



Jul 1st, 12:00 AM

## Assessment of data availability influence on integrated urban drainage modelling uncertainty

Gabriele Freni

G. Mannina

G. Viviani

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Freni, Gabriele; Mannina, G.; and Viviani, G., "Assessment of data availability influence on integrated urban drainage modelling uncertainty" (2008). *International Congress on Environmental Modelling and Software*. 91.

<https://scholarsarchive.byu.edu/iemssconference/2008/all/91>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# Assessment of data availability influence on integrated urban drainage modelling uncertainty

G. Freni<sup>a</sup>, G. Mannina<sup>a</sup>, G. Viviani<sup>a</sup>

<sup>a</sup> *Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali, Università di Palermo, Viale delle Scienze, 90146 Palermo, ITALY (freni/mannina/gviv@idra.unipa.it)*

**Abstract:** In urban water quality management, several models are connected and integrated for analysing the fate of pollutants from the sources on the urban catchment to the final recipient; classical problems connected with the selection and calibration of parameters are amplified by the complexity of the modelling approach increasing their uncertainty. The present paper aims to study the influence of dataset extension on the modelling response uncertainty with respect to the different integrated modelling outputs (both considering quantity and quality variables). At this scope, a parsimonious home-made integrated modelling approach has been used allowing for analysing the combinative effect between sewer system, treatment plant and receiving water body; the uncertainty analysis approach has been applied to an experimental catchment in Bologna (Italy). The number of available data points has been fictitiously reduced obtaining datasets with extension ranging between 25% and 100% of the measured data. For each of the datasets, uncertainty analysis has been performed and its propagation from upstream submodel to the downstream ones has been assessed. The results are interesting and show a strong influence of dataset extension on model uncertainty.

**Keywords:** Environmental modelling, Integrated urban drainage systems, Uncertainty analysis, Receiving water body, Waste water treatment plant.

## 1. INTRODUCTION

Nowadays, water quality improvement, mainly focused on the overall management of river basins, is gaining more importance during the last years [Blösch, 1999]. As a result of the publication of the European water framework directive (WFD), integrated management of river basins has become an important research area [Lindenschmidt, 2006]. The main elements of integrated urban water system are: catchment area and sewer system (SS); waste water treatment plant (WWTP); receiving water body (RWB). Dry weather flow along with surface runoff is transported from the catchment area to the WWTP for the removal of pollutants and subsequent release into the receiving water. Only if the amount of runoff exceeds the given hydraulic capacity of the plant, a mixture of wastewater and rainwater is discharged to the receiving water directly.

Indeed, due to the fact that integrated approaches are basically a cascade of sub-models (simulating SS, WWTP and RWB), uncertainty produced in one sub-model propagates to the following ones depending on the model structure, the estimation of parameters and the availability and uncertainty of measurements in the different parts of the system. For this reason, an important issue in integrated modelling is connected with the balance between sub-models so that each of them will not introduce too much uncertainty depending on the available field data. Furthermore, such problem, coupled with the consequent uncertainty

accumulation, may lead to get very uncertain and sometimes useless results [Willems, 2000].

Regarding field data for integrated modelling calibration/validation, a limited data base is generally available especially looking at the water quality one [Vanrolleghem et al., 1999]. Indeed, this consideration depends on the fact that contemporary monitoring campaigns on the different compartments of the integrated urban water system are complex and they require large technical and economic efforts. The situation is even worse when looking at wet weather impact on RWB because of the small scale at which polluting impacts take part adding more difficulties in providing reliable monitoring campaigns. As a matter of fact, several Authors highlight data availability as one of the main limitation to integrated modelling application and they adopt hypothetic or semi-hypothetic case studies where the different parts of the system are simulated using data coming from real and well documented case studies not really linked together [Schütze et al. 1999; Rauch and Harremoës, 1996].

The present paper aims to study the dataset extension influence on the modelling response uncertainty and, at this scope, a parsimonious home-made integrated model has been used allowing for analysing the combinative effect between SS, WWTP and RWB [Mannina et al., 2004, 2005]. In details, the uncertainty analysis has been applied to an experimental catchment in Bologna (Italy).

The dataset has been fictitiously reduced in the extension by uniformly decreasing the number of data points available for each of the monitored events as it will be better discussed in the following paragraphs.

