Uncertainty propagation throughout an integrated water-quality model

Gabriele Freni
G. Mannina
G. Viviani
Uncertainty propagation throughout an integrated water-quality model

G., Freni and G., Mannina, G., Viviana

Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali, Università di Palermo, Viale delle Scienze, 90146 Palermo, ITALY (mannina@idra.unipa.it)

Abstract: In integrated urban drainage water quality models, due to the fact that integrated approaches are basically a cascade of sub-models (simulating sewer system, wastewater treatment plant and receiving water body), 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. Uncertainty basically propagates throughout a chain of models in which simulation output from upstream models is transferred to the downstream ones as input. The overall uncertainty can differ from the simple sum of uncertainties generated in each sub-model, depending on well-known uncertainty accumulation problems. The present paper aims to study the uncertainty propagation throughout an integrated urban water-quality model. At this scope, a parsimonious bespoke integrated model has been used allowing for analysing the combinative effect between different sub-models. Particularly, the different parts of the quantifiable uncertainty have been assessed and compared by means of the variance decomposition concept. The integrated model and the methodology for the uncertainty decomposition have been applied to a complex integrated catchment: the Nocella basin (Italy). The results show that uncertainty contribution, due to the model structure, is higher with respect to the other sources of uncertainty.

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

1. INTRODUCTION

Since the Water Framework Directive (WFD) enactment, several researchers have been working to develop integrated approaches for water basin management (among others, Rauch et al., 2002; Mannina et al., 2006; Willems, 2008). Indeed, WFD implicitly requires a holistic approach of the whole system, namely: sewer system (SS), wastewater treatment plant (WWTP) and receiving water body (RWB), in order to properly design as well as manage water resources and protect the environment from pollution. A modelling framework of the overall integrated system may allows all domains of the catchment to be modelled and is essential for describing water states in both the temporal and spatial dimensions. An integrated approach generally requires the management of a complex model where several sub-models are linked together and processes and interactions within and between them take place. However, high complexity in general should be avoided and it is not desirable (Chapra, 2003; Rechow, 2003; Beven, 2006; Willems, 2008). Complex models usually require large databases for calibration and they can deliver relevant amounts of uncertainty to the modelling outputs; the study of the interconnection between model complexity and data availability is of paramount interest. In order to reduce model complexity, the system is generally separated into small systems. As a matter of fact, Schmitt and Huber (2006) suggest to divide the whole system into smaller subsystems and to define appropriate system boundaries. With this respect uncertainty assessment may constitute an optimal solution. As a matter of fact, uncertainty quantification of a mathematical model enables one to gain insights about the significance level of the results.
provided by the model. Further, quantitative uncertainty analysis can provide an illuminating role for problems where data are limited and where simplifying assumptions have been used in order to help identify the robustness of the conclusions and to help target data-gathering efforts (Frey, 1992). In the last years, a large debate rose in literature about the most appropriate methodologies for analysing model uncertainty and about the compromise between initial hypotheses of methods and, on the other hand, reliability of uncertainty analysis results. An example is given by the debate about the use of “formal” Bayesian methods and “non-formal” ones like Generalised Likelihood Uncertainty Estimation: the first providing more rigorous and reliable framework but requiring the definition of a formal error model; the latter generally computing larger uncertainties but relaxing many of the Bayesian hypotheses (Beven and Bynely, 1992; Stedinger et al., 2008; Vrougt et al. 2009).

In urban drainage, especially for water quality modelling, the state-of-the-art is far less advanced compared to other fields. Deletic et al. (2009) pointed out that the state of knowledge regarding uncertainties in urban drainage models is poor, in part due to a lack of clarity and/or consensus on the way in which the results of model uncertainty analyses are obtained, presented and used. Among the possible reasons for the current lack of quantification of model uncertainty in the urban drainage modelling field, the high computational effort (e.g., for a Monte Carlo simulation) required by mathematical models has been cited as the main impediment (Muschalla et al., 2009). Indeed, urban integrated models are basically a cascade of sub-models (simulating SS, WWTP and RWB), and the computational time required for the entire model may constitute a limitation, especially in cases where several Monte Carlo simulations are run (Mannina and Viviani, 2010).

