5 mm (i.e. sampled from the uniform distribution [-0.5, 0.5]). Ten disturbed rainfall datasets were generated using this procedure - RREr (x10). The '*Random spatial distribution*' was used to understand spatial errors areal approximations of rainfall using one rainfall gauge and was applied by multiplying all rainfall measurements within an event by a random number in the range of [-0.7, 1.3]; this number was kept constant within an event, but was re-sampled between events. Again, this procedure was used to create ten disturbed rainfall datasets - RainSD (x10). The '*Systematic constant offset*' was used to simulate a worst case scenario, where systematic errors are not corrected during the monitoring program and hence applied to all of the measured dataset (e.g. incorrectly calibrated bucket which was never re-calibrated, incorrect positioning of the rainfall gauge under a tree, etc.). Two disturbed rainfall datasets were generated: one where all rainfall rates in the measured dataset were reduced by 30% and the other where the rainfall rates were increased by 30%.

The '*Rain gauge maintenance scenarios*' were used to simulate 'good' and 'poor' calibration methods, including: the time period between re-calibration (t_{reset}), the direction of the time drift of the rain gauge logger between the re-calibration (t_{drift}) and the re-calibration error for the bucket volumes (bucket volume error). The error in the 0.2 mm nominal volume of the tipping bucket used in this study was represented by an offset of ±30%. The disturbed rainfall intensity was estimated according to the relationship between the rainfall intensities and the number of tips:

$$I^{*} = \pm 30\% \ \alpha \ N^{\beta} \qquad \text{Equation 1}$$

where I^{*} is the disturbed rainfall intensity; ±30% represents the error in the nominal volume of a 0.2 mm tipping bucket; N is the tipping rate (number/min); and, α and β are parameters depending on the tipping bucket. Here these values were assumed according to the literature (Simic and Maksimovic, 1994; Molini et al., 2005a; Molini et al., 2005b) and were equal to 0.185 and 1.047, respectively. Four disturbed datasets were generated for these scenarios (Table 2). As an example, for the Rain1 scenario, the disturbed dataset was generated by applying the -30% form of Equation 4, applying a time drift of 0.14 min/day and considering that the rain gauge was re-calibrated every month (which characterises the best case scenarios). In total, the eight input rainfall error scenarios generated 26 disturbed rainfall datasets, which were then applied to the modelling procedure outlined in Figure 1. These were compared with the base-case scenario when using the raw measured rainfall data (i.e. no errors).

Table 2 Summary of the tested error scenarios for the KAREN rainfall runoff model.

Calibration (flow) data error models. The measured flow data was also disturbed by both random and systematic errors to create O*, using seven scenarios (Table 2). The uncertainty sources associated with the Doppler area-velocity method are: measuring the channel's cross sectional area, depth and velocity (Harmel et al., 2006). For the 'Random error only' scenario, each measured flow in the time series was disturbed by a random error term derived using the Law of Propagation of Uncertainty (fully described in Taylor and Kuyatt, 1994) as per McCarthy et al. (2008). As with the input data, ten disturbed datasets were generated for this scenario - FREr (x10). '*Systematic constant offset*' was considered a worst-case scenario, where the flow measurements were always either underestimated or overestimated by ±30% (Harmel et al., 2006) and were not corrected for the entire monitoring period (hence producing two disturbed calibration datasets). The '*Flow gauge maintenance scenarios*' were developed to test the influence of 'good' and 'poor' calibration and maintenance regimes, including the systematic effects of: the level drift of the instrument (assumed to occur linearly with time; ±2 and ±10 mm/mth for best and worst case scenarios, respectively; Harmel et al., 2006), which occurs between the maintenance interval (t_{reset}), the error which occurs when the level sensor is recalibrated at each t_{reset} (±5 and ±20 mm for best and worst case scenarios, respectively; Harmel et al., 2006) and the velocity error which would occur if the probe is incorrectly positioned within the pipe (±10% and ±30% for best and worst case scenarios, respectively). Four disturbed datasets were generated for these scenarios; the disturbed water depth was calculated according to the equation:

$$h^*(t) = h + (C + a.t) \qquad \text{Equation 2}$$

where h^* is the disturbed water depth in mm; h is the measured water depth in mm; C is the level drift in mm; a is the drift rate in mm/mth; and, t is the time in months, after re-calibration; As an example, Flow1's disturbed dataset was generated by applying the following equation:

$$h^*(t) = h + (U[-5mm, +5mm] - 2t) \qquad \text{Equation 3}$$

and applying constant and linear noise of -10% to the measured velocities. In this scenario it is assumed that level sensor is recalibrated every 1 month. For the seven scenarios tested, 16 disturbed datasets were generated and propagated through the process outlined in Figure 1. These were compared with the base-case scenario when using the raw measured flow data (i.e. no errors).

Combined input (rainfall) and calibration data (flow) error models. Rainfall (input) and flow (calibration) data scenarios were combined to evaluate their joint impact on the model sensitivity and uncertainty (Table 2). '*Systematic constant offsets*' scenarios were generated by combining the ±30% rain and flow scenarios. '*Rain and flow gauges maintenance scenarios*' were developed by combining the individual rain and flow worst-case scenarios.

Error models for the water quality modelling

Input (modelled flow) data error models. The input data required for the water quality model is a time series of KAREN's modelled flows. Ten scenarios were developed to incorporate the possible errors of these modelled flows. The '*Input parameter error scenario*' was used to understand how various parameter sets which could equally calibrate KAREN impact on the water quality model. As such, instead of using only the 'optimal' parameter set as input in the water quality model, six sub-sets of KAREN parameters sampled from their PDs determined by Bayesian inference of the rainfall runoff model were used to model flows that were subsequently used as inputs for the water quality model (WQPar (x6) in Table 3).

For the '*Random error only*' scenario, 10 sets of modelled flows with KAREN when using the most likely parameter sets obtained in each of the 10 realisations for the rainfall random error (RREr (x10) in Table 2) were used – WQMFREr (x10). '*Systematic constant offset*' scenarios were developed using the worst-case scenarios from KAREN testing, where the modelled flows using the four disturbed calibration datasets generated with the combined rain and flow ±30% scenarios were used. '*Modelled flows with rain and flow gauges maintenance*' scenarios were used to test the influence of the systematic effects and inappropriate calibration/maintenance of measurement devices associated with the modelled flows. For that, the flows modelled using the most likely parameter values obtained with the four '*Rain and flow gauges maintenance scenarios*' in Table 2 were used as input to the water quality model.

Table 3 Summary of the tested error scenarios for the water quality model (see Table 2 for further explanations for some model errors).

Calibration (TSS) data error models. The measured discrete TSS samples were disturbed by both random and systematic errors, using seven scenarios (see Table 3). '*Random error only*' was estimated according to the uncertainty values presented in literature (e.g. Bertrand-Krajewski et al., 2003) and the TSS concentrations were disturbed with a value sampled from a uniform distribution in the range of [-0.28, 0.28] (mg/L). As with the previous data, ten disturbed datasets were generated for this scenario – TSSREr (x10). Two '*Systematic constant offset*' TSS data sets were generated to test the influence of systematic errors associated with the TSS measurements (e.g. positioning of the sample suction tubing placed at either the top or bottom of the water cross-section). In these scenarios, measured TSS concentrations were either all reduced or increased by 20%. The '*Discrete samples – combining all the systematic sources*' scenario was developed by combining the values reported in the literature for the key error sources (e.g. Harmel et al., 2006; Rode and Suhr, 2007; McCarthy et al., 2008); the best case scenarios were generated by compiling mean values (final values of -4.9% and +25% were applied systematically to the entire concentration dataset) and the worst case scenarios were generated by compiling the extreme reported values (final values of -9.8% and +40% were applied to the entire concentration dataset).

Combined input (modelled flow) and calibration (TSS) data error models. Four '*Systematic constant offset*' scenarios were created (Table 3) where KAREN's modelled flows were generated with rainfall (input) and measured flow (calibration) data with a systematic error of ±30%, while the TSS were underestimated or overestimated by 20%. The '*Modelled flows with rain and flow gauges maintenance scenarios combined with systematic constant offset*' were generated to assess impact of systematically faulty rain and flow measures (poorly calibrated gauges) combined with TSS errors (Table 2 and Table 3 explain the tested combinations).

3 Results and discussion

The overall efficiency of both the rainfall runoff and the water quality models, for calibration period, is represented by the Nash-Sutcliffe efficiency criterion (E) in Table 4 and 5, respectively. These values correspond to the optimised MICA parameters, which were obtained when the minimum weighted least-square likelihood function was achieved (E_{WLS}). The maximum Nash-Sutcliffe efficiency values recorded by the MICA process is also noted in these tables (E_{accep}); while these do not represent the 'best' parameter set according to the chosen likelihood function (i.e. weighted least squares - WLS), it still represents a 'behavioural' parameter set.

The E_{WLS} value for KAREN with undisturbed input and calibration datasets (NoError) was 0.56; when the input and calibration data were disturbed, this efficiency was maintained in all except, five scenarios. As such, none of the data errors has effect on the model performance (in case of pure random errors) or the model can compensate for most of the input and calibration data error scenarios tested; for example, the +30%R-30%F scenario (where rainfall was overestimated by 30% and flow was underestimated by 30%) yielded the same model efficiency as when using the undisturbed dataset. This demonstrates the flexibility of these models to compensate for systematic errors (i.e. model parameters are adapted). When the uncertainty becomes too large (i.e. worst-case scenarios associated with significant drifts in the data), KAREN is no longer able to compensate hence producing significantly lower E values; e.g. for the Rain4Flow4 combination scenario (Table 2) which does not only contain a constant systematic error but also a growing error between maintenance intervals which cannot be compensated by parameter adaption.

The E_{WLS} for the build-up wash-off model were consistently low, independent of whether or not the input or calibration data were disturbed (E_{LS} between 0.12 and 0.19). These low values might reflect structural errors in the model, which limits the model's performance. There were just three scenarios which yielded a significantly lower efficiency values, all again representing worst-case situations where the model is no longer able to compensate for large measurement errors.

Table 4: Overall efficiency for the different scenarios of rainfall runoff model KAREN - E_{WLS} stands for the maximum efficiency corresponding to the minimum weighted least-squares; E_{accep} stands for the maximum efficiency from all the MICA accepted parameter sets.

Table 5 Overall efficiency for the different scenarios of the build-up/wash-off stormwater quality model - E_{WLS} stands for the maximum efficiency corresponding to the minimum weighted least-square; E_{accep} stands for the maximum efficiency from all the MICA accepted parameter sets; RC stand for the runoff coefficient.

The overall model efficiency varied in the validation data period. The fact that the model performed different in the validation period is not surprising and can be explained by different factors related, for example, to the form of the likelihood function used for calibration associated with different climate conditions during calibration and validation period, and to the level of parameter compensation in the different scenarios. Considering the NoError scenario, the Nash-Sutcliffe efficiency obtained with the parameter set corresponding to the minimum weighted least-square likelihood function (in the calibration period - E_{WLS}) dropped from 0.58 in the calibration to 0.52 in the validation. And this could be explained by the fact that the validation data period was much drier then calibration data period and the model that was calibrated with a least square likelihood function could not predict the lower flows from the validation data period (please refer to Dotto el al, 2011 for further discussion). An example of a more drastic drop was found for the -30%R+30%F scenario, in which the same efficiency measure dropped from 0.58 to 0.1. This drastic droop can be explained by the fact that the model that was calibrated to deal with a much higher amount of runoff (in -30%R+30%F) was obviously not able to predict the low flows observed during the validation period. The opposite was observed for Rain4, in which the model was calibrated to less runoff and therefore was able to better predict the lower flows corresponding to the validation period (in which the Nash-Sutcliffe increased from 0.4 to 0.5).

3.1 Model performance and parameter sensitivity

The results, for both rainfall runoff and water quality models, are presented according to the impact of the error types (i.e. No error, Random error and Systematic error) on each of the calibration parameters of the two models. For each error type, we start by discussing the impacts of input data errors, then calibration data errors, and finish by discussing the impacts of the joint propagation of input and calibration data errors.

3.1.1 No error analysis

The PDs of parameters revealed that KAREN is quite sensitive to the effective *EIF*, *TOC* and *ev*, while less influenced by *li* (see Figure 2), which is in line with previous studies (Dotto et al, 2012). The build-up/wash-off model is sensitive to M_0, k_2, k_3 *and* r, but not very influenced by k_1. In addition, it was verified that M_0 and k_2 are very correlated mainly due to the model structure arrangement as previously described by Kanso et al. (2003).

3.1.2 Random error impact in input and calibration data

Rainfall runoff model - KAREN. Random errors in measured input and calibration data did not significantly impact on the model's performance (Tables 4 and 5) and sensitivity. For example, E_{WLS} ranged from 0.55 to 0.58 for the 10 sets of data generated with KAREN rainfall random errors. The PDs of KAREN parameters for the rainfall random errors scenario have very similar shape as those for the 'NoError' scenario (Figure 2, top). The same was observed when disturbing the rainfall input for spatial distribution errors (RainSD) and the flow data for random errors in measurements (FREr). While Kleidorfer et al. (2009) propagated the random input data errors in the same stormwater model in a slightly different way, the conclusion, that random errors do not represent a major impact in the model sensitivity, was the same. For this reason, the random errors in both input and calibration data were not assessed further (i.e. were not combined with the systematic errors).

Water quality model - Build-up/wash-off. Similarly, the build-up wash-off model PDs were unchanged when applying random errors to the input and calibration data (Figure 2, bottom). Again, Kleidorfer et al. (2009) found similar results when analysing TSS and TN loads with a simple regression equation. As such, these errors were not combined with the systematic errors.

Figure 2 Histograms for KAREN parameters (top) obtained from the 'NoError' scenario and from the 10 sets of data generated with KAREN rainfall random errors – RREr (x10); and histograms for the build-up/wash-off model parameters (bottom) obtained from the 'WQNoError' scenario and from the 10 sets of data generated with TSS random errors – TSSREr (x10) (see Table 2 for abbreviations).

3.1.3 Systematic errors in input and calibration data

Rainfall runoff model - KAREN. Impact of systematic errors on parameter sensitivity (visualised in the parameter distribution) is shown in Table 6. When rainfall is systematically overestimated (e.g. +30%Rain), or flow is underestimated (e.g. -30%Flow), *EIF* decreases or increases, (to compensate for these errors), respectively. These shifts were either compensated or accentuated when combining these rainfall and flow errors; for example, in the -30%R+30%F and +30%R-30%F scenarios, pronounced shifts in *EIF* were observed, while in the -30%R&F and +30%R&F scenarios little differences were seen in the PDs as compared with the 'NoError' scenario. These results reflect the ability of the model parameter *EIF* to entirely compensate for systematic errors found in the measured datasets. Similar trends were observed for the *EIF* PDs in the sensor maintenance scenarios; when the maintenance scenario overestimated rainfall (Rain2 and Rain 4), or underestimated flow (Flow1 and Flow3), the resultant *EIF* PD was shifted to accommodate these changes. There is little difference between worst-case and best-case rainfall sensor maintenance scenarios (i.e. PDs are the same for Rain1 vs. Rain 2, Rain2 vs. Rain 4), probably because these scenarios only differ by their recalibration frequency (1 month versus 6 months). Larger differences were seen in PDs for best and worst case flow sensor maintenance scenarios, most likely because of the differences in the magnitude of the systematic velocity errors. *EIF* was also adjusted when both input (rainfall) and calibration (flow) data errors were combined. For example, if both rain and flow were underestimated by 30% (-30%R&F), *EIF* kept fairly constant reflecting the combined reduction in rain and flow volumes. This is an example of how different sources of uncertainties compensate for each other, which is again similar to the results found by Kleidorfer et al. (2009).

Other parameters were less influenced than *EIF*, perhaps because *EIF* is the most influential parameter in KAREN. *TOC* was not influenced when only the rainfall data was systematically disturbed by ±30% (which is not surprising since no timing errors were introduced by these scenarios). Contrarily, the *TOC* PDs were impacted in all scenarios where flow was systematically disturbed. For the rainfall gauge sensor maintenance scenarios, longer periods between maintenance events (i.e. higher t_{reset} = 6 months) resulted in larger time drifts in the measured data which is therefore reflected by the change in the *TOC* parameter. If the rainfall logger has a slow clock (i.e. Rain1 and Rain3) *TOC* increases to account for this drift (and vice-versa in Rain2 and Rain4). The best-case flow sensor maintenance scenario had little impact on *TOC*, again because timing issues were not introduced. The worst-case flow sensor maintenance scenarios did impact *TOC*, probably because the incremental error added/subtracted to/from the level measurements altered the measured hydrographs which changed the probability distribution of *TOC*. When both input and calibration datasets were systematically disturbed, the resulting PDs are similar to the ones obtained for the flow scenarios.