## 2. MODELS AND METHODS

### 2.1 The adopted model

In order to carry out the survey, a home-made quality-quantity integrated urban drainage model has been employed. For sake of conciseness, in the following will be discussed only the model structure remanding to the literature for further details Mannina et al. [2004] and Mannina [2005]. The model is able to estimate both the interactions between the different systems (SS, WWTP and RWB) and the modifications, in terms of quality, that urban stormwater causes inside the RWB. Such a system is made up mainly of three sub-models:

- the rainfall-runoff and flow propagation sub-model, which is able to evaluate the quality - quantity features of SS outflows;
- the WWTP sub-model, which is representative of the treatment processes;
- the RWB sub-model that simulates the pollution transformations inside the river.

The first sub-model, reproducing the physical phenomena which take place both in the catchments and in the sewers, allows to determine the hydrograph and pollutograph in the sewer. This sub-model is divided into two connected parts: a hydrological - hydraulic module, which calculates the hydrographs at the inlet and at the outlet of the sewer system, and a water quality module, which calculates the pollutographs at the outlet for different pollutant species (TSS, BOD and COD). The hydrological - hydraulic module starts to evaluate the net rainfall, from the measured hyetograph, by a loss function (taking into account surface storage and soil infiltration). From the net rainfall, the model simulates the net rainfall-runoff transformation process and the flow propagation with a cascade of one linear reservoir and a linear channel (representing the catchment) and a linear reservoir (representing the sewer network). The solid transfer module reproduces the build-up and wash-off of pollutants from the catchment and the propagation of solids in the sewer network considering also their sedimentation and re-suspension.

The second sub-model is aimed to the analysis of WWTP during both dry and wet weather periods. The WWTP inflow has been computed taking into account the presence of a CSO device and its efficiency. The WWTP sub-model simulates the behaviour of the part of the

plant composed by an activated sludge tank and a secondary sedimentation tank. For the activated sludge tank model, mass balance equations derived from Monod's theory have been used in order to reproduce BOD removal. On the other hand, the sedimentation tank has been simulated using the solid flux theory according to the methodology proposed by Vitasovic et al. [1997]. In particular, the solids concentration profile has been obtained by dividing the settler into 50 horizontal layers of constant thickness. Within each layer the concentration is assumed to be constant and the dynamic update is performed by imposing a mass balance for each layer. Further, the sedimentation model proposed by Takács et al. [1991] has been employed in order to evaluate the different types of sedimentation.

The third sub-model examines the assessment of RWB. The simplified form of the De Saint Venant equation (cinematic wave) is used for the quantity module and water de-oxygenation and re-oxygenation phenomena are simulated by using the advection - dispersion equation; this approach allows for evaluating the effects of stormwater on the RWB, both at a single event scale and during long term simulation. The pollutant sources, which cause a worsening in the characteristics of RWB, are mainly two: the WWTP continuous discharges and the intermittent discharges coming from the combined storm overflows (CSOs). The sub-model also allows for simulating the effects of the insertion of CSOs control measures for temporary accumulation of stormwater during the rainfall event.

## **2.2 Uncertainty analysis based on Generalised Likelihood Uncertainty Estimation (GLUE) approach**

The GLUE methodology [Beven and Binley, 1992] is aimed to the identification of modelling uncertainty related to parameter estimation. In the present study, it will be used for assessing the effect of imperfect knowledge of the system on this specific kind of uncertainty. System knowledge will be fictitiously "altered" by reducing the number of available data points in the measurements dataset: in each system compartment (SS, CSO and RWB), data availability will be reduced to 25%, 50% and 75% respectively. Imperfect knowledge in each system compartment will be analysed separately in order to understand the weight to be given to measuring campaigns in the parts of the system. For applying the GLUE methodology, the model has been solicited by randomly sampled parameter sets throughout Monte Carlo simulations. By means of a likelihood measure,  $E$ , parameter sets are classified and sets with poor likelihood weights, with respect to a user-defined acceptability threshold ( $Tr$ ), are discarded as "non-behavioural". All parameters sets coming from the behavioural simulation runs are retained and their likelihood weights are re-scaled so that their cumulative total sum is equal to 1. The likelihood measure,  $E$ , represents the ability of the model to fit real data. On the other hand, the acceptability threshold,  $Tr$ , represents a user-defined critical value indicating the minimum value of  $E$  that each modelling simulation should have to be representative of the model behaviour with respect to the analysis aim.  $Tr$  is usually set equal to zero. In the present study, the Nash and Sutcliffe efficiency index has been used as likelihood measure [Nash and Sutcliffe, 1970]. Nash-Sutcliffe likelihood measure is analytically defined in the range  $]-\infty; 1]$  although behavioural simulations are defined in the range  $]Tr; 1]$ .