In this context the paper presents the uncertainty assessment and the uncertainty propagation throughout a bespoke integrated model developed in previous studies (Mannina, 2005). Particularly, the variance decomposition concept has been applied for the uncertainty quantification as well as propagation. The methodology and the model has been applied to a real case study, the Nocella catchment (IT), for which an extensive field gathering campaign was carried out and quantity/quality data were therefore available.

2. MATERIALS AND METHODS

2.1 Integrated urban drainage model

For the simulation of the whole system, a previously developed, bespoke model was adopted (Mannina, 2005). The model is able to estimate both the interactions between the three components of the system (SS, WWTP and RWB) and the modifications, in terms of quality, that urban stormwater causes inside the RWB. The general structure of the integrated model consists of three sub-models; each sub-model is divided into quantity and quality modules for the simulations of the hydrographs and of the pollutographs, respectively. The modelling structure can be adapted to the specific application by removing or duplicating sub-models or parts of them, such as the stormwater tank or the Combined Sewer Overflow (CSO). The quantity module of the SS sub-model is described by a cascade of a linear reservoir and a channel, representing the catchment, and a linear reservoir, representing the sewer network. Initially, the net rainfall is computed by subtracting both continuous and initial losses; the latter are modelled assuming constant initial depression storage and a constant runoff coefficient. For the quality module of the SS sub-model, several processes were considered, both on the catchment and in the sewer, as well as during dry and wet weather. In particular, the build-up and wash-off processes for pollutants were considered according to the classical approaches proposed by Alley and Smith (1981) and Jewell and Adrian (1978). Solids deposition in the sewer during dry weather was evaluated by adopting an exponential law depending basically on the duration of the antecedent dry weather and on sewer network characteristics (Bertrand-Krajewski, 1992). Regarding the erosion of sewer sediments (Mannina and Viviani, 2010 Crabtree, 1989), their cohesive behaviour was considered by assuming the bed sediment structures hypothesised by Skipworth et al. (1999). The pollutographs at the outlet of the sewer system were evaluated by hypothesising the complex catchment sewer network as a reservoir and by considering the transport capacity of the flow. Finally, the WWTP inflow
was computed by taking into account the presence of a CSO device, representing its efficiency by the introduction of two dilution coefficients.

The WWTP sub-model simulates the most sensitive units that can be affected by an increase of pollutant load inflow; more specifically, the activated sludge tank and the settler. In particular, the flow substrate and microbial density in the activated sludge tank were calculated with mass balances based on Monod's theory. Conversely, the sedimentation tank performance was simulated using the solid-flux theory according to the methodology proposed by Takács et al. (1991). In particular, the solids concentration profile was obtained by dividing the settler into horizontal layers of constant thickness. Within each layer, the concentration was assumed to be constant, and the dynamic update was performed by imposing a mass balance for each layer. The settling velocity function proposed by Takács et al. (1991) was employed. Regarding the RWB sub-model, the exemplified form of the Saint-Venant equation (kinematic wave) for the quantity module and the dispersion advection equation for the quality module were adopted (Brown and Barnwell, 1987).

2.2 The Case study

The analysis was applied to a complex integrated catchment: the Nocella catchment, which is a semi-urbanised catchment located nearby Palermo in the northwestern part of Sicily (Italy). The entire natural basin is characterised by a surface area of 99.7 km² and has two main branches that flow primarily east to west.

The two main branches join together at 3 km upstream from the river estuary. The southern branch is characterised by a smaller elongated basin and receives water from a large urban area characterised by relevant industrial activities partially served by a WWTP and partially connected directly to the RWB. The northern branch was monitored in the present study. The basin closure is located 9 km upstream of the river mouth; the catchment area is 66.6 km². The cross-section closing the catchment is equipped with a hydro-meteorological station (Nocella a Zucco).