The initial loss (*li*) reveals PDs that are distributed among realistic values for this parameter (i.e. values between 0.5 and 2 mm are usually adopted in similar urban storm models; (e.g. Chiew and McMahon, 1999). Introducing systematic errors into the rainfall data (i.e. ±30%Rain) slightly influenced the *li* PDs, while systematic flow errors (i.e. ±30%Flow) seemed to have no observable influence. When both rainfall and flow datasets were systematically disturbed simultaneously (i.e. ±30%R&F, -30%R+30%F and +30%R-30%F), the resultant *li* PDs resembled those of the rainfall only scenarios. Similar results were found when testing various maintenance scenarios; overestimated rainfall led to higher *li* values and vice-versa for both best and worst case scenarios. While the best-case flow sensor maintenance scenarios (Flow1 and Flow2) yielded no effect on PDs of *li*, the worst-case scenarios, including the combined input and calibration data errors (e.g. Rain4Flow4) significantly impacted them. It is hypothesised that these results are directly linked to *EIF*'s behaviour, which compensated for the dramatic increase in flow measurements in Flow4 and its related scenarios; indeed, it is common to see a relationship between *EIF* and *li*, as they are intrinsically linked in the model structure (Boyd et al., 1993).

The *ev* parameter reflected some of the model limitations. While common values for evaporation in Melbourne are around 4 mm/day, the parameter PDs were in the range of 0.01 and 0.15 mm/day for most of the scenarios. In fact, *ev* was expected to increase when the rainfall was overestimated (+30%Rain, Rain2 and Rain4). While the +30%Rain scenario does illustrate such change, Rain2 and

Rain4 scenarios indicated that when rainfall was overestimated and the time drift was introduced in the data error more parameter interaction was identified (mainly between *ev* and *li*) and the calibrated values were far from the expected ones. In general the shape of *ev* PDs, and most likely values, did not change much within the various scenarios.

In summary, the results from this sensitivity analysis showed that the systematic errors in input and calibration data (individual or combined) influenced most of the parameters' distributions. It also demonstrated that the parameters were able to compensate for most of the scenarios, mainly for the best case ones. The fact that the model parameters' compensate for most rainfall errors was also found by Younger et al. (2009) when evaluating the impact of rainfall data error in a hydrologic rainfall runoff model. They found that while the peak flows changed significantly for the different scenarios, the model performance did not. Furthermore, the propagation of input and calibration data errors provided new information about the model structure and its parameters.

Table 6 Rainfall runoff model KAREN - Histograms of parameter PDs for the different error scenarios. The variation in the x axis is to facilitate the visualisation of specific scenarios.

Water quality model - Build-up/wash-off. The impact of systematic errors on parameter sensitivity is shown in Table 7. Results from the input data error scenarios showed that M_0 is directly linked to the runoff volume; its PDs shifted towards higher values in all input data error scenarios where modelled flows were overestimated (WQ+30%R&F WQ-30%R+30%F, WQRain3Flow4 and WQRain4Flow4) and towards lower values when modelled flows were underestimated (-30%R&F WQ+30%R-30%F, WQRain3Flow3 and WQRain4Flow3). Similarly, M_0 was sensitive to calibration data error, and varied according to the increase (or decrease) in the TSS concentrations for each scenario. The impact of combined input and calibration data errors on M_0 parameter indicated a clear 'additive' effect for different sources of uncertainties. For example, the combined increase in both volumes and TSS concentration (WQ+30%RF+20%TSS) shifted M_0 for the 10% difference as the model does not need to compensate for the +20% in both data. Following the same pattern, M_0 shifts the most when the volumes are increased and TSS decreased (WQ+30%RF-20%TSS).

The PDs for the k_1 parameter were skewed towards the upper limit of the parameter value (i.e. 5 day^{-1}) for most of the scenarios, which reflects the modelling boundary conditions (i.e. model structure limitation). However, testing conducted with MICA confirmed that this parameter will always skew toward the upper bound of its boundary condition (data not shown). As values of k_1 higher than 5day^{-1} are unrealistic (i.e. quick build-up of mass on the surface), 5 was still used as the upper bound. These results suggest that the build-up of TSS occurs extremely quickly, which confirms that the wash-off during typical rainfall events remove only a very small portion of solids accumulated in the surface (e.g. Vaze and Chiew, 2003). Results from the scenarios where Flow4 was included were different; i.e. the model was insensitive to k_1 indicating that the errors were compensated by the other parameters. It is hypothesised that these results would be different if other pollutants were being modelled, for example the antecedent conditions in the catchment influences the amount of pathogens available on the surface (McCarthy et al., 2011).

Results from the individual input and calibration data error scenarios, and also from the combined error scenarios indicated that the peaks in the PDs of k_2 shifted to the opposite direction of shift in M_0 PDs. Again, the high correlation between these two parameters is illustrated. Moreover, the model sensitivity to k_2 (Table 7) confirms the wash-off variability within the event, as larger amounts of TSS are washed in the beginning of the event.

PDs of k_3 varied with the input data scenarios and did not change at all for the calibration data error scenarios. This is because the k_3 parameter is linked to the input data in Equation 2. Besides k_3 only responds to changes in variability of the data; this means that only scaling the input or output data (constant offset scenarios) does not impact on this parameter. In fact, this wash-off exponent is related to the kinetic energy of the rainfall, represented here using the modelled effective runoff rates, in mm/hr, from impervious areas. k_3 ranges from 0.25 in WQ+30%RF and WQRain4Flow4 to 0.42 in Rain3Flow3+20%TSS, which is in accordance to the 0.29 value found by McCarthy et al. (2011) for sediment transport. For the combined input and calibration data error scenarios, k_3 only ranged for the scenarios in which the volumes were overestimated (e.g. WQ+30%RF-20%TSS and WQ+30%RF+20%TSS).

While the change in the translation parameter *r* was not significant for most of the input, calibration and combined error scenarios, it was extreme for the specific cases in which the flow worst case scenario in the rainfall runoff model was included (input data error: WQRain3Flow4 and WQRain4Flow4; combined data error: WQRain3Flow4±20%TSS). These major shifts are due to the fact that Flow4 was the scenario that implied the worst *'Flow gauge maintenance scenario'*. The 6 months period for re-calibration of the sensor (both timing and level/velocity) significantly changed the properties and location of the modelled hydrograph. As a consequence *r* was affected. For example, *r* shifted from 3 timesteps (18 mins) in the 'NoError' scenario to more than 10 timesteps (1 hour) in WQRain3Flow4 and WQRain4Flow4.

The five parameters of the build-up/wash-off model were not significantly impacted when the flows modelled with sub-sets of parameters from KAREN PDs were used to generate TSS concentrations (Figure 3). This might be because the PD generated for the most sensitive parameter *EIF* in the 'NoError' scenario was very narrow (e.g. Figure 2), and thus different combinations of KAREN model parameters had very little effect on the model flows.

Figure 3 Histograms for the build-up/wash-off model parameters obtained with the input parameter error scenarios; i.e. Modelled flows with KAREN with sub-sets of parameters from the PDs - WQPar (x6).

In summary, the results indicated that while the sensitivity of the model to its parameters did not alter significantly (i.e. the shape of the parameter distributions remained the same) the model parameters were able to compensate for the errors in measured input and calibration data (i.e. parameter values or the position of the distributions changed).

Table 7 Build-up/wash-off - Histograms of parameter PDs for the different modelled flows, TSS concentrations and combined systematic error scenarios. The variation in the x axis is to facilitate the visualisation of certain scenarios.

3.2 Modelling uncertainty

Figure 4 and Figure 5 present the predictive uncertainty for an event with different error scenarios for KAREN and the build-up/wash-off model, respectively. Table 8 shows a summary of the observations (corresponding to each scenario) which fall within the uncertainty bounds for example scenarios (representing some of the different worst case scenarios). In general, the coverage from parameter uncertainties for KAREN was very low. This is a reflection of the shape and form of the parameter distributions shown in Figure 4; indeed, KAREN's most influential parameters have a very narrow distribution for most of the scenarios. This was thought to be caused by KAREN's limited ability to represent pervious surface flows and baseflow; without these processes, the model was unable to predict these lower flows (which dominate the dataset in terms of absolute number of data points) meaning that its coverage was very low (i.e. coverage equally weights all data points in the dataset). Furthermore the fact that the number of observations within the parameter uncertainty bands is the same or very similar for the 'NoError' scenario and the -30%Rain, Rain4 and, -30%R+30%F indicates that the effect of measured data uncertainty is small.

Figure 4 KAREN - predictive uncertainty for an event sample hydrograph for the calibration data period. The dark dots represent the observed data, the black line is the modelled data with the optimised parameter set, while the two shaded areas (of different grey) show the predictive uncertainty due to parameters and the total predictive uncertainty associated with the total error related to the modelling residuals.

Figure 5 Build-up/wash-off - predictive uncertainty for an event sample pollutograph. The dark dots represent the observed data, the black line is the modelled data with the optimised parameter set, while the two shaded areas (of different grey) show the predictive uncertainty due to parameters and the total predictive uncertainty associated with the total error related to the modelling residuals.

The total uncertainty associated with both models varied with the different scenarios. The percentage of observations within the total uncertainty bound for the build-up/wash-off model varied and was almost linearly correlated with the relative number of observations within the parameter uncertainty limits (**Error! Reference source not found.**). KAREN on the other hand, did not present such a clear pattern. It seems that the total uncertainty was intrinsically related to the scenario characteristics. For example, for Flow3 scenario, which is the worst case scenario underestimating flows, the coverage of the parameter uncertainty increased when compared to the 'NoError' scenario, while the total uncertainty decreased.

Table 8 Summary of the observations within the uncertainty bounds for sample scenarios (validation period results for the rainfall runoff model are given between brackets).

As seen in Table 8 and Figure 6, the coverage from parameter uncertainties was lower during the validation data period, which indicates that some parameter compensations obtained in the calibration process are not good as the simple model (i.e. not based on the detailed physics of the urban drainage systems) is not able to extrapolate the calibration results into the future.

Figure 6 KAREN - predictive uncertainty for an event sample hydrograph for the validation data period. The dark dots represent the observed data, the black line is the modelled data with the optimised parameter set, while the two shaded areas (of different grey) show the predictive uncertainty due to parameters and the total predictive uncertainty associated with the total error related to the modelling residuals.

4 Conclusions

The paper presented the application of a simple approach for global assessment of uncertainties in urban drainage models, which propagates errors in input and calibration data and evaluates how they impact the model calibration performance, sensitivity and predictive uncertainty. The approach was tested for a coupled urban stormwater model (a simple rainfall runoff model coupled with a commonly used build-up/wash-off model).

Results suggested that random errors in all input and calibration data had minor impact on the model performance and sensitivity. Systematic errors in input and calibration data on the other hand, influenced the model sensitivity (represented by the parameter distributions). In most of the scenarios (especially those where uncertainty in input and calibration data was represented using 'best-case' assumptions), the errors in measured data were fully compensated by parameter calibration. For example, when rainfall was systematically under or overestimated, the effective impervious area parameter varied systematically to compensate for the changes in the input data. In addition the model predictive uncertainty was also compensated in most of the cases as the number of observations within the parameter uncertainty bound kept fairly constant. It should then be noted that if the model parameters were considered initially as reflecting reality, this representation was reduced when input and calibration data errors were considered. Parameters were unable to compensate only in some of the scenarios where the uncertainty in the input and calibration data were represented using extreme worst-case scenarios. As such, in these few worst case scenarios, the model performance was reduced considerably. These cases were generally linked to scenarios in which mainly the time drifts in the battery logger device and calibration of water column levels were ignored for long periods. Results suggested that re-calibration once a month is sufficient.

The fact that uncertainties were assessed in such 'ill-posed' water quality models is a major weakness of this research as the obtained results are likely to be compromised. Nevertheless, the combination of evaluating 'ill-posed' models with such a large dataset allowed us to confirm that the model structure is the main reason for the poor performance of water quality models, and not the lack of measured data

The results obtained with the assessment of uncertainties in the build-up/wash off model are likely to be compromised because the model is 'ill-posed' (i.e. cross correlated parameters and model performance suggests the structure is not appropriate to represent stormwater sediment levels). While this is a limitation of this study, the combination of evaluating an 'ill-posed' model with a large dataset allowed us to confirm that the model structure is the main reason for the poor performance of water quality models, and not the lack of measured data. In this context, it seems that deterministic approaches currently used to model water quality could be re-considered and that the stochastic nature of the pollution generation process could be taken into account when modelling stormwater quality.

5 References

Bertrand-Krajewski, J. L., Bardin, J. P., Mourad, M. and Beranger, Y. (2003). Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. *Water Science and Technology* **47**(2): 95-102.

Beven, K. and Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* **6**(3): 279-298.

Beven, K. J., Smith, P. J. and Freer, J. E. (2008). So just why would a modeller choose to be incoherent? *Journal of Hydrology* **354**(1-4): 15-32.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*. **Series B 26**(2): 211-252.

Boyd, M. J., Bufill, M. C. and Knee, R. M. (1993). Pervious and impervious runoff in urban catchments. *Hydrological Sciences* **38**(6): 463-478.

Butts, M. B., Payne, J. T., Kristensen, M. and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* **298**(1-4): 242-266.

Chaubey, I., Haan, C. T., Grunwald, S. and Salisbury, J. M. (1999). Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology* **220**(48-61).

Chiew, F. H. S. and McMahon, T. A. (1999). Modelling runoff and diffuse pollution loads in urban areas. *Water Science and Technology* **39**(12): 214-248.

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. (2012). Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth, Parts A/B/C* **42–44**(0): 3-10.

Doherty, J. (2003). Model Independent Markov Chain Monte Carlo Analysis - MICA, Watermark Numerical Computing and US EPA, US.

Doherty, J. and Welter, D. (2010). A short exploration of structural noise. *Water Resources Research* **46**(W05525, doi:10.1029/2009WR008377).

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., Freni, G., Rauch, W. and Deletic, A. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research* **46**(8): 2545-2558.

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.-E., Quirmbach, M., Vaes, G. and Vieux, B. (2004). Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology* **299**(3-4): 186-202.

Feyen, L., Vrugt, J. A., Nualláin, B. Ó., van der Knijff, J. and De Roo, A. (2007). Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. *Journal of Hydrology* **332**(3-4): 276-289.

Francey, M., Fletcher, T. D., Deletic, A. and Duncan, H. (2010). New insights into water quality of urban stormwater in South Eastern Australia. *Journal of Environmental Engineering* **136**(4): 381-390.

HACH, C. (2008). Sigma 950 Flow Meter - Instrument Manual. 2009, from http://www.hach.com/fmmimghach?/CODE%3A3314-2008-1117081%7C1.

Harmel, R., Cooper, R., Slade, R., Haney, R. and Arnold, J. (2006). Cumulative uncertainty in measured stream flow and water quality data for small watersheds. *American Society of Agricultural and Biological Engineers* **49**(3): 689-701.

Haydon, S. and Deletic, A. (2009). Model output uncertainty of a coupled pathogen indicator-hydrologic catchment model due to input data uncertainty. *Environmental Modelling and Software* **24**(3): 322-328.

Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B. and Chebbo, G. (2003). Bayesian approach for the calibration of models: application to an urban stormwater pollution model. *Water Science and Technology* **47**(4): 77-84.

Kleidorfer, M., Chen, S., Dotto, C. B. S. and McCarthy, D. (2012). *Impact of objective function choice on model parameter sensitivity*. 9th International Conference on Urban Drainage Modelling, Belgrade, Serbia.

Kleidorfer, M., Deletic, A., Fletcher, T. D. and Rauch, W. (2009). Impact of input data uncertainties on stormwater model parameters. *Water Science and Technology* **60**(6): 1545 -1554.

Kotlash, A. R. and Chessman, B. C. (1998). Effects of water sample preservation and storage on nitrogen and phosphorus determinations: implications for the use of automated sampling equipment. *Water Research* **32**(12): 3731-3737.

Krzysztofowicz, R. and Kelly, K. (2000). Hydrologic Uncertainty Processor For Probabilistic River Stage Forecasting. *Water Resources Research* **36**(11): 3265-3277.

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology* **331**(1-2): 161-177.

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* **211**(1-4): 69-85.

Leecaster, M. K., Schiff, K. and Tiefenthaler, L. L. (2002). Assessment of efficient sampling designs for urban stormwater monitoring. *Water Research* **36**(6): 1556-1564.

Matott, L. S., Babendreier, J. E. and Purucker, S. T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research* **45**(6): W06421.

McCarthy, D. T. (2008). Modelling microorganisms in urban stormwater. *Civil Engineering*. Melbourne, Monash University, Australia. **PhD thesis:** 488.

McCarthy, D. T., Deletic, A., Mitchell, V. G. and Diaper, C. (2011). Development and testing of a model for Micro-Organism Prediction in Urban Stormwater (MOPUS). *Journal of Hydrology* **409**(1-2): 236-247.

McCarthy, D. T., Deletic, A., Mitchell, V. G., Fletcher, T. D. and Diaper, C. (2008). Uncertainties in stormwater E. coli levels. *Water Research* **42**(6-7): 1812-1824.

Métadier, M. and Bertrand-Krajewski, J. L. (2012). *Interest of Bayesian learning principle for stormwater quality modelling based on turbidity time series*. 9th International Conference on Urban Drainage Modelling, Belgrade, Serbia.

Molini, A., Lanza, L. G. and Barbera, P. L. (2005a). The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications. *Hydrological Processes* **19**: 1073-1088.

Molini, A., Lanza, L. G. and La Barbera, P. (2005b). Improving the accuracy of tipping-bucket rain records using disaggregation techniques. *Atmospheric Research* **77**(1-4): 203-217.