Treating the distribution of likelihood values as a probabilistic weighting function for the predicted variables, it is possible to assess the uncertainty associated with the predictions, conditioned on the definition of the likelihood function, of the input data and model structure. A method of deriving predictive uncertainty bands using the likelihood weights from the behavioural simulations has been shown by Beven and Binley [1992]. The uncertainty bands are calculated using the 5% and 95% percentiles of the predicted output likelihood weighted distribution. Wider bands mean higher uncertainty in the estimation of the modelling output and thus lower confidence in the model results.

Leaving details to literature [among others: Beven and Binley, 1992; Gupta et al., 2005; Freni et al., 2008a], GLUE procedure can be applied according to the following steps:

1. "Uncertain" parameters are selected: the adopted integrated model is controlled by 27 parameters. As pointed out in other studies, the model is not sensitive to some of them in the presented case study and, thereafter, nineteen parameters have been considered in

the present study [Freni et al., 2008b]. Their brief explanations are provided in previous studies [Mannina, 2005; Freni et al., 2008a, b].

2. variation ranges are selected for each model parameter which can be considered affected by uncertainty: the variation range for each of the selected parameters has been obtained by the calibration of the five fully monitored events [Freni et al., 2008b]; more specifically, initially wide ranges have been assigned for uncertain parameters based on both their physically feasible ranges and previous literature researches. Thereafter, five calibration processes has been carried out based on data coming from single fully monitored events. Parameters variation ranges were thus refined limiting them between the maximum and the minimum value obtained on the different calibration processes.
3. random sets are drawn from the parameter variation ranges by Monte Carlo sampling according to an *a priori* distributions: in this case, the uniform distribution has been assumed not having adequate prior knowledge for selecting another distribution; correlation between parameters, that may affect their mutual distribution, has been neglected because no evidence has been reported in the model calibration process showing the relevance of such aspect [Freni et al., 2008b].
4. the model is run for each random parameter set and likelihood measure (in this study, Nash-Sutcliffe efficiency index) is calculated for all analysed modelling variables (discharges and water quality variables at the SS ending pipe, at CSO outlet and at the RWB closing cross-section); if likelihood measure is positive, the model simulation is considered “behavioural” and saved for the following analysis steps;
5. step 2 is repeated until a user-specified number of behavioural simulations is reached (the present study is based on 1,000 behavioural simulations) linking the parameter set and the correspondent modelling outputs to a likelihood measure value;
6. analogously to the procedure used in statistics for obtaining cumulated probability from probability density, likelihood measures can be cumulated for each parameter and for each modelling output obtaining cumulated likelihood distributions;
7. for each modelling output, 5% and 95% percentiles cumulated likelihood distribution represent the uncertainty bands; for each model parameter, 5% and 95% percentiles represent the region in the parameter space where most likely the “real” parameter values should be (with a confidence level equal to 0.1);

