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

A B S T R A C T

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.

Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.

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.
   E-mail address: ana.deletic@monash.edu (A. Deletic).

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).
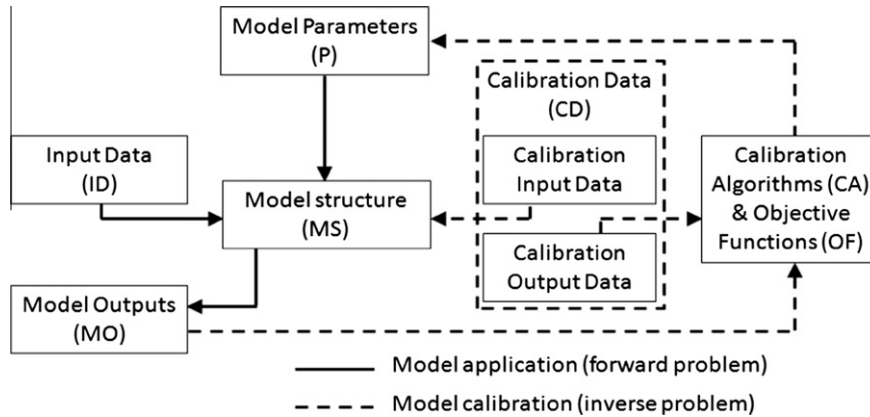
Fig. 1. General modelling framework.

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
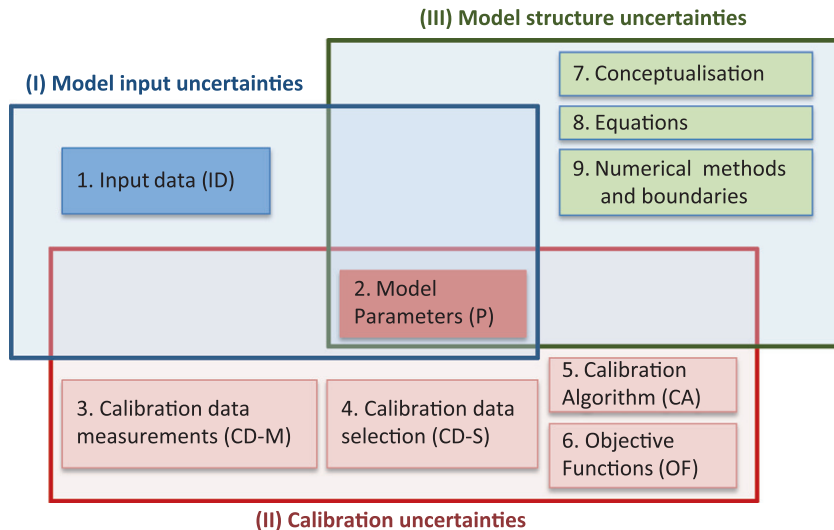

Fig. 2. The key sources of uncertainties in urban drainage models and links between them.

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

$$ID_{FACTOR} = f(\delta, \varepsilon) \tag{1}$$

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 apporprate 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.
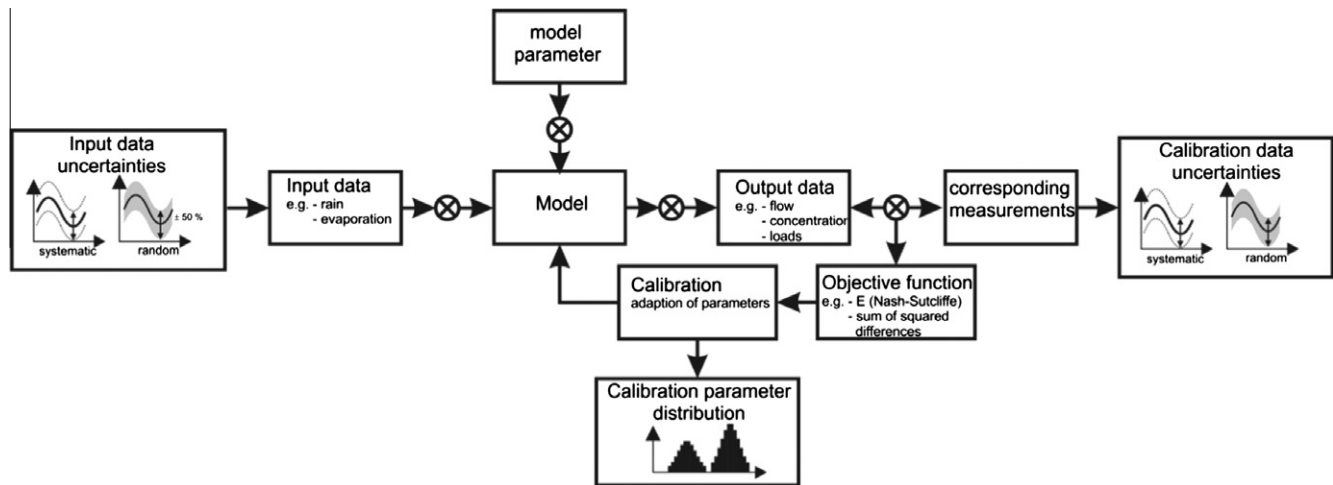
Fig. 3. A total error framework for urban drainage models.

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 modeling. 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.