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

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

Rauch, W. and Kinzel, H. (2007). KAREN - User Manual, hydro-IT GmbH, Innsbruck, (in German).

Rauch, W., Thurner, N. and Harremoes, P. (1998). Required accuracy of rainfall data for integrated urban drainage modeling. *Water Science and Technology* **37**(11): 81-89.

Renard, B., Kuczera, G., Kavetski, D., Thyer, M. and Franks, S. (2008). *Bayesian Total Error Analysis for Hydrologic Models: Quantifying Uncertainties Arising from Input, Output and Structural Errors*. Water Down Under 2008 Modbury, SA: Engineers Australia ; Causal Productions

Rode, M. and Suhr, U. (2007). Uncertainties in selected river quality data. *Hydrology and Earth System Sciences* **11**: 863-874.

Sartor, J. D. and Boyd, G. B. (1972). Water pollution aspects of street surface contaminants. *Rep. EPA-R2-72-081*. Washington, D. C., U.S. Environ. Prot. Agency.

Sevruk, B. (1982). Methods of correction for systematic error in point precipitation measurement for operational use,. *WMO Operational Hydrology Report 21* **21**(589): 91.

Sevruk, B. (2002). WMO questionnaire on recording precipitation gauges: state-of-the-art. *Water Science and Technology* **45**(2): 139-145

Sherman, L. K. (1932). Stream-flow from Rainfall by the Unit-Graph Method. *Engineering News-Record* **108**: 501-505.

Simic, M. and Maksimovic, C. (1994). Effect of the siphon control on the dynamic characteristics of a tipping bucket raingauge. *Hydrological Sciences* **39**(1): 35-46.

Taylor, B. N. and Kuyatt, C. E. (1994). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. USA, NIST Technical Note 1297, National Institute of Standards and Technology.

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

USEPA (2007). Storm Water Management Model (SWMM) - User Manual version 5.0. US Environmental Protection Agency (USEPA). *US Environmental Protection Agency - EPA/600/R-05/039*. Washington DC.

Vaze, J. and Chiew, F. H. S. (2003). Comparative evaluation of urban storm water quality models. *Water Resources Research* **39**(10): 1280.

Vrugt, J. A., Braak, C. J. F. t., Diks, C. G. H., Robinson, B. A., Hyman, J. M. and Higdon, D. (2009). 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): 271-288.

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J. and Yang, H. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology* **358**(1-2): 1-23.

Younger, P. M., Freer, J. E. and Beven, K. J. (2009). Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. *Hydrological Processes* **23**(14): 1988-2003.

Table 1 The governing equations of the build-up/wash-off model.

Process and model equation	Unit	Equation no.
Build-up during dry weather:		
$M(t_d) = M_0 \cdot \left(1 - e^{-\frac{6k_1}{1440}t_d}\right)$	(kg)	1
Wash-off during wet weather:		
$C(t) = k_2 \cdot M(t_d) \cdot \frac{q}{RC}(t + r)^{k_3}$	(mg/L)	2
$W(t) = 10^{-6} * C(t) * Vol(t)$	(kg)	3

$M(t_d)$ is the mass of solids which accumulate on the surface during dry weather periods (t_d) within a timestep in kg; M_0 is the maximum amount of accumulated solids in kg, k_1 represents the accumulation constant in day^{-1}; C is the concentration of pollutants in the runoff at the time t in mg/L; k_2 and k_3 are the wash-off coefficient and exponent, respectively; q is the runoff in mm/h, RC is the catchment runoff coefficient; and, r in number of timesteps in min is used to correct for the fact that pollutograph precedes the runoff hydrograph (Chiew and McMahon, 1999); $W(t)$ is the wash-off of TSS from the surface ; and, Vol is the runoff volume in L.

Table 2 Summary of the tested error scenarios for the KAREN rainfall runoff model.

Base scenario				**Name**
No error				**NoError**
Input data error scenarios - Rainfall error scenarios				
*Random error only**	Random number sampled from U[-0.5, 0.5] for each rain timestep			**RREr (x10)**
*Random spatial Distribution**	Random offset sampled from U[-0.3, 0.3] for each rain event			**RainSD (x10)**
Systematic constant offset	-30% offset to entire rainfall dataset			**-30%Rain**
	+30% offset to entire rainfall dataset			**+30%Rain**
Rain gauge maintenance and bucket volume error scenarios				
t_{reset}	t_{drfit}	*Bucket volume error*		
1month	-0.14min/day	-30%		**Rain1**
1month	+0.14min/day	+30%		**Rain2**
6 months	-0.14min/day	-30%		**Rain3**
6 months	+0.14min/day	+30%		**Rain4**

Calibration data error - Flow error scenarios				
*Random error only**	Law of Propagation of Uncertainty			**FREr (x10)**
Systematic constant offset	-30% offset to entire flow dataset			**-30%Flow**
	+30% offset to entire flow dataset			**+30%Flow**
Flow gauge maintenance scenarios				
t_{reset}	Height drift	*Re-calibration shift*	*Velocity error*	
1 month	-2mm/mth	U[-5,5]	-10%	**Flow1**
1 month	2mm/mth	U[-5,5]	10%	**Flow2**
6 months	-10mm/mth	U[-20,20]	-30%	**Flow3**
6 months	10mm/mth	U[-20,20]	30%	**Flow4**

Combination of input (rainfall) and calibration (flow) error scenarios		
Systematic constant offsets	-30% offset to entire rainfall & flow datasets	**-30%R&F**
	+30% offset to entire rainfall & flow datasets	**+30%R&F**
	-30% rainfall & +30% flow	**-30%R+30%F**
	+30% rainfall & -30% flow	**+30%R-30%F**
Rain and flow gauges maintenance scenarios	Rain3 & Flow3	**Rain3Flow3**
	Rain3 & Flow4	**Rain3Flow4**
	Rain4 & Flow3	**Rain4Flow3**
	Rain4 & Flow4	**Rain4Flow4**

*10 random sets were generated and 10 MICA realisations were performed

Table 3 Summary of the tested error scenarios for the water quality model (see Table 2 for further explanations for some model errors).

Scenario	Description	Name
Base scenario		**Name**
No error		**WQNoError**
Input parameter error scenarios		
Modelled flows with KAREN with sub-sets of parameters from the PDs		**WQPar (x6)**
Input data error scenarios - Modelled flows with KAREN		
*Random error only**		**WQMFREr (x10)**
Systematic constant offsets		**WQ-30%R&F**
		WQ+30%R&F
		WQ-30%R+30%F
		WQ+30%R-30%F
Modelled flows with rain and flow gauges maintenance scenarios		**WQRain3Flow3**
		WQRain3Flow4
		WQRain4Flow3
		WQRain4Flow4
Calibration data error - TSS Concentrations		
*Random error only**	Random number sampled from U[-0.28,0.28] for each discrete sample	**TSSREr (x10)**
Systematic constant offset	-20% offset to entire concentration dataset	**-20%TSS**
	+20% offset to entire concentration dataset	**+20%TSS**
Discrete samples – combining all the systematic sources	-4.9% offset to entire concentration dataset	**TSS1**
	+25% offset to entire concentration dataset	**TSS2**
	-9.8% offset to entire concentration dataset	**TSS3**
	+40% offset to entire concentration dataset	**TSS4**
Combination of input (modelled flow) and calibration (TSS) scenarios		
Systematic constant offsets		**WQ-30%R&F-20%TSS**
		WQ-30%R&F+20%TSS
		WQ+30%R&F-20%TSS
		WQ+30%R&F+20%TSS
Modelled flows with rain and flow gauges maintenance scenarios combined with systematic constant offset		**Rain3Flow3-20%TSS**
		Rain3Flow3+20%TSS
		Rain3Flow4-20%TSS
		Rain3Flow4+20%TSS

*10 random sets were generated and 10 MICA realisations were performed

Table 4 Overall efficiency for the different scenarios of rainfall runoff model KAREN - E_{WLS} stands for the maximum efficiency corresponding to the minimum weighted least-squares; E_{accep} stands for the maximum efficiency from all the MICA accepted parameter sets.

Rain Error Scenarios		
		E_{WLS} [E_{accep}]
NoError		*0.58[0.76]*
Random error		
	RREr	from 0.55 to 0.58 [0.7]
Systematic error		
Best case scen.	Rain1	0.55[0.7]
	Rain2	0.55[0.67]
Worst case scen	Rain3	0.57[0.67]
	Rain4	**0.40[0.54]**
	-30%Rain	0.57[0.65]
	+30%Rain	0.56 [0.65]

Flow Error Scenarios		
		E_{WLS} [E_{accep}]
NoError		*0.58[0.76]*
Random error		
	FREr	from 0.55 to 0.58 [0.7]
Systematic error		
Best case scen.	Flow1	0.50[0.65]
	Flow2	0.58[0.75]
Worst case scen	Flow3	**0.38[0.54]**
	Flow4	**0.37[0.57]**
	-30%Flow	0.58[0.67]
	+30%Flow	0.58[0.62]

Combination Scenarios	
	E_{WLS} [E_{accep}]
NoError	*0.58[0.76]*
Rain3Flow3	0.36[0.47]
Rain3Flow4	0.39[0.52]
Rain4Flow3	**0.17[0.33]**
Rain4Flow4	**0.16[0.41]**
-30%R&F	0.58[0.59]
+30%R&F	0.58[0.65]
-30%R+30%F	0.58[0.60]
+30%R-30%F	0.58[0.59]

Table 5 Overall efficiency for the different scenarios of the build-up/wash-off stormwater quality model - E_{WLS} stands for the maximum efficiency corresponding to the minimum weighted least-square; E_{accep} stands for the maximum efficiency from all the MICA accepted parameter sets; RC stand for the runoff coefficient.

Input Parameter Error Scenarios	**RC**	**E_{WLS} [E_{accep}]**
WQNoError	0.31	*0.16[0.19]*
WQPar1	0.31	0.18[0.20]
WQPar2	0.29	0.17[0.20]
WQPar3	0.32	0.16[0.19]
WQPar4	0.34	0.16[0.19]
WQPar5	0.31	0.16[0.19]
WQPar6	0.31	0.16[0.20]

Modelled Flows with Rain&Flow Error Scenarios	**RC**	**E_{WLS} [E_{accep}]**
WQNoError	0.31	*0.16[0.19]*
Random error		
WQMFREr	0.31	0.14 from to 0.17 [0.19]
Systematic error		
WQRain3Flow3	0.29	0.16[0.16]
WQRain4Flow4	0.92	**0.12**[0.13]
WQRain3Flow4	0.99	0.14[0.15]
WQRain4Flow3	0.16	0.21[0.17]
WQ-30%RF	0.34	0.16[0.18]
WQ+30%RF	0.32	0.17[0.19]
WQ-30%R+30%F	0.58	0.16[0.18]
WQ+30%R-30%F	0.17	0.16[0.16]

TSS Concentration Scenarios		**RC**	**E_{WLS} [E_{accep}]**
WQNoError		0.31	*0.16[0.19]*
Random error			
	TSSREr	0.31	from **0.11** to 0.17 [0.20]
Systematic error			
Best case scen.	TSS1	0.31	0.17[0.18]
	TSS2		0.15[0.17]
Worst case scen.	TSS3	0.31	0.17[0.19]
	TSS4		**0.12**[0.15]
	-20%TSS	0.31	0.16[0.19]
	20%TSS		0.14[0.16]

Combination Scenarios	**RC**	**E_{WLS} [E_{accep}]**
WQNoError	0.31	*0.16[0.19]*
Combination Scenarios		
Rain3Flow3-20%TSS	0.29	0.16[0.18]
Rain3Flow3+20%TSS		0.14[0.16]
Rain3Flow4-20%TSS	0.92	0.14[0.16]
Rain3Flow4+20%TSS		0.14[0.16]
WQ-30%RF-20%TSS	0.34	0.18[0.19]
WQ-30%RF+20%TSS		0.14[0.17]
WQ+30%RF-20%TSS	0.32	0.19[0.21]
WQ+30%RF+20%TSS		**0.13**[0.15]

Table 6 Rainfall runoff model KAREN - Histograms of parameter PDs for the different error scenarios. The variation in the x axis is to facilitate the visualisation of specific scenarios.

Scenarios —Noerror	***EIF*** **(%)**	***TOC*** **(min)**	***li*** **(mm)**	***ev*** **(mm/day)**

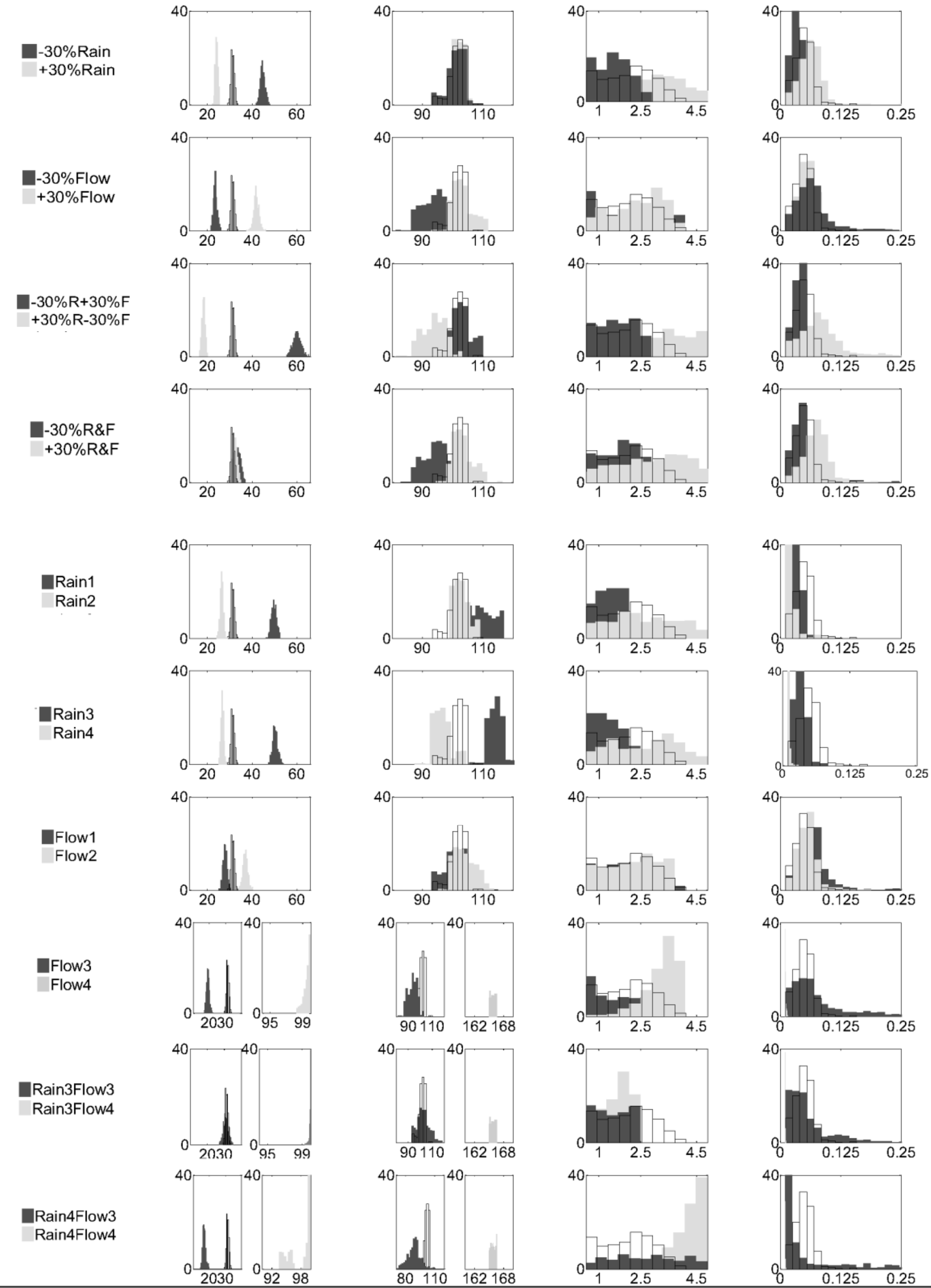
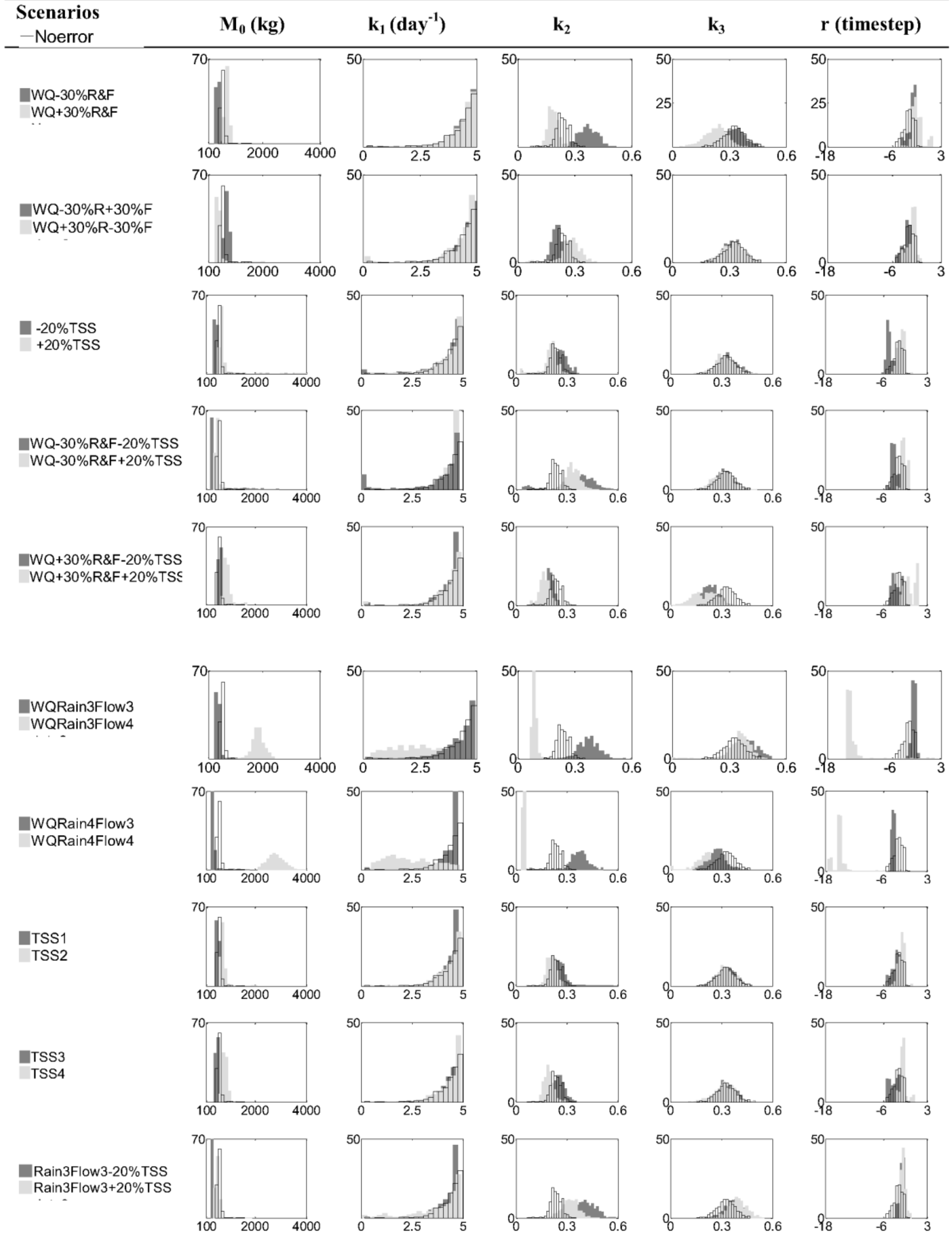