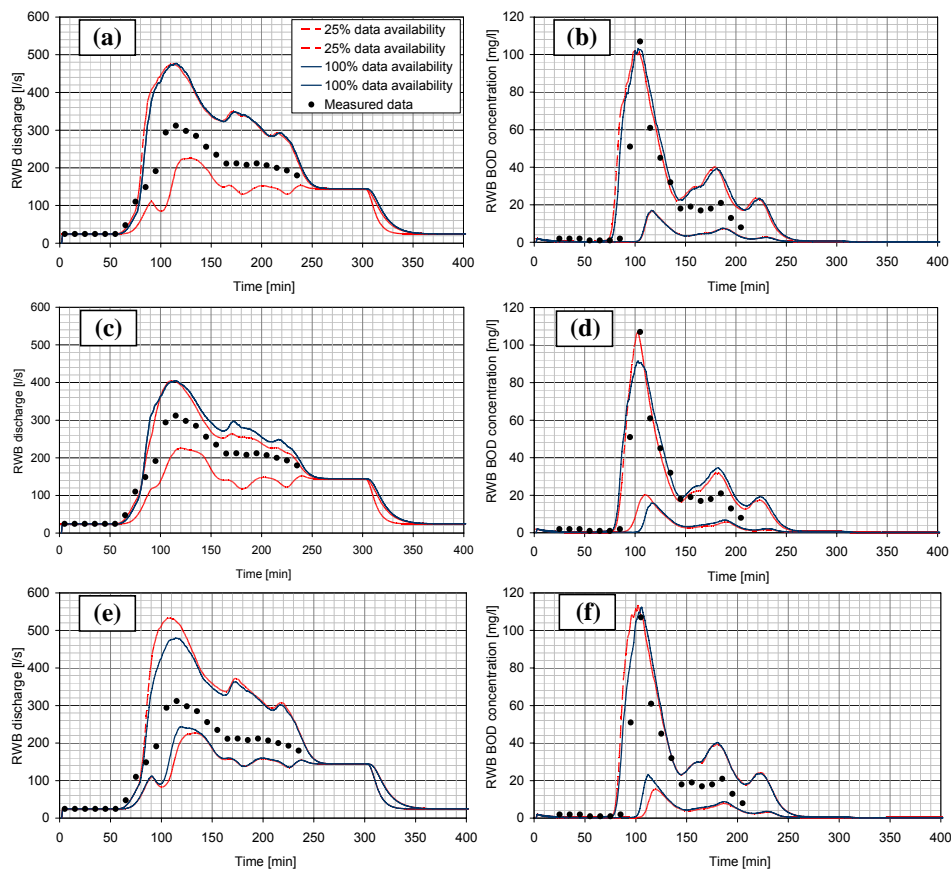
### 3. THE CASE STUDY

The integrated model and the uncertainty analysis have been applied to the catchment of the Savena river (Italy). The sewer system and the river studied in this work concern a part of the sewer network of Bologna, studied within the European Union research project INNOVATION 10340I [Artina et al., 1999]. The studied river reach is long about 6 km and it receives discharge from 6 CSOs deriving from the Bologna sewer network. The sewer network is a part of the combined system serving the whole city of Bologna, which can be considered as hydraulically divided into many independent catchments, all connected to a WWTP. The part of Bologna connected to the studied river has an area of more than 450 ha, with an impervious percentage of about 66% and about 60,000 inhabitants. During experimental survey, carried out within the INNOVATION European Research Project, from December 1997 to July 1999, about 50 events have been recorded, but, for only 5 of these, water quality aspects have been analysed regarding both RWB and SS. The monitoring infrastructure consisted of 3 raingauges, 8 sonic level gauges (6 in the drainage system and 2 in the receiving river) and 6 automatic 24-bottles sampler (3 in the sewer system and 3 in the river). The study has been focused on BOD<sub>5</sub>, TSS, COD and DO, even if analogous considerations may be extended to other parameters. In this study, only a part of the Savena river has been simulated (400 meters downstream the CSO No. 6) because the contribution of this CSO to river pollution has been determined much more relevant in respect to all the others. The contribution of other polluting sources has been considered by monitoring river pollution load in the first cross-section upstream of CSO

No. 6 and introducing this information as input in the models. Savena is an ephemeral river since there are wide flow variations during the different seasons and the river base flow is comparable with the CSO discharge. Further details on the monitoring campaign can be found in Artina et al. [1999].

#### 4. RESULTS ANALYSIS

The comparison among system knowledge scenarios will be provided according to three criteria: the effect on parametric uncertainty will be assessed by comparing uncertainty bands for different modelling outputs; the effect on single parameters weight in uncertainty propagation will be identified by analysing likelihood distributions; the effect on model calibration effectiveness will be derived from modelling efficiency distributions. Figure 1 shows uncertainty bands obtained the RWB downstream cross-section reducing data availability reducing data availability in different upstream sub-models: data availability is reduced in one sub-system per time so trying to separate the influence of each sub-system in the overall modelling uncertainty. Uncertainty bands width demonstrates that SS sub-model and RWB sub-model have higher importance in uncertainty propagation.