The river reach receives wastewater and stormwater from two urban areas (Montelepre, with a catchment surface area equal to 70 ha, and Giardinello, with a surface area of 45 ha) drained by combined sewers. Both urban areas are characterised by concrete sewer pipes with steep slopes. The Montelepre sewer is characterised by circular and oval pipes with maximum dimensions of 100 cm × 150 cm. The sewer system serves 7,000 inhabitants, and it is characterised by an average dry weather flow of 12.5 L/s (the water supply is 195 L/capita/d), and an average dry weather BOD concentration of 223 mg/L. The Giardinello sewer is characterised by circular pipes with a maximum diameter of 800 mm. The served population is 2,000 inhabitants, and it has an average dry weather flow of 2.5 L/s (here, the water supply is 135 L/capita/d) and an average dry weather BOD concentration of 420 mg/L. The calculated BOD unit loading factors for the two urban catchments are 35 and 45 g/capita/d for Montelepre and Giardinello, respectively. These values are lower than those typically observed in Italy (60 g/capita/d), likely due to the industrial activities present in the urban catchments; the lower concentration of BOD in Montelepre’s urban catchment is also due to the presence of an infiltration flow into the sewer system.

Each sewer system is connected to a WWTP protected by CSO devices. The WWTPs are characterised by simplified, activated sludge processes with preliminary mechanical treatment units, an activated sludge tank and a final circular settler. According to the modelling scheme, particular attention in data acquisition was given to the activated sludge tank and the sedimentation tank. Moreover, such units are the most sensitive to flow and concentration variations during wet weather periods; wet weather loads greatly affect activated sludge settling tanks and can significantly affect the effluent quality.

Rainfall was monitored by four rain gauges distributed over the basin: the Montelepre rain gauge is operated by Palermo University and is characterised by a 0.1-mm tipping bucket and a temporal resolution of 1 minute; the other three rain gauges are operated by the Regional Hydrological Service and they are characterised by a 0.2-mm tipping bucket and a temporal resolution of 15 minutes. The hydro-meteorological station (Nocella a Zucco) located at the catchment end is characterised by an ultrasonic level gage operated by the
Regional Hydrological Service and has a temporal resolution of 15 minutes. Rainfall data for yearly maximum intensity events are available for all the rain gauges from 1955 to the present without a gap. The instruments were integrated by Palermo University by installing an area–velocity submerged probe that provides water level and velocity data with a 1-minute temporal resolution. An ultrasonic external probe was used to obtain a second water level measurement for validation and as a backup in case the submerged probe failed; an automatic 24-bottle water quality sampler was used for water quality data collection. The water quality parameters monitored were Total Suspended Solids (TSS), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Ammonia-Nitrogen (NH₄-N), Total Kjeldahl Nitrogen (TKN) and Phosphorus (P); the Dissolved Oxygen (DO) was only monitored for the river. All analyses were carried out according to Standard Methods (APHA, 1995). The monitoring campaign was used for model calibration under the present conditions; details of the calibration process can be found in Freni et al. (2010). The model parameters were calibrated for each of the sub-models by means of Monte Carlo Analysis, randomly varying parameters in user-defined ranges and minimising the variance of the model output errors based on the available water quantity and water quality measurements (Freni et al., 2009).

2.3 The variance decomposition concept and quantification of different uncertainty sources

The uncertainties for each sub-model can be decomposed into model input and model-related uncertainties. Model input uncertainties are due to errors in the data used as boundary and initial conditions in the model. Model uncertainties are due to the structure of the model, which includes the equations and algorithms used for the simulations and the coupling of the models and the parameters used to control the equations. As pointed out by Willems (2008), model structure uncertainties can be seen as the remaining uncertainties in the model output after use of error-free input in the model and after the most optimal calibration of the model parameters to the available measurements (e.g., for a given model structure, by optimising the selected goodness-of-fit statistics). Usually, when the comparison of different model structures is not within the scope of the study, model structure and model parameter uncertainties are jointly analysed. In such cases, parameters are assumed to be the only source of uncertainty and structural uncertainty is implicitly distributed among the parameters (Freni et al., 2010-2009). Such a hypothesis was maintained in the present study, which was based on the analysis of uncertainties related to input and calibration data and model parameters.