Table 7 Build-up/wash-off - Histograms of parameter PDs for the different modelled flows, TSS concentrations and combined systematic error scenarios. The variation in the x axis is to facilitate the visualisation of certain scenarios.



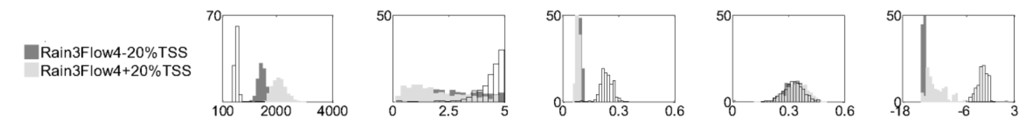


Table 8 Summary of the observations within the uncertainty bounds for sample scenarios (validation period results for the rainfall runoff model are given between brackets).

Rainfall Runoff KAREN

	NoError	-30%Rain	Rain4	Flow3	-30%R+30%F
Observations within the *paramete*r uncertainty bound (%)	9 [8]	9 [1]	6 [2]	20 [2]	9 [2]
Observations within the *total* uncertainty bound (%)	63 [30]	87 [98]	91 [30]	59 [30]	84 [100]

Build-up/Wash-off

	WQNoError	WQRain3Flow3	TSS4	WQ-30%RF+20%TSS
Observations within the *paramete*r uncertainty bound (%)	48	15	53	19
Observations within the *total* uncertainty bound (%)	72	40	79	42

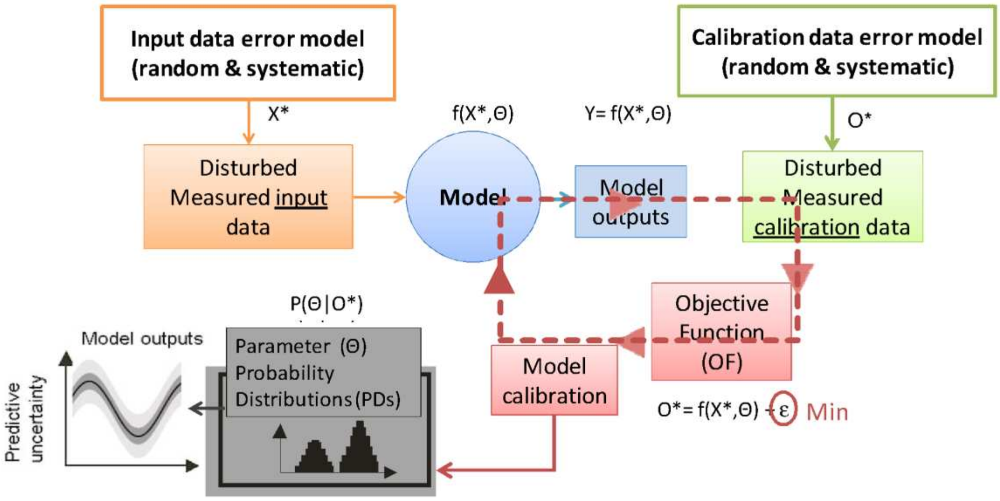


Figure 1 Propagation of input and calibration data errors through models (after Deletic et al, 2012).

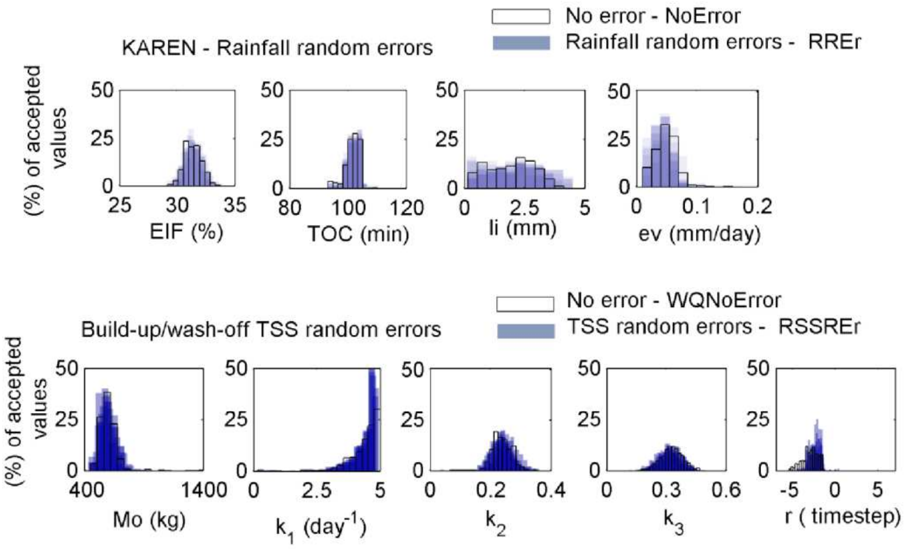


Figure 2 Histograms for KAREN parameters (top) obtained from the 'NoError' scenario and from the 10 sets of data generated with KAREN rainfall random errors – RREr (x10); and histograms for the build-up/wash-off model parameters (bottom) obtained from the 'WQNoError' scenario and from the 10 sets of data generated with TSS random errors – TSSREr (x10) (see Table 2 for abbreviations).

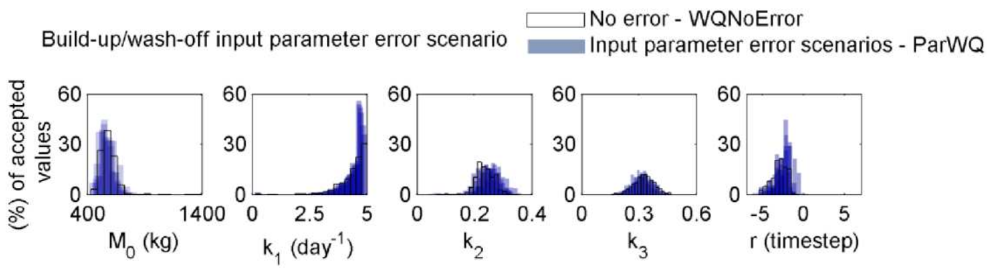


Figure 3 Histograms for the build-up/wash-off model parameters obtained with the input parameter error scenarios; i.e. Modelled flows with KAREN with sub-sets of parameters from the PDs - WQPar (x6).

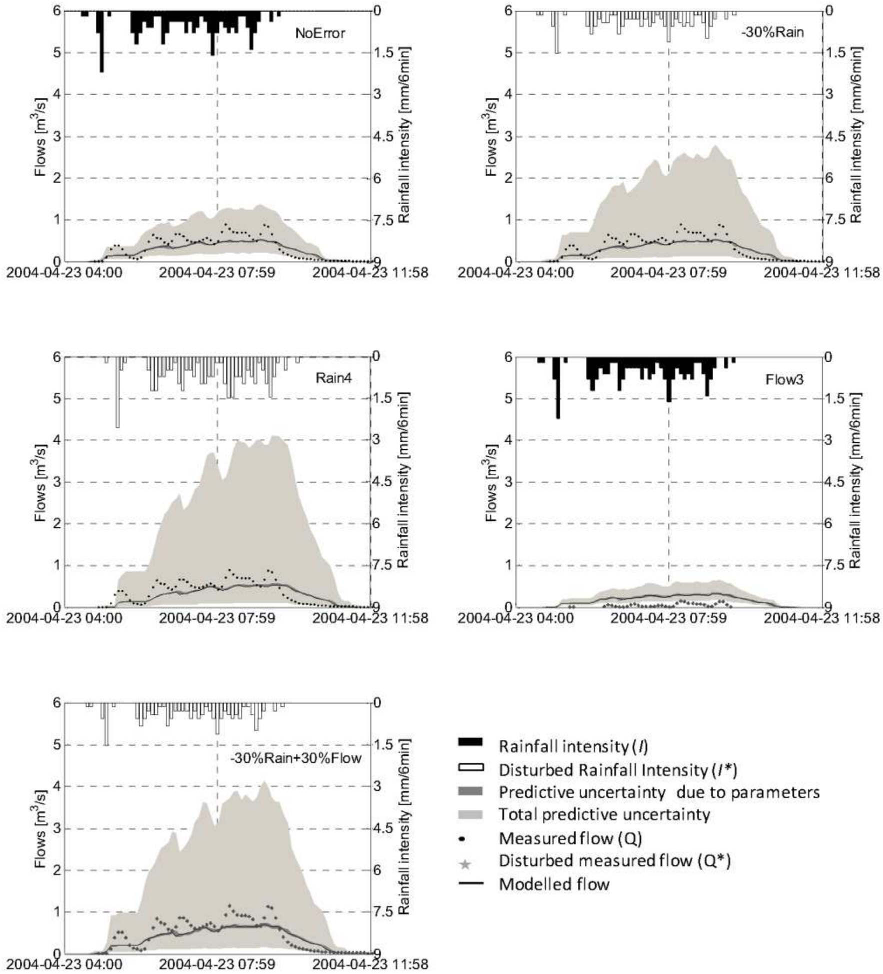


Figure 4 KAREN - predictive uncertainty for an event sample hydrograph. The dark dots represent the observed data, the black line is the modelled data with the optimised parameter set, while the two shaded areas (of different grey) show the predictive uncertainty due to parameters and the total predictive uncertainty associated with the total error related to the modelling residuals.

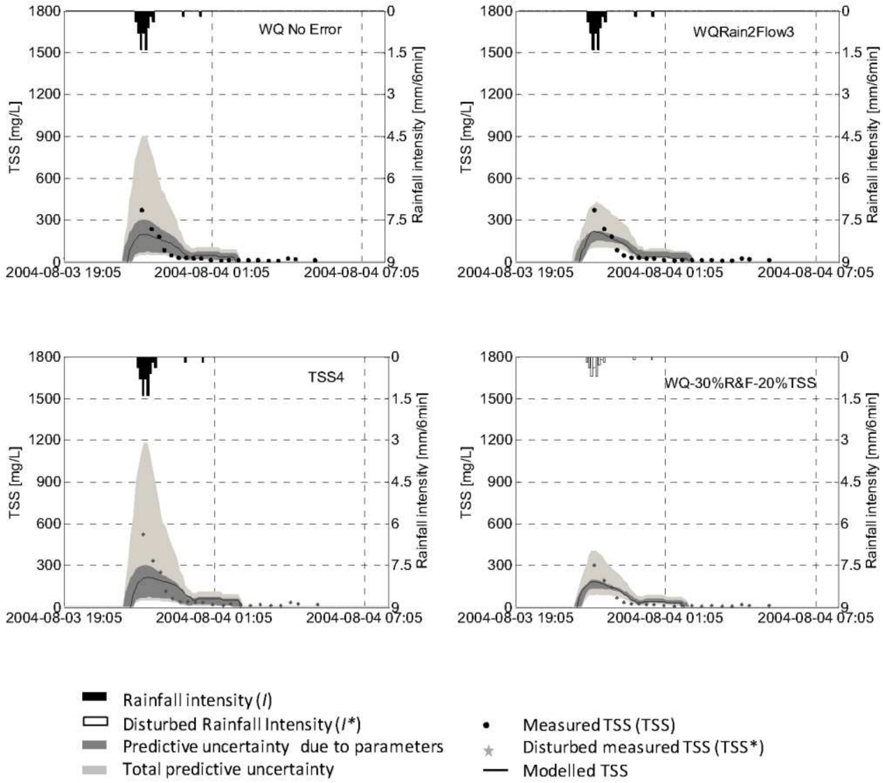


Figure 5 Build-up/wash-off - predictive uncertainty for an event sample pollutograph. The dark dots represent the observed data, the black line is the modelled data with the optimised parameter set, while the two shaded areas (of different grey) show the predictive uncertainty due to parameters and the total predictive uncertainty associated with the total error related to the modelling residuals.

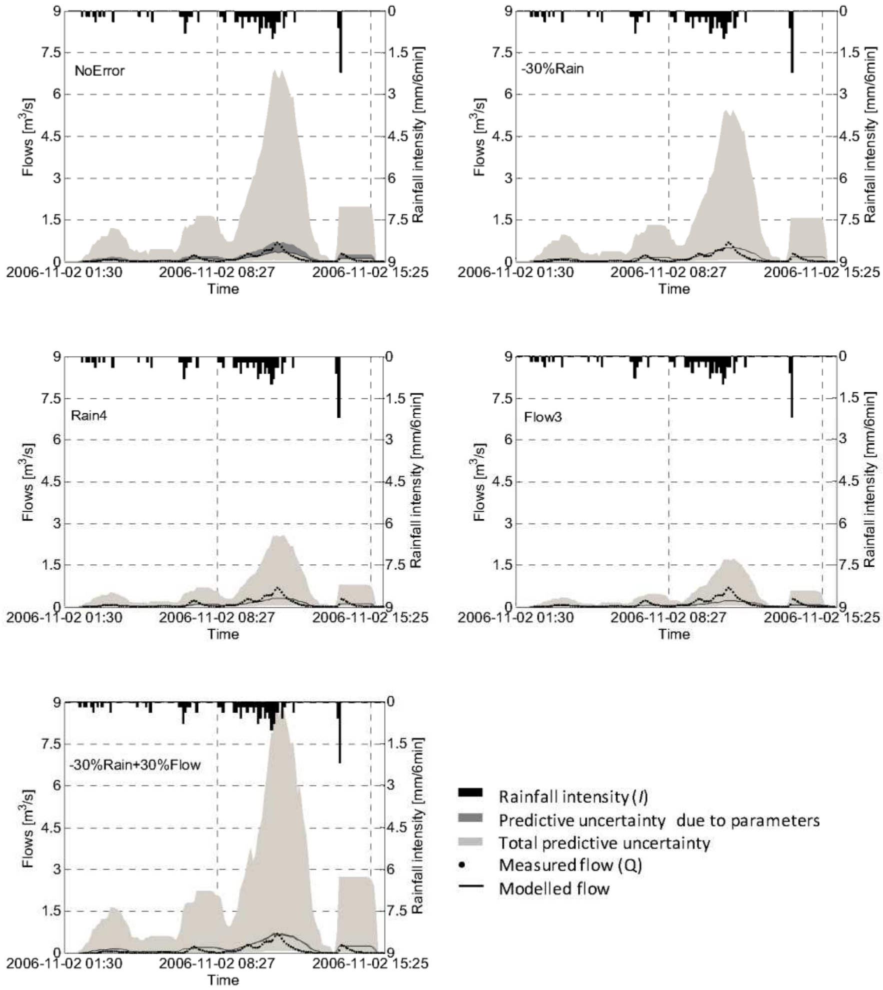


Figure 6 KAREN - predictive uncertainty for an event sample hydrograph for the validation data period. The dark dots represent the observed data, the black line is the modelled data with the optimised parameter set, while the two shaded areas (of different grey) show the predictive uncertainty due to parameters and the total predictive uncertainty associated with the total error related to the modelling residuals.