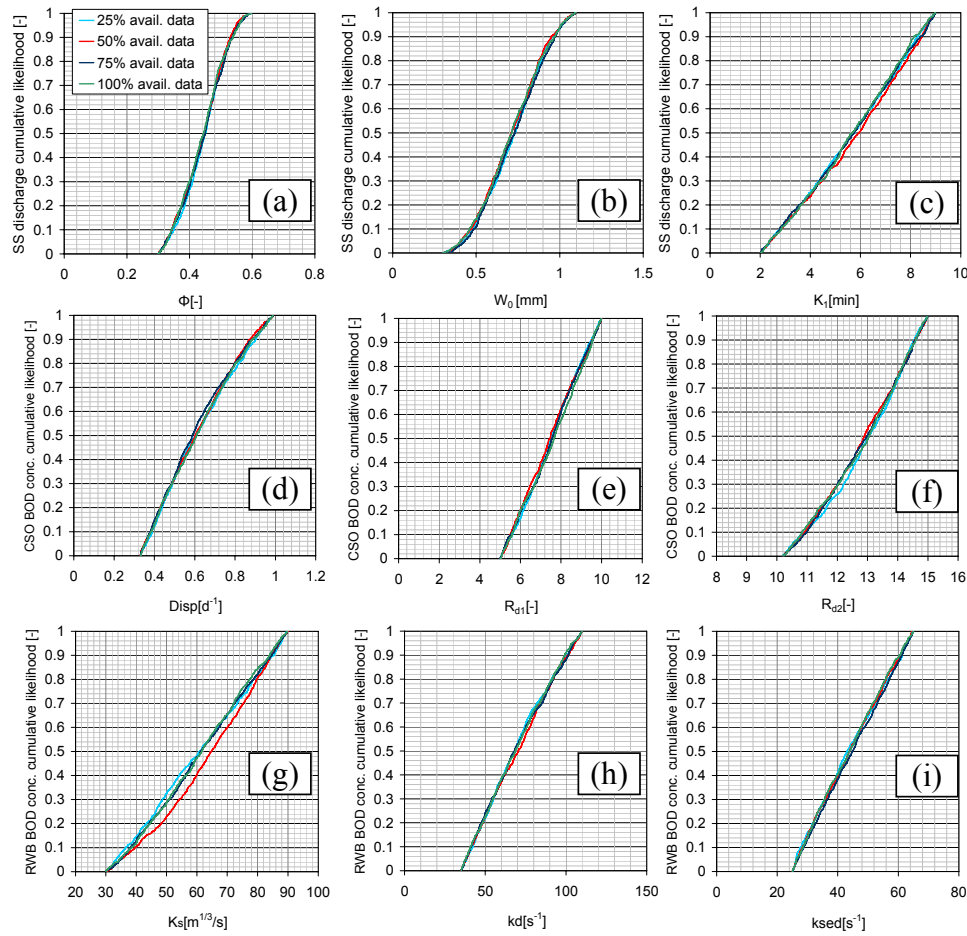


**Figure 1.** Uncertainty bands in the RWB modelling outputs obtained reducing data availability in SS sub-model (a-b), in CSO sub-model (c-d) and in RWB sub-model (e-f).

Uncertainty bands generated during CSO analysis (Figures 1c-1d) are generally smaller independently from data availability. Data availability in SS sub-model practically does not affect modelling output uncertainty (Figures 1a-1b); this behaviour can be explained by the uncertainty mitigation effect provided by the two downstream sub-models that can supply to the imperfect SS knowledge with their data. The effect of data availability is progressively higher moving to downstream sub-models (Figures 1c-1d and 1e-1f):

reducing data availability to 25% percent of measured data, uncertainty bands maximum width is increased by 30% for RWB discharge and 20% for BOD concentrations.