The variance of the total uncertainty in the model output variables was calculated as the variance of the errors in the model results after comparison with observational data; the variance of the observational errors was subtracted from this variance. The variance of the other model structure-related uncertainties was then quantified as the variance due to the variation of only the model parameters (a lumped approach) and also as a rest term in the description of the total variance, making use of the concept of variance decomposition (a distributed approach):

\[
\sigma^2_{\text{tot,Y}} = \sum_{i=1}^{n} \sigma^2_{\text{inp,Y}} + \sigma^2_{\text{str,Y}}
\]

where \(\sigma^2_{\text{inp,Y}}\) is the variance of the total uncertainty in the model output variable Y after subtracting the variance of the observational errors, \(\sigma^2_{\text{str,Y}}\) is the source variance of the uncertainty contribution by model input variable \(X_i\) (\(i=1, \ldots, n\); \(n\) is the total number of input variables) and \(\sigma^2_{\text{str,Y}}\) is the source variance of the contribution of the model structure-related uncertainty.

Once quantified the error for each submodels, these are thereafter propagated throughout the different submodels that constitute the integrated model. To accomplish such a goal, different techniques may be employed. Among such techniques, the Monte Carlo method may be employed.

Similarly to Willems (2008), a Box–Cox (BC) transformation (Box and Cox, 1964) was applied to the all sub-model outputs Y for which a total variance was calculated:

\[
\text{BC}(Y) = (Y^{\lambda} - 1) / \lambda
\]
where the parameter $\lambda$ (0 $<\lambda$ $\leq 1$) is calibrated to reach homoscedasticity in the errors (variance of errors nearly independent on the model output magnitude). The parameter was calibrated in this study by trial-and-error after a visual inspection of the homoscedasticity of the model errors. The BC transformation was applied to the model input and the output variables before calculating the variances $\sigma^2_{\text{err}(X_i)}$ and $\sigma^2_{\text{var}}$ of Eq. (1).

After this transformation, standard deviations or confidence interval widths were obtained that were nearly uniform for all time steps. The overall uncertainty was then represented by the average of the variances for all time steps (after BC transformation). Using this methodology and the above-mentioned measurements, total uncertainty was quantified for the output variables for the different sub-models.

Model input uncertainties were quantified based on detailed investigations of the input data. For rainfall, which is the main model input and the driving force of the temporal variability of the system processes, uncertainties in the calibration curves for the rain gauges were investigated at the Hydraulics Laboratory of Palermo, and the following error formulation was found:

$$H = H_{\text{real}} \cdot (1 + \text{err} \cdot H_{\text{real}}^a)$$  \hspace{1cm} (3)

where $H$ is the rainfall depth taking into account the error, $\text{err}$ is the error randomly generated considering a normal distribution with mean equal to zero and standard deviation equal to 0.035, $H_{\text{real}}$ is the real rainfall depth (unknown) and $a$ is a shape coefficient equal to 0.2332. The error curve parameters were obtained by calibration over several rain gauges similar to those installed in the analysed case study.

In order to conduct the random error simulation, the time series (the input series in this case) were separated into “independent” storm events, i.e., events leading to separate or independent sewer runoff events (events were separated by a dry weather flow period equal to or larger than the concentration time of the sewer network).

Errors in calibration data were analysed by means of a normal statistical distribution with a null mean. The uncertainty in this flow monitoring was assessed after testing the depth–velocity devices in the Hydraulics Laboratory of Palermo. The standard deviation of the flow per monitor ranged from 20% of the flow value for water levels lower than 5 cm to 6% for water levels between 5 cm and 50 cm. The standard deviation of the water quality measurement errors was assessed based on values found in the literature (Ahyerre et al., 1998; Bertrand-Krajewski et al., 2001; Kanso et al., 2003; Willems, 2008) to be as high as 30–40% for BOD and 15–20% for the other variables considered.

In the present study, structural uncertainties for each sub-model were related to parameters, assigning to each an uncertainty quota connected to model algorithms and equations. Parameter error distributions were assumed uniform, and thus the parameters had the same probability of taking any value within a specified range. The ranges were determined in a previous study by means of a model calibration based on several monitored events (Freni et al., 2009). These values were also employed for the variation of the model parameters in the simultaneous assessment of the structural uncertainty by means of the Monte Carlo runs. The propagation of the random input errors to the model output variables was performed by Monte Carlo simulation. Random simulations (1,000 runs) were carried out with the stochastic model input error, and the propagated errors on the model output variables were calculated for each time step. With this procedure, distributions of random errors were obtained, reflecting the uncertainty in the model output variables caused by the total model input uncertainty. These distributions and corresponding error variances or confidence intervals could be obtained for each time step or averaged over all time steps in the simulation period to obtain a “mean overall uncertainty” estimate.