7.7 Conclusions

This chapter started with a summary of the main sources of uncertainties in the measured variables that are commonly needed for the application of urban drainage models. Specifically, the uncertainties associated with as rainfall, flow rates and TSS concentrations were reviewed. This was followed by an investigation of the impact of these input and calibration measured data uncertainty (here, estimated through the error models) on the model sensitivity and predictive uncertainty.

Results suggested that random errors in all input and calibration data had a minor impact on the model performance and sensitivity. Systematic errors in input and calibration data influenced the model sensitivity (represented by the parameter distributions). In most of the scenarios (especially those where uncertainty in input and calibration data was represented using 'best-case' assumptions), the model performance was fully compensated by the parameters. For example, when rainfall was systematically under or overestimated, the effective impervious area parameter varied systematically to compensate for the changes in the input data. In addition the model predictive uncertainty was also compensated in most of the cases as the number of observations within the parameter uncertainty bound was kept fairly constant. It should then be noted that if the model parameters were considered initially as reflecting real characteristics of the catchment (i.e. not only mere calibration parameters values), this representation was reduced when input and calibration data errors were considered. Parameters were unable to compensate only in some of the scenarios where the uncertainty in the input and calibration data were represented using extreme worst-case scenarios. As such, in these few worst case scenarios, the model performance was reduced considerably. These cases were generally linked to scenarios in which the time drifts in the battery logger device was ignored for long periods, which indicates that rain and flow gauges should be regularly recalibrated. From the results presented, it is suggested that re-calibration once a month is sufficient.

The results obtained with the assessment of uncertainties in the build-up/wash-off model are likely to be compromised because of its 'ill-posed' nature. While this is a limitation of this study, the assessment of the uncertainties in such a widely used model (with a large number of events) confirmed that the pollution generation processes in the catchment are quite variable. This suggests that the determinist approaches currently used to model water quality should be re-considered and that the stochastic nature of the pollution generation processes should be taken into account when modelling stormwater quality.

7.8 References

Achleitner, S., Moderl, M. and Rauch, W. (2007). CITY DRAIN © – An open source approach for simulation of integrated urban drainage systems. *Environmental Modelling and Software* **22**(8): 1184-1195.

Ahyerre, M., Chebbo, G., Tassin, B. and Gaume, E. (1998). Storm water quality modelling, an ambitious objective? *Water Science and Technology* **37**(1): 205-213.

Beck, M. B. (1987). Water Quality Modeling: A Review of the Analysis of Uncertainty. *Water Resources Research* **23**(8): 1393-1442.

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

Bertrand-Krajewski, J. L., Bardin, J. P., Mourad, M. and Beranger, Y. (2003). Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. *Water Science and Technology* **47**(2): 95-102.

Bertrand-Krajewski, J. L. and Muste, M. (2007). Chapter 6: Understanding and managing uncertainty. *in: Flectcher, T.D. and Deletić, A. (Eds.) Data Requirements for Integrated Urban Water Management.* 18 London, UK, Taylor and Francis, Urban Water series - UNESCO IHP**:** 65-90.

eWater CRC (2012). Model for urban stormwater improvement conceptualisation (MUSIC) (Version 5.0), Canberra: eWater Cooperative Research Centre.

Fletcher, T. (2008). Personal communication regarding 'Estimation of systematic errors in depth measurements when dopplers are used for monitoring of flows in stormwater pipes'.

Gordon, J. D., Newland, C. A. and Gagliardi, S. T. (2000). Laboratory Performance in the Sediment Laboratory Quality-Assurance Project, 1996–98. U.S. Geological Survey Water-Resources Investigations Report 99-4184. Denver, Colorado.**:** 39.

Harmel, R., Cooper, R., Slade, R., Haney, R. and Arnold, J. (2006). Cumulative uncertainty in measured stream flow and water quality data for small watersheds. *American Society of Agricultural and Biological Engineers* **49**(3): 689-701.

Harmel, R. D., Smith, D. R., King, K. W. and Slade, R. M. (2009). Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications. *Environmental Modelling and Software* **24**(7): 832-842.

Haydon, S. and Deletic, A. (2009). Model output uncertainty of a coupled pathogen indicator-hydrologic catchment model due to input data uncertainty. *Environmental Modelling and Software* **24**(3): 322-328.

Huang, J., Tu, Z., Du, P., Lin, J. and Li, Q. (2010). Uncertainties in stormwater runoff data collection from a small urban catchment, Southeast China. *Journal of Environmental Sciences* **22**(11): 1703-1709.

Kotlash, A. R. and Chessman, B. C. (1998). Effects of water sample preservation and storage on nitrogen and phosphorus determinations: implications for the use of automated sampling equipment. *Water Research* **32**(12): 3731-3737.

Maksimovic, C., Buzek, L. and Petrovic, J. (1991). Corrections of rainfall data obtained by tipping bucket rain gauge. *Atmospheric Research* **27**(1-3): 45-53.

Martin, G. R., Smoot, J. L. and White, K. D. (1992). A Comparison of Surface-Grab and Cross Sectionally Integrated Stream-Water-Quality Sampling Methods. *Water Environment Research* **64**(7): 866-876.

McCarthy, D., Bach, P. M. and Deletic, A. (2009). Conducting a microbial budget - a literature review, Monash University, Australia.

McCarthy, D. T. (2008). Modelling microorganisms in urban stormwater. *Civil Engineering*. Melbourne, Monash University, Australia. **PhD thesis:** 488.

McCarthy, D. T., Deletic, A., Mitchell, V. G., Fletcher, T. D. and Diaper, C. (2008). Uncertainties in stormwater E. coli levels. *Water Research* **42**(6-7): 1812-1824.

Molini, A., Lanza, L. G. and Barbera, P. L. (2005a). The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications. *Hydrological Processes* **19**: 1073-1088.

Molini, A., Lanza, L. G. and La Barbera, P. (2005b). Improving the accuracy of tipping-bucket rain records using disaggregation techniques. *Atmospheric Research* **77**(1-4): 203-217.

Niemczynowicz, J. (1986). The Dynamic Calibration of Tipping-Bucket Raingauges. *Nordic Hydrology* **17**: 203-214.

Pavlyukov, Y. (2007). Precipitation measurement with automated tipping-bucket rain gauges. *Russian Meteorology and Hydrology* **32**(11): 711-718.

Prodanovic, D., R. (2009). Personal communication on the 'Estimation of systematic errors in depth and velocities measurements when dopplers are used for monitoring of flows in stormwater pipes'.

Rode, M. and Suhr, U. (2007). Uncertainties in selected river quality data. *Hydrology and Earth System Sciences* **11**: 863-874.

Sevruk, B. (1982). Methods of correction for systematic error in point precipitation measurement for operational use,. *WMO Operational Hydrology Report 21* **21**(589): 91.

Sevruk, B. (2002). WMO questionnaire on recording precipitation gauges: state-of-the-art. *Water Science and Technology* **45**(2): 139-145

Simic, M. and Maksimovic, C. (1994). Effect of the siphon control on the dynamic characteristics of a tipping bucket raingauge. *Hydrological Sciences* **39**(1): 35-46.

Stransky, D., Bares, V. and Fatlm, P. (2006). *The Effect of Rainfall Measurement Uncertainties on Rainfall-Runoff Processes Modelling.* 7th International Conference on Urban Drainage Modelling and the 4th International Conference on Water Sensitive Urban Design, Melbourne, Australia.

Taylor, B. N. and Kuyatt, C. E. (1994). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. USA, NIST Technical Note 1297, National Institute of Standards and Technology.

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

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M. and Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.* **44**: W00B09.

Wang, J., Fisher, B. L. and Wolff, D. B. (2008). Estimating Rain Rates from Tipping-Bucket Rain Gauge Measurements. *Journal of Atmospheric and Oceanic Technology* **25**(1): 43-56.

WP Software (1995). *XP-SWMM stormwater management model with XP graphical interface, User's manual, Version 2*. Tampa, Forida and Belconnen, USA.

Chapter 8

Discussion, conclusions and further investigation

8.1 Introduction

The objective of this research project was to advance the current understanding of the uncertainties in urban rainfall runoff and pollution generation models in order to better define their reliability. The study concentrated on the main following areas: (i) identifying suitable global uncertainty analysis method(s) to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models; (ii) exploring parameter calibration, model sensitivity and the resulting predictive uncertainties in stormwater models with different level of complexities; and, (iii) investigating the impact of measured input and calibration data uncertainty on the performance, sensitivity and predictive uncertainty of stormwater models. This chapter begins with an assessment of the strengths and weaknesses of the presented research (Section 8.2). The key conclusions gained from the study (i.e. the major findings from each of the chapters in the thesis) are then presented in Section 8.3. Finally, opportunities for further work are discussed in Section 8.4.

8.2 Strengths and weaknesses of the evidence

Methods for global sensitivity and uncertainty analysis

The comparison of methods for global sensitivity and uncertainty analysis contribute to better understand the limitations and advantages of each method. A strength of this research is that the tested methods were based on the current available knowledge on uncertainty of stormwater models, as well as on the state of the art approaches used for uncertainty analysis in related fields (including both 'formal' and 'informal' Bayesian approaches). The work was done in close collaboration with another four research groups who are all recognised as international leaders in the urban drainage research.

Method assumptions – checking and validating

The likelihood function in the adopted formal Bayesian approach assumes that the model errors (or residuals between the measured and modelled values) are independent, homoscedastic and normally distributed. Some of the main weaknesses of this thesis are that such assumptions were not verified at the start of the activities and that the normality assumption was only verified in the later stages. Nevertheless, the fact that we studied the impact of the normality assumption on the results (Chapter 6), and even attempted to develop a more rigorous way of applying the methodology (Chapter 7), is a strength of this research (as this is often not addressed in the similar studies).

Another weakness of the current project is that the other assumptions regarding the error model were not verified: model residuals were assumed to be independent (not correlated) and homoscedastic. This is a major limitation as the model residuals are very likely to be correlated. However, the serial correlation between the data points was not considered in this study because the current used methods to account for autocorrelation (e.g. first order model) were proven to be ineffective for small timesteps as used in this research. The homoscedasticity assumption of the model errors does not seem much of a problem as the data transformation resulted in normal distributions (with constant mean and standard deviation).

Parameter calibration, model sensitivity and the resulting predictive uncertainties

This thesis was the first in the field to compare the sensitivity and predictive uncertainty of conceptual stormwater models with different levels of complexity. It also provided a comprehensive analysis of model parameters and their interactions. Such analysis offered practical recommendations for developing new stormwater models and for refining of existing ones.

Another strength of this research was the ability to assess uncertainties in stormwater modelling on five urban catchments (with different areas, land-uses and levels of development) using a long term and high resolution dataset (mainly for water quality data, which used to be scarce so far). Testing the models using this large dataset meant that the models had to demonstrate their ability (or inability) of representing the variability in the data with the possible contexts and ranges of events and values (e.g. if the model is able to explain the hydrological responses from a specific urban catchment during the different seasons).

A weakness was that the number of models tested was limited (although it is also acknowledged it is hard to cover all available models in the one thesis). Further, although the tested pollution generation models were chosen for being the most adopted in urban drainage models, the results demonstrated that they were not able to represent reality. The fact that uncertainties were assessed in such 'ill-posed' water quality models is a major weakness of this research as the obtained results are likely to be compromised. Nevertheless, the combination of evaluating 'ill-posed' models with such a large dataset allowed us to confirm that the model structure is the main reason for the poor performance of water quality models, and not the lack of measured data (as was the case previously in the urban drainage modelling community).

Impact of measured input and calibration data uncertainty

This thesis tested a new framework for joint assessment of the impacts of key sources of uncertainties on the modelled results. Although the application of the framework to assess the input and calibration data errors demands significant computational time, the applied approach is relatively simple and far less complex than other similar frameworks (that are in fact very rare), and could be adopted and/or adapted for assessing the errors in other conceptual based models with a low number of calibration parameters. However, the procedure is still too computationally demanding to allow its application for more deterministic models.

One main limitation of this framework is that the sources of errors are lumped, and therefore the changes in one specific source cannot be accounted independently. Again, assessment of uncertainties in the 'ill-posed' water quality model is a weakness of this research task. However, the strength of the study is that it confirmed the suggestion that the pollution generation processes in the catchment are stochastic, and that the physically based conceptual approaches (currently used to model water quality) should be re-considered.

8.3 Conclusions

Global sensitivity analysis methods in stormwater modelling

On the search for suitable global uncertainty analysis methods to perform parameter calibration, model sensitivity and uncertainty analysis in stormwater models, the main contribution of this thesis was that the application of different methods produced fairly similar results. It should be noted that these results might be severely compromised as the tested uncertainty analysis methods rely on a number of assumptions and subjectiveness. As such, the appropriateness of these methods is questionable.

Results from this research task recommended that the selection of the most appropriate method for uncertainty estimation is a trade-off between the need for a strong theory-based description of uncertainty (but limited by the requirements on prior knowledge about the structure of the model errors - Bayesian approach, MICA), simplicity (but limited by the subjectivity - GLUE) and computational efficiency (also affected by subjectivity - AMALGAM and SCEM-UA). It is also suggested that different evaluation scenarios should be analysed (i.e. different catchments, models, data, etc). Modellers should also select the method which is most suitable for the system they are modelling (e.g. complexity of the model's structure including the number of parameters), their skill/knowledge level, the available information, and the purpose of their study.

The Bayesian approach was suggested to be more suitable for uncertainties in stormwater modelling because it is more efficient (or at least less time consuming) to evaluate models with larger number of parameters and also because of its statistical rigour. Further application of the tested Bayesian approach confirmed the potential of the method to assess different urban drainage models (i.e. with different level of complexities) in urban catchments of different sizes and land-use types.

However, the likelihood functions in the applied Bayesian approach assumes that the model errors (residuals) are normally distributed. This study demonstrated that this assumption is often not met in stormwater modelling (i.e. model residual are not normally distributed). In order to verify the normality assumption, a data transformation approach was adopted. While Box-Cox transformation solved the normality issue, it drastically influenced the sensitivity of the model parameters. Consequently we decided to assess the impacts of verifying the normality assumption of the model errors on the model parameter sensitivity and its associated predictive uncertainty.

In addition, all observed data have uncertainty, and this should be taken into account in the likelihood function so that the parameters are estimated appropriately; indeed, it is important that the function places more emphasis on data which has lower uncertainty. Weighting strategies were used to re-adjust how the likelihood function emphasises various parts of the dataset to (1) consider measured data uncertainty and (2) compensate for the Box-Cox transformation which had adjusted the emphasis in an undesirable way.

It was found that the overall efficiency of the models was different and that the changes in parameter distributions were significant between the scenarios in which the normality assumption of the residuals was verified or not. The main reason for such results is the fact that the data transformation used to meet the normality assumption altered the data, which then influenced the emphasis on various parts of the hydrograph.

Another interesting finding from the thesis was that pursuing the normality assumption by adequate data transformation seemed to better calibrate the models. It was found that when the normality assumption was achieved, most of the model's processes were activated (resulting in more sensitive parameters), while only few parameters drove the outputs when the normality assumption was not achieved (i.e. only some parameters were sensitive).

Parameter calibration, model sensitivity and the resulting predictive uncertainties

The comparison between two catchment rainfall runoff models with different levels of complexity demonstrated that both models performed similarly and that the effective impervious fraction is the most important parameter in runoff prediction. Other key parameters are those related to the time of concentration. It is interesting to note that the calibrated parameter values were different for each model, which demonstrated that parameters estimated for one model cannot be transferred to other models without a new model calibration, even if they represent the same physical background. However, this is not surprising as only simplified conceptual models were tested. In addition, the analysis indicated that the pervious area parameters played a secondary role when modelling highly urbanised catchments. This suggests that, for practical applications, parameters relating to the pervious areas do not have to be calibrated and default values could be used when applied to urbanised catchments.

The most widely adopted water quality models, the build-up/washoff and simple regression equations, were tested. Even with the robust calibration and parameter sensitivity approach used, it was clear that these models poorly represent reality and their predictions presented a high level of uncertainty. This opposes to most of publications that showed a good agreement between measured data and simulated data, but is explained by the fact that such studies were based on a few single events. While the two water quality models tested were both sensitive to wet weather related parameters, the build-up/wash-off model was not very sensitive to the dry weather related parameters. Recommendations were made to aid the improvement of existing models and also the development of new model formulations. For example, results indicated that future work on the development of better water quality models should focus on formulations that use routed variables (e.g. routed rainfall or runoff) rather than 'unrouted' variables.

The uncertainty analysis showed that the total predictive uncertainty bands were larger than the uncertainty bands resulting from parameter uncertainty only. While the total uncertainty bands covered the bands resulting from parameter uncertainty only, it was observed that the contribution of the later was minor. This indicates that the predictive uncertainty due to other sources is more important (e.g. measured input data including spatial rainfall distribution, model formulation and assumptions and selected objective function).

In summary, one of the main contributions of this work was to demonstrate the limitations of the currently used stormwater models as they all presented large uncertainty, mainly on the side of water quality modelling.