Data availability may affect the importance of single parameters in modelling uncertainty propagation: one parameter, that can greatly affect modelling efficiency when all measured data are available, can lose its importance when data availability is lower thus contributing to produce scattered modelling output and consequently increase uncertainty. The change in parameters behaviour can be analysed by the mean of likelihood distributions (Figure 2). The analysis of all parameters demonstrated that data availability does not affect parameters behaviour in terms of modelling uncertainty contribution. Figure 2 shows likelihood distributions for some of the modelling parameters adopted as examples. In particular, in Figure 2, the following model parameters are reported: the initial hydrological losses ( $W_0$ ), the catchment runoff coefficient ( $\Phi$ ), the catchment reservoir constant ( $K_1$ ), the decay rate in the Alley-Smith model (Disp), the CSO dilution factors ( $R_{d1}$ - $R_{d2}$ ), the river bed roughness (Gauckler-Strickler) ( $K_s$ ), the deoxygenation coefficient ( $K_d$ ) and the sediment oxygen demand ( $k_{sed}$ ).

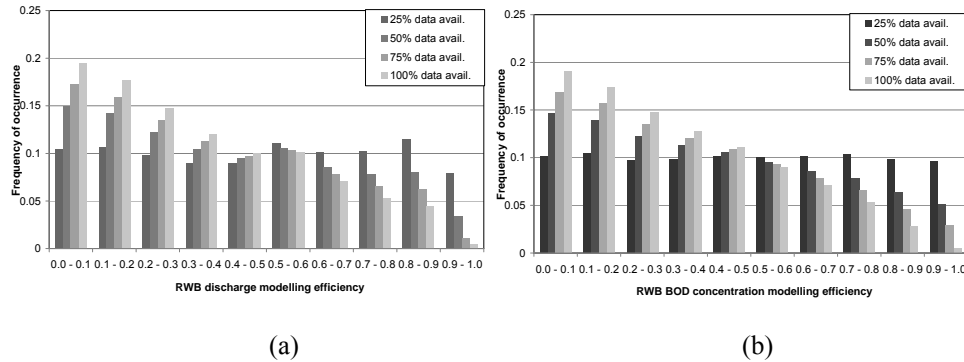


**Figure 2.** Cumulated likelihood distributions for some relevant parameters in SS sub-model (a-c), in CSO sub-model (d-f) and in RWB sub-model (g-i).

The distributions obtained for all data availability scenarios are substantially superimposed with some differences only in RWB sub-model that justify the global increase in modelling uncertainty presented in Figure 1.

Another important effect of data availability has been detected by Beven and Freer [2001]: the basic idea of GLUE is that given our inability to represent exactly in a mathematical model how nature works, there will always be several different models that mimic equally well an observed physical process (such as sewer discharge or pollutant concentration).

Such concept takes to model equifinality (more than one model structure can fit the monitored reality) and to model parameters identifiability problems (once fixed the modelling structure, several sets of parameters are able to give similar model outputs). Such behaviour can be amplified by lack of data for efficient characterization of over-parameterized models. The increase in model identifiability problems can be demonstrated by examining modelling efficiency frequency density plots (Figure 3).



**Figure 3.** Efficiency frequency density plots for the estimation of RWB discharge (a) and of RWB BOD concentration (b)

Such diagrams show the frequency that a model simulation efficiency can be included in a specific interval and it gives an idea of the shape of the modelling efficiency surface (i.e. the surface representing modelling efficiency as a function of parameters values in the parameters confidence region). If frequency is uniformly distributed in all intervals, the shape of efficiency surface is not steep and several parameters sets can be found obtaining a good modelling performance; if frequency is progressively reducing from lower efficiency intervals to higher ones, the efficiency surface is steep and the identification of a single calibration parameters set is easier. Figure 3 shows that data availability affects parameters identifiability problems of the adopted model especially for water quality aspects that are probably characterised by higher uncertainty.

## 5. CONCLUSIONS

The study analysed the effect of data availability on urban drainage integrated modelling reliability. Such effect has been analysed by the mean of GLUE uncertainty analysis approach. The study allowed for the following considerations:

- Modelling uncertainty is affected by data availability even if some differences can be highlighted among different sub-models; when RWB analysis is the aim of the study, imperfect knowledge in upstream sub – systems (SS and CSO) may be partially corrected by the model thanks to the higher amount of information available in the downstream one.
- The importance of single parameters in uncertainty propagation is not greatly affected by data availability so increasing modeller confidence in the fact that a first modelling analysis when low amount of data is available can help in focusing further monitoring campaigns towards the most sensitive parts of the system.
- Data availability greatly can greatly modelling parameter identifiability thus reducing modeller confidence in the calibration process; this aspect also affects the possibility that initial calibration (based on low amount of data) may be confuted by following data acquisition.