Impact of input and calibration data uncertainties on stormwater models

The impact of the input and calibration data uncertainties on stormwater models in terms of performance, sensitivity and predictive uncertainty was assessed by means of a rather simple approach for global assessment of uncertainties in urban drainage models. A coupled urban stormwater model (a simple rainfall runoff model was coupled with a commonly used build-up/wash-off model) and error models were developed to estimate the uncertainty associated with input and calibration data.

It was demonstrated that random errors in all input and calibration data had a minor impact on the model performance and sensitivity. Another finding was that systematic errors in input and calibration data influenced the model sensitivity (represented mainly by the position of the peak of parameter distributions). In most of the scenarios, the model performance was fully compensated by the parameters. Parameters were unable to compensate only in some scenarios where the uncertainty in the input and calibration data were represented using extreme worst-case scenarios. As such, in these few worst case scenarios, the model performance was reduced considerably.

Results specifically from the water quality modelling suggested that the model sensitivity was not significantly impacted by the calibration data errors, which might be due the fact that the model cannot reproduce TSS concentrations even when the 'true' measured dataset is used. One of the main contributions of this research task was to demonstrate that the current <u>main</u> limitation in water quality modelling is related to poor model structure, and that, even though it is still important, the main limitation may not be errors in measured data.

8.4 Future work

This research has provided a comprehensive overview of the different sources of uncertainties in stormwater models (with different level of complexities) and how the different sources impact on parameter sensitivity and the resulting predictive uncertainty. Future research, however, is needed to extend the understanding of the different sources of uncertainties in stormwater models.

Methods for global sensitivity and uncertainty analysis in stormwater modelling

Results from this thesis suggested that the methods currently used to assess uncertainties in stormwater models are uncertain themself, mainly because of their subjectivity (e.g. user defined thresholds). In the author's opinion, there is a lack of appropriate uncertainty analysis methods, and therefore future studies should focus on the development of more robust and far less subjective global sensitivity and uncertainty analysis methods. In addition, attention should be paid to the development of methodologies to assess uncertainties in the absence of measured data (i.e. ungauged systems), that in fact is the main problem in hydrological modelling field.

The main issue with most of the Bayesian approaches is related to the fact that they rely on a number of assumptions related to the structure of the model errors (residuals). The data transformation used here to meet the normality assumption altered the data by influencing the emphasis on various parts of the hydrographs (to obtain a more even distribution of the flows). This means that, in the transformed space, the chosen least square based likelihood function no longer focuses on the peaks (as initially desired), but on medium and low flows. Future work should always ensure that all assumptions are verified, for such future studies should focus on other data transformation methods, and on the investigation of alternative formal likelihood functions to accommodate correlated and non - normal model residuals.

Sources of uncertainties in stormwater modelling

The impact of calibration data availability (i.e. different sections of the calibration dataset) on the model sensitivity and predictive uncertainty should be evaluated and incorporated into the global approach for modelling uncertainties. Besides contributing to a better understanding of the impact of calibration data availability on the total model uncertainty, this would be useful to guide future applications of the model, mainly when only a limited number of events is available in the calibration dataset.

Future work should focus on the evaluation of structural errors (mainly related to the model conceptualisation, equations, numerical methods and boundaries) as they seem to be a major contribution to the total model uncertainty.

The proposed framework used to evaluate the impact of the input and calibration data on the sensitivity and uncertainty of the tested stormwater models should be applied to other stormwater models. In addition, alternative error models to estimate the errors associated with measured data should be developed and tested. These would be useful to further validate the application of the method for a range of models and data.

Improving water quality modelling

It is very likely that semi-physically based conceptual models are just not able to cope with the stochasticity of the pollution generation processes. As such, further work should focus into development of radically different water quality models to those currently used in practice.

Engagement with stormwater professionals

In the author's opinion, it is crucial that the researchers working with the assessment of uncertainties in stormwater models should engage with other stormwater professionals (e.g. decision makers, planners, etc.) in order to discuss the best ways to communicate and use the quantified uncertainty. This would be mainly beneficial to provide strategic directions for the future of uncertainty assessment in the urban drainage field.

Appendix A

Journal paper co-authored by the candidate

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. (2012). Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth*, Parts A/B/C 42–44(0): 3-10.

Physics and Chemistry of the Earth 42–44 (2012) 3–10

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

journal homepage: www.elsevier.com/locate/pce

Assessing uncertainties in urban drainage models

A. Deletic [a,*], C.B.S. Dotto [a], D.T. McCarthy [a], M. Kleidorfer [b], G. Freni [c], G. Mannina [d], M. Uhl [e], M. Henrichs [e], T.D. Fletcher [a], W. Rauch [b], J.L. Bertrand-Krajewski [f], S. Tait [g]

[a] Department of Civil Engineering, Centre for Water Sensitive Cities, Monash University, Victoria 3800, Australia
[b] Unit of Environmental Engineering - University Innsbruck, Technikerstrasse 13;6020 Innsbruck, Austria
[c] Facoltà di Ingegneria ed Architettura, Università di Enna "Kore", Cittadella Universitaria, 94100 Enna, Italy
[d] Dipartimento di Ingegneria Civile, Ambientale ed Aerospaziale, Università di Palermo, Viale delle Scienze, 90128 Palermo, Italy
[e] Laboratory for Water Resources Management, Muenster University of Applied Sciences, Department of Civil Engineering, Corrensstr. 25, FRG 48149 Muenster, Germany
[f] Université de Lyon, INSA Lyon, Laboratoire de Génie Civil et d'Ingénierie Environnementale, 34 avenue des Arts, F-69621 Villeurbanne cedex, France
[g] Pennine Water Group, The University of Bradford, Richmond Road, Bradford BD7 1DP, United Kingdom

ARTICLE INFO

Article history:
Received 31 January 2011
Accepted 13 April 2011
Available online 19 April 2011

Keywords:
Urban drainage models
Uncertainties
Input data
Calibration data
Sensitivity analysis

ABSTRACT

The current state of knowledge regarding uncertainties in urban drainage models is poor. This is in part due to the lack of clarity in the way model uncertainty analyses are conducted and how the results are presented and used. There is a need for a common terminology and a conceptual framework for describing and estimating uncertainties in urban drainage models. Practical tools for the assessment of model uncertainties for a range of urban drainage models are also required to be developed. This paper, produced by the International Working Group on Data and Models, which works under the IWA/IAHR Joint Committee on Urban Drainage, is a contribution to the development of a harmonised framework for defining and assessing uncertainties in the field of urban drainage modelling. The sources of uncertainties in urban drainage models and their links are initially mapped out. This is followed by an evaluation of each source, including a discussion of its definition and an evaluation of methods that could be used to assess its overall importance. Finally, an approach for a Global Assessment of Modelling Uncertainties (GAMU) is proposed, which presents a new framework for mapping and quantifying sources of uncertainty in urban drainage models.

1. Introduction

Uncertainty is intrinsic in any modelling process and originates from a wide range of sources, from model formulation to the collection of data to be used for calibration and verification. Uncertainties cannot be eliminated, but their amplitude should be estimated and, if possible, reduced. It is necessary to understand their sources and impact on model predictions. For example, the confidence level of a model's predictions should be included in every modelling application. Beven (2006) reported that there are many sources of uncertainty that interact non-linearly in the modelling process. However, not all uncertainty sources can be quantified with acceptable levels of accuracy, and the proportion of uncertainty sources being ignored may be high in environmental modelling investigations (Harremoës, 2003; Doherty and Welter, 2010).

In the literature, the following sources of uncertainties are listed (e.g. Butts et al., 2004): (i) model parameters, (ii) input data, (iii) calibration data, and (iv) model structure. The impacts of calibration methods and data availability are also recognised. Each of these sources is discussed below.

When dealing with complex urban drainage models, calibration may lead to several equally plausible parameters sets, reducing confidence in the model predictions (Kuczera and Parent, 1998). The concept that a unique optimal parameter set exists should therefore be replaced by the concept of "equifinality" (Beven, 2009) in which more than one parameter set may be able to provide an equally good fit between the model predictions and observations. Many published studies have dealt with the impact of uncertainties on model parameters, also known as sensitivity analysis (Kanso et al., 2003; Thorndahl et al., 2008; Dotto et al., 2009). Some studies used the results of a model sensitivity analysis to produce parameter probability distributions (PDs), which reflect how sensitive the model outputs are to each parameter (e.g. Marshall et al., 2004; Dotto et al., 2010a; McCarthy et al., 2010); while other studies used the sensitivity analysis to screen parameters for further analysis (e.g. Reichl et al., 2006; Haydon and Deletic, 2007). In most cases, model sensitivity results were also used to estimate confidence intervals around the model's outputs (e.g. Yang et al., 2008; Li et al., 2010).

* Corresponding author.
[redacted]

1474-7065/$ - see front matter Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.pce.2011.04.007

Impacts of input data uncertainties on urban drainage modelling are far less understood. Their importance, however, is widely studied in related areas (Kuczera et al., 2006). For example, the impact of systematic rainfall uncertainties on the performance of non-urban catchment models was recognised and assessed by Haydon and Deletic (2009). Work has also been completed on the propagation of input data uncertainties through urban drainage models (Rauch et al., 1998; Bertrand-Krajewski et al., 2003; Korving and Clemens, 2005). However, in these studies, the models were first calibrated assuming that measured inputs and outputs were true (no-error), and the impacts of input data uncertainties were then propagated through the models, while keeping the model parameters fixed. Recently, Kleidorfer et al. (2009a) and Freni et al. (2010) attempted to assess how input data uncertainties impact model parameters, investigating the interactions between these two sources of uncertainties. Freni and Mannina (2010) attempted to isolate the contribution of different sources of uncertainty in a complex integrated urban drainage model.

Research on the impact of calibration data on the accuracy of drainage models has focused on the effectiveness of the calibration and verification processes. Many studies have examined how to divide the available data into calibration and verification sets (McCarthy, 1976; Klemes, 1986; Vaze and Chiew, 2003; Wagener et al., 2004). A few recent papers (e.g. Mourad et al., 2005; Dotto et al., 2009) evaluated how the number of events used in calibration and verification of urban drainage models impacts on their predictive uncertainty. On the other hand, there is little work reporting on how uncertainties in measured calibration data impact on the model's predictive capacity. However, large uncertainties in measured urban discharges and water quality have often been reported (e.g. Bertrand-Krajewski, 2007; McCarthy et al., 2008), thus clearly demonstrating that calibration data sets may in themselves be a significant source of uncertainty in the model calibration process. In fact, McCarthy (2008) showed the influence of calibration data uncertainty on the calibration of a simple rainfall-runoff model.

There are many studies on the effectiveness of calibration algorithms. For example, Gaume et al. (1998) showed that different calibration methods can lead to different parameter sets, which demonstrate a similarly good fit between measured and simulated data. This can occur as a result of difficulties in finding a global minima, especially for systems where the objective/criteria function surface is nonlinear. It is evident that these problems become more important as model complexity increases (Silberstein, 2006), or where models are ill-posed (Dotto et al., 2009). Therefore it is not surprising that some calibration algorithms simply cannot find the global minima in rather complex urban drainage models (Kanso et al., 2003).

Given the wide range of communities and applications in which uncertainty is studied, there is no general consensus in the literature with regard to the terminology used. For example, the terms "sensitivity" and "uncertainties" are often used interchangeably and yet have distinctly different meanings. A further example is that some input variables that could be measured, but are also refined through calibration processes (such as, effective imperviousness in rainfall-runoff modelling), are sometimes regarded as fixed inputs and at other times as model parameters. These terminology problems need to be addressed so as to improve the communication between research groups, thus the coherence and applicability of future studies.

Despite previous work attempting to unify definitions and approaches of uncertainty evaluation (e.g. Walker et al., 2003), no universal framework and methodology for categorising and assessing modelling uncertainties has been accepted. Indeed, Montanari (2007) stated that uncertainty assessment in hydrology suffers from a lack of a coherent terminology and hence a systematic approach.

This paper is a contribution in the debate to develop common terminology and a conceptual framework for accounting for uncertainties in urban drainage modelling. It outlines a Global Assessment of Modelling Uncertainties (GAMU), which presents a new framework for mapping and quantifying sources of uncertainty in urban drainage models.

2. Methods

The International Working Group on Data and Models, which works under the IWA/IAHR Joint Committee on Urban Drainage (JCUD), conducted several workshops focused on uncertainties in monitoring and modelling of urban drainage systems:

(1) 'Integrated Urban Water Management Modelling: Challenges and Developments', Melbourne, Australia, 2006, in conjunction with the 7th Urban Drainage Modelling and 4th Water Sensitive Urban Design conferences (7UDM & 4WSUD);
(2) 'Uncertainties in data and models', Lyon, France, 2007, as part of the 6th Novatech conference; and,
(3) 'Challenges in monitoring and modelling of stormwater treatment systems', Edinburgh, UK, 2008 as part of the 11th International Conference on Urban Drainage (11ICUD).

This paper represents the outcome of these workshops. The literature, guidelines and standards on uncertainties in measurements (Bich et al., 2006; ISO, 2008, 2009a,b) were also consulted, as well as recent relevant work on uncertainties. This paper thus presents a review of the state of the art, and an attempt to harmonise concepts, definitions and protocols.

3. Proposed framework for a Global Assessment of Modelling Uncertainties (GAMU)

The first step in the proposed uncertainty framework is to map the sources of uncertainty and their links; their contribution and significance are then evaluated. Finally, the propagation of all sources simultaneously provides an analysis of their effect on the model sensitivity and consequently on the uncertainty of the model predictions.

3.1. Mapping uncertainties

The majority of urban drainage models require calibrating prior to use. This calibration process is referred to as the 'inverse problem' (Gallagher and Doherty, 2007), whereby parameter values are determined from measured calibration input data, calibration output data and the model structure by applying an objective function. When using models for prediction outside of calibration, or when models are simply used with estimated parameter values (from expert knowledge, literature or defaults), the process is known as the 'forward problem'.

A generic modelling framework was therefore adopted, for which the following information is needed (Fig. 1): model structure MS (i.e. relationships, linkages and numerical methods), input data ID (e.g. rainfall or potential evapotranspiration time series) and model parameters P (e.g. effective impervious area, linear reservoir lag-time parameters in rainfall-runoff conceptual models). For the inverse problem, the following information is also needed: calibration input data (e.g. rainfall intensity time series), measured calibration output data (e.g. flow time series), calibration algorithms CA and objective functions OF selected by the modeller according to the model requirements (e.g. sum of the squared errors).

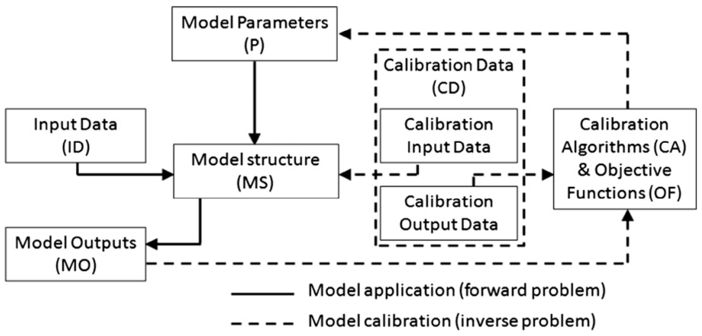


Fig. 1. General modelling framework.

Three key groups of uncertainty sources mapped in this framework are outlined below and in Fig. 2.

(I) *Model input uncertainties*: Inputs that are required to run either a calibrated or a non-calibrated model can be grouped into the following categories, which their associated uncertainties should be propagated through the model:

1. Input data (ID) – both random and systematic effects have to be assessed in the input data collection process (these may be described statistically using the actual measurement information or simply estimated).
2. Model parameters (*P*) – uncertainty in their calibrated values or estimates.

(II) *Calibration uncertainties*: That are related to the processes and data used in model calibration. This source is mainly due to:

3. Calibration data uncertainties due to measurement errors in both inputs and outputs (CD-M), that are dependent on the quality of the monitoring program and instruments used in the collection of the data sets, including the temporal resolution of the time series, data collection and validation procedures and data manipulation protocols.
4. Selection of appropriate calibration input and output datasets (CD-S), which is linked to the choice of the calibration variable (e.g. the of use concentrations or loads to calibrate a water quality model) and the amount of data available for calibration (e.g. number of storm events, length of time series).
5. Calibration algorithms (CA), which depends on the algorithm used for finding the appropriate sets of parameters.
6. Objective functions (OF) used in the calibration process; these need to be appropriate for the modelling application.

(III) *Model structure uncertainties*: Which depend on how well the model simulation represents the systems and processes. These can include:

7. Conceptualisation errors, such scale-issues or omitting key processes.
8. Equations, which could be ill posed and thus inadequately represent the process.
9. Numerical methods and boundary conditions, which can be ill defined leading to inaccurate solutions (e.g. numerical dispersion or instabilities).

Fig. 2 indicates that sources of uncertainties are interlinked. For example, uncertainties in input data and calibration procedures will at the same time impact on the model's sensitivity to its calibration parameters. In fact, all identified sources of uncertainties

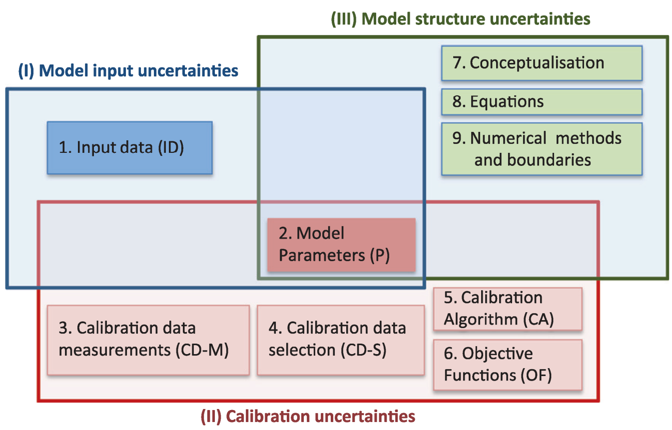


Fig. 2. The key sources of uncertainties in urban drainage models and links between them.

will impact the model parameter values. Further, the model development and calibration process needs to be strongly related to the model application. A model used to predict average annual discharge might be built and calibrated differently to a model used to predict hydrographs and pollutographs. As discussed in the Introduction, the model structure (e.g. conceptualisation, choice of equations and numerical methods) impacts on this process, since ill-posed models are notoriously difficult to calibrate. Therefore, in Fig. 2, the model parameter uncertainties are placed at the intercept of all three categories.

3.2. Model input uncertainties

In general, depending on their type and use in the model, model inputs can either be measured or estimated. The two identified sources (Fig. 2) are discussed below in detail.

Source (1): Input data uncertainties (ID) are defined as uncertainties in any input data that can be either measured or estimated. In the first case, input data are measured using appropriate monitoring protocols and instruments (e.g. rainfall intensities measured by a rain gauge). Uncertainties in the measured input data are generally caused by (i) systematic and/or (ii) random errors. If input data are not directly measured, their uncertainty can be elucidated using accepted statistically based methods (Garthwaite et al., 2005): in both cases, they can be described by probability distributions. For example, typical probability distributions for measured and estimated input data are Gaussian and uniform PDs, respectively. In urban drainage applications, effective impervious area is one of the most common inputs that can be estimated using GIS/terrain maps associated with drainage plans, but is often also used as a calibration parameter (see Source (2) below) depending on the modelling approach. It is frequent in urban drainage modelling that some input data, although theoretically measurable, are either estimated or replaced by the use of a model parameter which is then calibrated.

Uncertainties in measured input data can be characterised and assessed according to international standards (ISO, 2007, 2008, 2009a,b) or related literature such as Bertrand-Krajewski and Muste (2007). In these standards, uncertainty is defined as the variable associated with a measurement result which characterises the dispersion of the values which could be reasonably attributed to the measured variable. As a first approximation in normal distributions, uncertainty can be considered as equivalent to the standard deviation. This probabilistic approach allows measurement result to be provided as a most probable mean value given with its 95% confidence or coverage interval, or as a most probable mean value given with its probability distribution (see ISO (2008, 2009a,b) for more details). In simple cases where normal distributions can be assumed, uncertainty is estimated as the standard deviation derived from repeated measurements. This is usually referred to as the Type A method to evaluate uncertainties. In most frequent cases in urban drainage, repeated measurements are not possible and uncertainties are estimated by means of two other methods: (1) the Type B method which applies the Law of Propagation of Uncertainties (LPU) when the required underlying hypotheses (linearity, normality, and narrow distributions) are verified, and (2) the Monte Carlo method which propagates probability distributions of any type (uniform, normal, log-normal, empirical, etc.) and is the most generic method with less restrictive hypotheses for its application. In this case, probability distributions are determined for each variable used in the measurement process from the best available knowledge.

Input data uncertainties are often propagated in model applications by methods based on Monte Carlo simulations. As a first step example, one may multiply the measured variable with the factor

$$\mathrm{ID}_{FACTOR} = f(\delta, \varepsilon) \tag{1}$$

in which δ is a systematic variability (e.g. an offset value, or results from an error model application) and ε is a random variability, ideally sampled from a distribution that represents random input uncertainties. This means that an input data error model with two additional model parameters δ and ε is introduced. The values for δ and the distribution for ε should be assessed using the best knowledge on the monitoring protocol applied (e.g. following ISO standard and by gathering additional data on possible systematic uncertainties); or it can be estimated together with model parameters in an inverse modelling approach. In the forward modelling approach, uncertainties in the input data can be propagated through the model to the output by using Monte Carlo methods. For example, for rainfall data, an ID_{FACTOR} can be assumed as a simple sum of δ, which is an approximated constant, and ε, which is sampled from a uniform distribution (e.g. Rauch et al., 1998; Haydon and Deletic, 2009). However, this approach is rather simplistic and the uncertainties in the input data are better modelled using our best knowledge about the measurement process (e.g. information on the accuracy in the equipment used, sampling procedure, etc.).

Both measured and estimated input data can be affected by additional "long-term prediction uncertainties" which occur when trying to predict long-term environmental change effects (e.g. land-use, climate change effects). Such predictions often contain substantial uncertainties, but as they cannot be compensated during model calibration they are not covered here.

It should be noted that the method described above differs from that typically used to quantify measurement uncertainty, since it is not only the measurement uncertainty that needs quantification, but rather how uncertainty in input data impacts model results. This difference is necessary since the assessment of measurement uncertainties requires that the measurements first be corrected for all recognised systematic errors (ISO, 2009a). ISO (2009a) states that since the measurements have been corrected for systematic errors using a calculated correction factor or offset value, they now contain (1) the random errors affecting the chosen correction value since it cannot be exactly known and (2) the same random errors that existed prior to the correction. As such, there is no difference in nature in the uncertainties derived from a random error and those originating from a correction factor used to adjust the dataset for systematic errors (hence both error types are to be propagated similarly).

In the case of model application (forward problem in Fig. 1), the propagation of uncertainties associated to input data is often processed to the PDs by means of Monte Carlo methods, where inputs are perturbed using, for example, Eq. (1) (or any other appropriate function) for thousands of possible realisations. In other words, the inputs are multiplied by ID_{FACTOR} and the model is run many times. The results are then represented by constructing mean and 95% confidence intervals for each model output. If the confidence intervals are small, it can be concluded that uncertainties do not significantly impact the model results, and vice versa. Small intervals are usually possible if input uncertainties are small, or if the model calibration compensates these uncertainties. As in all other analyses, it is important to propagate all inputs simultaneously because of possibilities that uncertainties in different variables are not independent and interact. Accounting for correlated input data and their correlated PDs is of particular importance when attempting to estimate an overall uncertainty.

Source (2): Model calibration parameter uncertainties (P). This is also referred to as the "sensitivity of a model to its parameters". The aim is to derive probability distributions for the given parameters, and the extent and shape of the confidence region of modelling predictions around a specified measured output variable. Since parameters in urban drainage models can be highly correlated (commonly the case for water quality models, e.g. Dotto et al., 2010b), it is essential to perform a global sensitivity of parameters where all parameters are varied simultaneously over the whole range of possible parameter values (as opposed to the local sensitivity analysis where sensitivity is only investigated at one point in parameter space and one-at-a-time (OAT) methods where one parameter is varied with others held fixed).

The literature on sensitivity of general hydrological models is extensive, and the key methods and concepts already used in water resources modelling are applicable to urban drainage. Many of these methods, applied in model calibration (inverse problem in Fig. 1), refer more or less strictly to Bayes' (1763) principle. They range from formal Bayesian approaches (e.g. Markov Chain Monte Carlo – MCMC, like MICA (Doherty, 2003) or DREAM (Vrugt et al., 2008)) to less formal likelihood methods (e.g. Generalized Likelihood Uncertainty Estimation; GLUE – (Beven and Binley, 1992)). According to Freni et al. (in press), the classical Bayesian method is more effective at discriminating models according to their uncertainty, but the GLUE approach performs similarly when it is based on the same founding assumptions as the Bayesian method. However, this conclusion is still debated (e.g. Beven, 2009; Vrugt et al., 2009).

The International Working Group on Data and Models is currently working on comparison of some of the most popular calibration and sensitivity analysis approaches, including: (1) GLUE developed by Beven and Binley (1992), (2) The Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) by Vrugt et al. (2003), (3) Amultialgorithm, genetically adaptive multiobjective method (AMALGAM) by Vrugt and Robinson (2007), and (4) The classical Bayesian approach based on a MCMC method (implemented in the software MICA by Doherty, 2003). These methods were tested for both a simple rainfall-runoff model (KAREN – Rauch and Kinzel, 2007) and a simple water quality model using the same datasets collected at a single site in Melbourne, Australia. Preliminary results showed that all methods tested are eligible to analyse uncertainties of urban drainage models, to estimate parameter sensitivity, parameter probability distributions and consequently uncertainty bands of model output. However, each method has its specific advantages and drawbacks. Special attention has to be given to the computational efficiency (i.e. number of iterations required to generate the PDs of parameters) as computational time is often a limiting factor of uncertainty analysis. So far it was found that MICA and AMLAGAM produce results quicker than GLUE. However, GLUE requires the lowest modelling skills and is easy to implement. An important step in the application of all methods is comprehensive posterior diagnostics of parameter distributions and uncertainty bands obtained to ensure that the distributions have converged and implicit assumptions are valid. Further investigations are being undertaken in order to provide insights on the advantages and disadvantages of different approaches.

3.3. Calibration uncertainties

Source (3): Measured calibration data uncertainties (CD-M) are uncertainties in the measured data collected for possible use during calibration (e.g. flow and pollutants times eries). As in all other measured variables, errors could be systematic and/or random, and probability distributions are used to describe their uncertainty, as for input data. So Eq. (1), or the other approaches discussed under Source (1), could be applied to estimate measured calibration data uncertainties.

It is well understood that the techniques used to measure urban discharges and associated water quality are of limited accuracy (e.g. Bertrand-Krajewski, 2007; McCarthy et al., 2008). However, the propagation of these uncertainties has not been widely applied in practice. Recently, Freni and Mannina (2010) assessed the different components of uncertainty in an integrated urban drainage model using a variance decomposition method. Interestingly, they found that the uncertainty contribution of calibration data progressively reduced from upstream to downstream sub-models as they became overwhelmed by other error sources. Others in the literature which have considered calibration data uncertainty usually assess model accuracy by plotting the uncertainty bars (usually 95% confidence interval or just one standard deviation) around the measured data, alongside the model outputs. In general, it is proposed that the model is doing well if its outputs fall within the uncertainty bars around the measured data. However, this cannot be regarded as a proper and rigorous propagation of calibration uncertainties. It is therefore proposed that this should be improved, and that the calibration data uncertainties be explicitly accounted for while the parameters are calibrated.

Source (4): Calibration data selection (CD-S) is focussed on using the appropriate calibration variables and associate data sets that will best suit the model application (e.g. selecting the right amount of data for model assessment). For example, there has been discussion on whether to calibrate load models using pollutant concentrations or fluxes, with fluxes most commonly used. McCarthy (2008) demonstrated that using instantaneous concentrations for calibration produced more accurate predictions than using instantaneous fluxes. This was thought to be caused by the fact that the flux calibration process is affected by poorer quality input data because measured flow rates are used to estimate measured fluxes, whilst modelled flow rates (which are calibrated to measured flow rates) are used in the prediction of modelled fluxes. However, Dembélé (2010) observed that calibrating various types of models for a wide range of pollutants using event loads gives more accurate predictions than calibrating them using event mean concentrations. This indicates conclusions based on some data sets, models and calibration variables are difficult to be generalised: more research is needed to identify the most appropriate calibration parameters to use.

If calibration data are not representative (i.e. do not represent all possible contexts and ranges of phenomena and values to be simulated by the model), the calibrated model parameters will not be accurately estimated for the range of applicability of the model (e.g. calibrating a rainfall-runoff model during summer periods will produce model parameters which will likely not reflect winter period processes). For example, Mourad et al. (2005) used a random sampling methodology to understand the impact of data availability (i.e. number of events) on the calibration of several urban stormwater quality models. They found that, in order to adequately calibrate these models, it was often required to use the majority (between 60% and 100%) of the available dataset during calibration.

In the case of spatially distributed systems, it is neither possible nor sensible to measure the complete system characteristics, and the question is raised about how many measurement sites are necessary. Kleidorfer et al. (2009b) evaluated the impact of the number of measurement sites used for calibration of combined sewer systems and showed that the number of required sites is influenced by the time period used for calibration. For example, a similar calibration performance can be reached when using 30% of all available sites for calibration and a time period of one year, as compared to using 60% of all available sites with five single events.

Furthermore, calibration data availability impacts not only the uncertainty of a model's prediction outside the calibration period (Vaze and Chiew, 2003; Mourad et al., 2005; McCarthy, 2008), but also the model's parameter probability distributions (McCarthy, 2008).

The assessment of this type of uncertainty on a model should be incorporated into the global approach for modelling uncertainties, and the method presented by Mourad et al. (2005) could be easily incorporated for this purpose. For example, for a rainfall-runoff model, a number of events could be randomly (or systematically) selected and these events could then be used to perform a sensitivity analysis of the model outputs to parameter values. These results could then be compared with the results obtained when all the data was used for the analysis, to determine the impact of data availability. For example, Dembélé (2010) applied the Leave-One-Out Cross Validation (LOOCV) method (Rudemo, 1982), which is particularly useful when only a limited number of events is available in the calibration dataset.

Source (5): Calibration algorithms (CA) used during model parameter optimisation can produce significant uncertainties in the model's predictive performance (Beven and Freer, 2001). There are many calibration algorithms available which can automatically calibrate model parameters. However, even when using such complex algorithms, which are capable of calibrating highly non-linear functions, there is never certainty that the best solution (or global optimum) will always be found (Beven and Freer, 2001; Wagener et al., 2004). This can be caused by several conditions, but calibration which results in a non-global optimum can often be the fault of the user, who has (1) incorrectly 'wrapped' the calibration algorithm around the chosen model, and/or set incorrect boundary conditions, or (2) chosen an algorithm which cannot solve the specified model (e.g. a linear algorithm used to solve a nonlinear function). Several tools can now calibrate models using a range of different algorithms, the results of which could be used to help quantify this type of uncertainty. Therefore, the best approach is to use several calibration algorithms for a specific model and its application and select the best outcome. Ideally, the algorithm or algorithms tested will have been selected based on the suitability of their criteria for the particular model. Another possibility is the use comprehensive uncertainty analysis techniques (see Source 2) to explore the likelihood surface in a wider range of the parameter space and to identify local minima which can cause problems in the calibration process.

Source (6): Objective functions (OF) used in the calibration process. Models are often calibrated without considering the implications of the selected criteria/objective function (see Wagener et al., 2004). Different objective functions can influence parameter distributions (magnitude and shape), and therefore impacting the apparent sensitivity of the modelled results to each parameter and the general uncertainty of model predictions. All objective functions sacrifice the fit of a certain portion of the dataset, to achieve a good performance in another portion (Wagener et al., 2004). McCarthy (2008) found that using a least-squares objective function to calibrate an urban rainfall-runoff model over-emphasised peak flow rates, resulting in poor predictive performance of events which only had smaller flows. However, changing this objective function to a less biased function (similar to Chi-squared) decreased the model's performance slightly for peak estimation, but substantially increased the accuracy of low flow estimation. The choice of objective function can also impact on how well the model will predict outside its calibration dataset, with certain objective functions resulting in better estimates of the parameter distributions. As such, it is essential that objective functions are matched to the purpose and requirements of the modelling application.

Most calibration tools (e.g. PEST – (Doherty, 2004); CALIMERO – (Kleidorfer et al., 2009a); KALIMOD – (Uhl and Henrichs, 2010)) and model uncertainty assessment tools (e.g. MICA, GLUE) can use alternate or multiple objective functions, and, as such, these tools should be used to assess the impact of different objective function choices on model results. It may also be considered that, for a given model, different sets of parameters could be applied for different contexts, e.g. one set for dry weather and another set for storm weather. With this approach, the aim is not to identify the unique model for all contexts, but to distinguish models for specific ranges of application.

3.4. Model structure uncertainties

Uncertainties are introduced through simplifications and/or inadequacies in the description of spatially and temporally distributed real-world processes. Three main sources (see Fig. 2) are identified, but it is possible that other factors could be causing inaccuracies, as well as coarse mistakes. Human error in model development (e.g. derivation of equations, coding, etc.) may be the major problem that cannot easily be evaluated. However, the authors recognise that it is very difficult, and sometimes not possible (e.g. in the case of human error), to distinguish between these causes. In general, it is a complex task, which requires a very advanced understanding of the processes of the system and model development. Even if the estimation of model structure uncertainty for a single model is not feasible and most of the time has to be assessed heuristically, we suggest to compare the performance of different models and thus establish which can better represent the system under investigation.

3.5. Global Assessment of Modelling Uncertainties (GAMU)

Assessing single sources of uncertainties independently from others is not appropriate, since there are often strong links between the sources (Fig. 2). Therefore, the approach for a Global Assessment of Modelling Uncertainties is recommended (Fig. 3) that has recently been proposed by Dotto et al. (2010b). The GAMU has three distinctive steps:

Step 1: Choosing analysis tools and datasets to minimise uncertainties: Each model application may require different analysis calibration tools/algorithms (CA), criteria/objective functions (OF), and datasets (CD-S) to minimise errors in the evaluation methods. Unfortunately, due to the long computational times required for detailed urban drainage models, it is very time consuming to determine the most apporpriate CA, OF and CD-S while still having to propagate the other uncertainties through the model (i.e. conduct Step 2 (below) for every possible CA, OF and CD-S). Therefore, it is necessary to select CA, OF and CD-S in a preliminary study. For example, it could be done by using simplified response surface based methods (Schellart et al., 2010) to estimate combined uncertainties. Tools such as CALIMERO or KALIMOD could be used to compare effectiveness of algorithms and OFs for the given model and its application, as well as to select adequate data sets for the next step of the analysis. It could be speculated that in this approach at least some uncertainnties due to sources CA, OF and CD-S will be minimised.