The study is based on a specific case study and model, thus requiring extended applications for generalising the study results. Nevertheless, the study suggests a method for evaluating the importance of data in integrated modelling identifying where larger efforts are needed with the aim of collecting more information.



## ACKNOWLEDGEMENTS

Authors gratefully acknowledge Prof. Sandro Artina and Dr. Marco Maglionico (D.I.S.T.A.R.T. – University of Bologna, IT) for providing data of Savena catchment.

## REFERENCES

- Artina S., Bardasi G., Borea F., Franco C., Maglionico M., Paoletti A., Sanfilippo U., Water quality modelling in ephemeral streams receiving urban overflows. The pilot study in Bologna, 8th International Conference Urban Storm Drainage, ICUSD, Sydney, 30 Agosto-3 Settembre 1999, Joliffe Ian B. & James E. Ball (eds.), Sydney, Australia, Vol. 3, 1589-1596, 1999.
- Beven K.J., Binley A.M., The future of distributed models - model calibration and uncertainty prediction, *Hydrological Processes*, , Vol. 6(3), pp. 279-298, 1992
- Beven K.J., Freer J. Equifinality, data assimilation and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, 249: 11-29, 2001
- Blöch, H., The European water framework directive: taking European water policy into the next millennium. *Water Science and Technology* 40 (10), 67-71, 1999.
- Freni, G., Mannina, G., Viviani, G., Uncertainty in urban stormwater quality modelling: The effect of acceptability threshold in the GLUE methodology, *Water Res.*, Volume 42, 8-9, 2061-2072, 2008a
- Freni, G., Maglionico, M., Mannina, G., Viviani, G., Comparison between a detailed and a simplified integrated model for the assessment of urban drainage environmental impact on an ephemeral river, *Urban Water J.* 5 (2), 21–31, 2008b.
- Gupta, H. V., Beven, K. J., Wagener, T., Model Calibration and Uncertainty Estimation.” *Encyclopedia of Hydrological Sciences*. Edited by M. Anderson. John Wiley & Sons, Ltd, 2005.
- Lindenschmidt K. E., The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling, *Ecological Modelling* 190, 72–8672-86, 2006
- Mannina, G., Integrated urban drainage modelling with uncertainty for stormwater pollution management, Ph.D. Thesis, Università di Catania, Italy, 2005.
- Mannina, G., Viviani, G., Freni, G., Modelling the integrated urban drainage systems, In: Bertrand-Krajewski, J.-L., Almeida, M., Matos, J., Abdul-Talib, S. (Eds.), *Sewer Networks and Processes within Urban Water Systems*. IWA Publishing, London, 3–12, 2004.
- Nash, J.E. and Sutcliffe, J.V., River flow forecasting through conceptual models. *J. of Hydr.*, 10(3), 282–290, 1970.
- Rauch W. and Harremoës P., The importance of the treatment plant performance during rain to acute water pollution. *Wat. Sci. Tech.*, 34 (3-4), 1-8, 1996.
- Schütze, M., Butler, D., and Beck, M. B., Optimisation of control strategies for the urban wastewater system. An integrated approach, *Water Science and Technology* ,39(9), 209-216, 1999.
- Takács, I., Patry, G.G., Nolasco, D. (1991). A Dynamic Model of the Clarification-Thickening Process. *Wat. Res.*, 25(10), 1263-1271.
- Vanrolleghem P., Schilling W., Rauch W., Krebs P., Aalderink H., Setting up measuring campaigns for integrated wastewater modelling” - *Wat. Sci. and Tech.*, vol. 39, n. 4, pp. 257-268, 1999.
- Vitasovic, Z. C., Zhou, S., McCorquodale, J. A., and Lingren, K., Secondary clarifier analysis using data from the Clarifier Research Technical Committee protocol. *Water Environment Research* 69, 999 – 1007, 1997.
- Willems, P.. Probabilistic immission modelling of receiving surface waters. Ph.D. Thesis, K.U. Leuven, 2000.