Step 2: Creating probability distributions of model parameters while simultaneously propagating all data uncertainties: The parameter PDs should be created by simultaneously propagating input data uncertainties (ID) and measured calibration data uncertainties (CD-M), as outlined in Fig. 3. The uncertainties in these data sets are assessed as outlined above; e.g. both the input data and calibration data uncertainties could be modelled by estimating their most probable parameters δ and ε in Eq. (1) and creating probability distributions of possible inputs and calibration data at any given time. The PDs of all model parameters are then generated using a Bayesian method (e.g. MICA, DREAM, GLUE, etc.) by sampling from the input and calibration data assumed distributions. In this approach, uncertainties due to Sources (5) and (6) (CA and OF) are replaced by uncertainties caused by the Bayesian method being used. Therefore, this leads to the fully calibrated model with the parameter PDs derived by taking into account uncertainties in inputs and calibration data, while using tools/algorithms that hopefully impose the smallest possible uncertainty. The process also yields information on the misfit between modelled and observed output datasets, known as residuals.

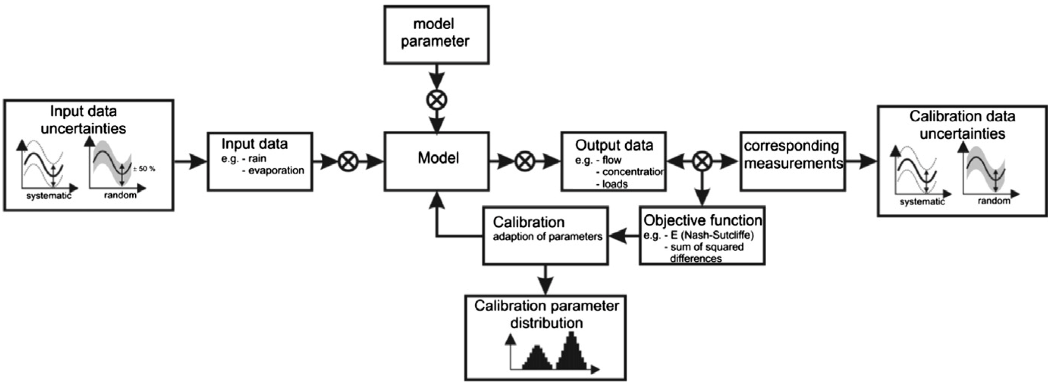


Fig. 3. A total error framework for urban drainage models.

The calibrated model is then used to determine model prediction uncertainties, typically for a dataset not used for calibration. This is done in the 'forward approach' (Fig. 1) where the model is applied to a new input dataset using the derived PDs of the model parameters to create the prediction bounds. The residuals from the calibration process are also used to understand the total predictive uncertainty, obtained by the addition of the error term to the simulated values.

Step 3: Comparing models: As discussed earlier, the authors are of the opinion that systematic and random effects due to model structure could be assessed only by comparing the performance of models applied for the same situation. Ideally, the proposed approach should be run for given models and situations and their effectiveness compared.

4. Conclusions

This paper presents an attempt of the JCUD International Working Group on Data and Models to develop and promote a framework for accounting and estimating the uncertainties in urban drainage models. The following key sources of uncertainties are accounted for: (I) *Model input uncertainties* including (1): input (measured and estimated) data uncertainties, (2): model parameter uncertainties; (II) *Calibration uncertainties* due to (3): measured calibration data uncertainties, (4): measured calibration data selection (availability and choice), (5): calibration algorithms, (6): objective functions used in the calibration process; and (III) *Model structure uncertainties* in conceptualisation, equations and numerical methods. They are highly interlinked, suggesting that assessing the impact of a single source is not going to be adequate and that simultaneous propagation of key sources of uncertainties is required. The importance of minimising uncertainties due to tools that are used in model assessment is also recognised. Framework for Global Assessment of Modelling Uncertainties (GAMU) is thus recommended, containing three major steps:

Step 1: Selecting analysis tools and data sets to minimise uncertainties;
Step 2: Creating probability distributions of model parameters while simultaneously propagating all data uncertainties; and
Step 3: Comparing different models for similar scenarios.

Due to the large computational times required for applying this approach, it is not expected that this method will be a standard procedure in everyday engineering practice. However, this method will contibute to an enhanced system understanding, and thus an improved assesment of the reliability of modelling results, especially when using new models or working under limited data availability.

References

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society London 53 (A), 370–418.

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

Bertrand-Krajewski, J.-L., Muste, M., 2007. Understanding and managing uncertainty. In: Fletcher, T.D., Deletic, A., (Eds.), Data Requirements for Integrated Urban Water Management, p. 333.

Bertrand-Krajewski, J.L., Bardin, J.-P., Mourad, M., Beranger, Y., 2003. Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. Water Science and Technology 47 (2), 95–102.

Beven, K., 2006. On undermining the science? Hydrological Processes 20, 3141–3146.

Beven, K., 2009. Comment on "Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?" by Jasper A. Vrugt, Cajo J.F. terBraak, Hoshin V. Gupta and Bruce A. Robinson. Stochastic Environmental Research and Risk Assessment 23 (7), 1059–1060. doi:10.1007/s00477-008-0283-.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6 (3), 279–298.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 246, 11–29.

Bich, W., Cox, M.G., Harris, P.M., 2006. Evolution of the 'guide to the expression of uncertainty in measurement'. Metrologia 43, S161–S166.

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. Journal of Hydrology 298 (1–4), 242–266.

Dembélé, A., 2010. MES, DCO et polluants prioritaires des rejets urbains de temps de pluie: mesure et modélisation des flux événementiels (TSS, COD and prioritypollutants in urbanwetweatherdischarges: monitoring and modelling of eventloads). PhD Thesis, INSA, Lyon, France, p. 286 (in French).

Doherty, J., 2003. MICA – Model-Independent Markov Chain Monte Carlo Analysis – User Manual, Watermark Numerical Computing, US, EPA.

Doherty, J., 2004. PEST – Model-Independent Parameter Estimation Version 10 – User Manual, Watermark Numerical Computing.

Doherty, J., Welter, D., 2010. A short exploration of structural noise. Water Resources Research 46, W05525. doi:10.1029/2009WR008377.

Dotto, C.B.S., Deletic, A., Fletcher, T.D., 2009. Analysis of parameter uncertainty of a flow and quality stormwater model. Water Science and Technology – WST 60 (3), 717–725.

Dotto, C.B.S., Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T., Rauch, W., 2010a. Stormwater quality models: performance and sensitivity analysis. Water Science and Technology – WST 62 (4), 837–843.

Dotto, C.B.S., Kleidorfer, M., McCarthy, D.T., Deletic, A., Rauch, W., Fletcher, T.D., 2010b. Towards global assessment of modelling errors. In: 6th International Conference on Sewer Processes and Networks (SPN). Gold Coast, Australia.

Freni, G., Mannina, G., 2010. Uncertainty in water quality modelling: the applicability of variance decomposition approach. Journal of Hydrology 394 (3–4), 324–333.

Freni, G., Mannina, G., Viviani, G., in press. Urban runoff modelling uncertainty: comparison among Bayesian and pseudo-Bayesian methods. Environmental Modelling & Software, 1–11.
Freni, G., Mannina, G., Viviani, G., 2010. The influence of rainfall time resolution for urban water quality modelling. Water Science and Technology – WST 61 (9), 2381–2390.
Gallagher, M., Doherty, J., 2007. Parameter interdependence and uncertainty induced by lumping in a hydrologic model. Water Resources Research 43, W05421, doi:05410.01029/02006WR005347.
Garthwaite, P.H., Kadane, J.B., O'Hagan, A., 2005. Statistical methods for eliciting probability distributions. Journal of American Statistical Association 100 (470).
Gaume, E., Villeneuve, J.P., Desbordes, M., 1998. Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. Journal of Hydrology 210, 38–50.
Harremoës, P., 2003. The role of uncertainty in application of integrated urban water modeling. In: International IMUG Conference. Tilburg, Netherlands.
Haydon, S., Deletic, A., 2007. Sensitivity testing of a coupled Escherichia coli – Hydrologic catchment model. Journal of Hydrology 338 (3–4), 161–173.
Haydon, S., Deletic, A., 2009. Model output uncertainty of a coupled pathogen indicator-hydrologic catchment model due to input data uncertainty. Environmental Modelling & Software 24 (3), 322–328.
ISO, 2007. ISO/TS 25377:2007. Hydrometric uncertainty guide (HUG). November, Geneva, Switzerland, p. 62.
ISO, 2008. ISO/IEC Guide 98-3/Suppl.1:2008(E). Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) Supplement 1: Propagation of distributions using a Monte Carlo method. December 2008, Geneva, Switzerland, p. 98.
ISO, 2009a. ISO/IEC Guide 98-1:2009(E). Uncertainty of measurement – Part 1: Introduction to the expression of the uncertainty in measurement. September 2009, Geneva, Switzerland, p. 32.
ISO, 2009b. ISO/IEC Guide 98-3/S1/AC1:2009(E). Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995), Supplement 1: Propagation of distributions using a Monte Carlo method, Technical corrigendum 1. May 2009, Geneva, Switzerland, p. 2.
Kanso, A., Gromaire, M.-C., Gaume, E., Tassin, B., Chebbo, G., 2003. Bayesian approach for the calibration of models: application to an urban stormwater pollution model. Water Science and Technology 47 (4), 77–84.
Kleidorfer, M., Deletic, A., Fletcher, T.D., Rauch, W., 2009a. Impact of input data uncertainties on stormwater model parameters. Water Science and Technology – WST 60(6), 1545–1554.
Kleidorfer, M., Möderl, M., Fach, S., Rauch, W., 2009b. Optimization of measurement campaigns for calibration of a conceptual sewer model. Water Science and Technology 59 (8), 1523–1530.
Klemes, V., 1986. Dilettantism in hydrology: transition or destiny? Water Resources Research 22 (9), 177–188.
Korving, H., Clemens, F., 2005. Impact of dimension uncertainty and model calibration on sewer system assessment. Water Science and Technology 52 (5), 35–42.
Kuczera, G., Kavetski, D., Franks, S., Thyer, M., 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters. Journal of Hydrology 331 (1–2), 161–177.
Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211 (1–4), 69–85.
Li, L., Xia, J., Xu, C.-Y., Singh, V.P., 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. Journal of Hydrology 390 (3–4), 210–221.
Marshall, L., Nott, D., Sharma, A., 2004. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall–runoff modeling. Water Resources Research 40.
McCarthy, D.T., 2008. Modelling microorganisms in urban stormwater. PhD thesis, Civil Engineering. Melbourne, Monash University, Australia, p. 488.
McCarthy, D.T., Deletic, A., Mitchell, V.G., Fletcher, T.D., Diaper, C., 2008. Uncertainties in stormwater E. coli levels. Water Research 42 (6–7), 1812–1824.
McCarthy, D.T., Deletic, A., Mitchell, V.G., Diaper, C., 2010. Sensitivity analysis of an urban stormwater microorganism model. Water Science and Technology 62 (6), 1393–1400.
McCarthy, P.J., 1976. The use of balanced half-sample replication in cross-validation studies. Journal of American Statistical Association 71, 596–604.
Montanari, A., 2007. What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. Hydrological Processes 21, 841–845.
Mourad, M., Bertrand-Krajewski, J.L., Chebbo, G., 2005. Stormwater quality models: sensitivity to calibration data. Water Science and Technology 52 (5), 61–68.
Rauch, W., Kinzel, H., 2007. KAREN – User Manual, Hydro-IT GmbH, Innsbruck (in German).
Rauch, W., Thurner, N., Harremoes, P., 1998. Required accuracy of rainfall data for integrated urban drainage modeling. Water Science and Technology 37 (11), 81–89.
Reichl, J.P.C., Chiew, F.H.S., Western, A.W., 2006. Model averaging, equifinality and uncertainty estimation in the modelling of ungauged catchments. In: iEMSs Third Biennial Meeting, Summit on Environmental Modelling and Software, International Environmental Modelling and Software Society, Burling.
Rudemo, M., 1982. Empirical choice of histograms and Kernel density estimators. Scandinavian Journal of Statistics 9, 65–78.
Schellart, A., Tait, S.J., Ashley, R.M., 2010. Towards the quantification of uncertainty in predicting water quality failures in integrated catchment model studies. Water Research 44, 3893–3904.
Silberstein, R.P., 2006. Hydrological models are so good, do we still need data? Environmental Modeling & Software 21 (9), 1340–1352.
Thorndahl, S., Schaarup-Jensen, K., Jensen, J.B., 2008. Probabilistic modelling of combined sewer overflow using the First Order Reliability Method. Water Science and Technology 57 (9), 1337–1344.
Uhl, M., Henrichs, M., 2010. KALIMOD: Calibration of hydrological process models, Manual Version 2.0. Muenster University of Applied Sciences, Laboratory of Water Resources Management, Germany (in German).
Vaze, J., Chiew, F.H.S., 2003. Comparative evaluation of urban storm water quality models. Water Resources Research 39 (10), 1280.
Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39 (8), 1201, doi:1210.1029/2002WR001642.
Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. In: Proceedings of the National Academy of Science of the United States of America (MEC).
Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44, W00B09.
Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment 23 (7), 1011–1026.
Wagener, T., Wheater, H.S., Gupta, H.V., 2004. Rainfall-Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London, UK.
Walker, W.E., Harremoës, P., Rotmans, J., Van, J.P., der Sluis, M.B.A., Van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P., 2003. Defining uncertainty–A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4 (1), 5–17.
Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology 358 (1–2), 1–23.

Appendix B

Glossary

B.1 Glossary

The following definitions are not new and are established in various guidelines (Taylor and Kuyatt, 1994; Kacker and Jones, 2003; Michelson et al., 2005; Bertrand-Krajewski and Muste, 2007; ISO, 2009a; b) and studies referenced thought the thesis.

Calibration is an iterative process of matching or minimizing the difference between observed values against simulated responses by means of an objective or a likelihood function.

Calibration algorithm is the method used to calibrate the model (i.e. to minimise the difference between measured calibration data and modelled values).

Calibration data is the measured input data required to calibrate the model (i.e. measured data to be compared to the modelled values).

Conceptual parameters are those quantities that are unobservable and can only be inferred through calibration.

Error is the difference between a true value and a modelled/observed value.

Input data is the measured input data required by a model.

Measurand is a particular quantity subject to measurement. A result of measurement is a value attributed to the measurand.

Model sensitivity is the sensitivity of the model outcomes to changes in the model parameters. It is also referred as parameter sensitivity.

Model structure refers to the formulation, assumptions and initial conditions if the model.

Model validation is the processes of assess how well (or not) the model can perform outside the calibration period.

Objective function is a function representing the errors between the measured calibration data and modelled values.

Physical parameters are parameters that can be estimated by measurements independently of observable catchment responses.

Random error is the result of a measurement minus the mean that would result from an infinite number of measurements of the same measure carried out under repeatability conditions (because only a finite number of measurements can be made, it is possible to determine only an estimate of random error).

Sensitivity analysis is the process of varying model calibration parameters within a reasonable interval and observing the relative change in model responses (the parameters that are most likely to significantly affect relevant outputs are determined).

Standard uncertainty is the uncertainty of the result of a measurement and is expressed by the standard deviation.

Systematic error is an error which results from some bias in the measurement process and is not due to chance, in contrast to random error.

Uncertainty characterizes the dispersion of the values within which the true value is believed to lie with a pre-established level of confidence (e.g. it can be expressed by a standard deviation, or a given multiple of it, or the width of a confidence interval).

Uncertainty analysis is the term used to describe the exercise of identifying the uncertainty in the model results.

B.2 References

Bertrand-Krajewski, J. L. and Muste, M. (2007). Chapter 6: Understanding and managing uncertainty. *in: Flectcher, T.D. and Deletić, A. (Eds.) Data Requirements for Integrated Urban Water Management.* 18 London, UK, Taylor and Francis, Urban Water series - UNESCO IHP**:** 65-90.

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

ISO (2009b). ISO/IEC Guide 98-3/S1/AC1:2009(E) Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM:1995), Supplement 1: Propagation of distributions using a Monte Carlo method, Technical corrigendum 1. May 2009, Geneva, Switzerland**:** 2.

Kacker, R. and Jones, A. (2003). On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent. *Metrologia* **40**: 235-248.

Michelson, D., Einfalt, T., Holleman, I., Gjertsen, U., Friedrich, K., Haase, G., Lindskog, M. and Jurczyk, A. (2005). Weather Radar Data Quality in Europe: Quality Control and Characterization. Review, COST Action 717 - Use of radar observations in hydrological and NWP models. Luxembourg.

Taylor, B. N. and Kuyatt, C. E. (1994). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. USA, NIST Technical Note 1297, National Institute of Standards and Technology.