A similar procedure was followed for the random simulation and propagation of the other uncertainty sources considered. Uncertainty sources were either analysed separately to obtain the partial contributions of each uncertainty source or jointly to obtain the total uncertainty considered for a specific modelling output.

Whenever the different types of model uncertainties are assessed separately it is possible to compare and quantify the contributions of the different sources of uncertainty to the total uncertainty in the model output (Willems, 2008). More specifically, it is of paramount
interest the comparison of the uncertainties resulting from the data (model inputs and parameter calibration) and the uncertainties resulting from the model structure.

It is advisable to have a balance between data and structure uncertainty. Indeed, whenever the data uncertainty is higher it is recommended to provide more attention on data collection than to research in an attempt to improve the model results. By comparing the contributions of the different uncertainty sources, efficient measures thus can be determined to reduce the total uncertainty in the model results (Willems, 2008).

3. ANALYSIS OF RESULTS

The variance decomposition outlined above was applied to the bespoke integrated urban drainage model for RWB quality modelling of a case study in Italy. The analysis was performed as a step-by-step process starting from the most upstream water quantity sub-model and then propagating the uncertainties to the downstream ones. Initially, the homoscedasticity was verified and corrected by B-C transformation. For each modelling output for which measures were available, the total variance of the errors was computed by jointly accounting for all sources of error defined in the previous paragraph (input, calibration data and model structure, which was limited to model parameters). Partial contributions were singled out by analysing one uncertainty source at a time; their sum, according to the variance decomposition equation, gave the total variance of errors for the analysed modelling output. Any differences between the total variances computed by the lumped and the distributed approaches may have been due to the presence of a correlation between uncertainty sources. Such a correlation was surely absent if measurement and model errors were considered, but could have been present if uncertainties in different sub-models were considered. In the present application, the analysis has been focused on RWB discharge and quality state. The uncertainty in the estimation of these variables has been used as the object of the study.

In Figures 1-3 the results in terms of uncertainty bounds for the different sources of uncertainty are reported. Particularly, Figure 1 shows the uncertainty bounds in the river cross-section taking into account only the rainfall monitoring uncertainty. Such uncertainty was taken into account by applying the error model presented in the previous paragraph and calibrated in Laboratory to measured rainfall. Conversely, Figure 2 shows the uncertainty bounds considering a variation of the model parameters. By comparing the two uncertainty bounds it emerges that the model parameter uncertainty bounds are wider both for the quantity (Figure 2a) and for the quality (Figures 2b-2c). Therefore, for the present application, the uncertainty due to the rainfall could be negligible.

Figure 1. Confidence limits due to rainfall uncertainty.

Figure 2. Confidence limits due to model parameter uncertainty.
As a confirmation of the previous statement, Figure 3 reports the uncertainty bounds for the case of both model parameter uncertainty and rainfall uncertainty. Although the uncertainty bounds are wider with respect to the ones of the model parameter uncertainty, the differences are not relevant. The uncertainty bounds are in general wider for the quantity rather than for the quality aspects confirming the higher uncertainty that relies on quality aspects. Thus the results suggest to provide much more efforts both on the modelling as well as gathering data for the quality processes.

![Figure 3. Confidence limits due to both model parameter uncertainty and rainfall uncertainty](image)

### 4. CONCLUSIONS

The uncertainty assessment and its propagation throughout an integrated bespoke urban drainage model has been performed. Model structure, input and parameter uncertainty has been assessed and compared at the final cross section of the modelled system: the RWB outlet. Particularly, the analysis was carried out analyzing both quantity and quality phenomena: discharges and BOD and DO concentrations, respectively. The study enabled to draw some interesting considerations:

- Model results revealed that when analysing water quality variables, water quantity sub-models always provide smaller contributions to uncertainty than do water quality ones; this may be due to the higher complexity of water quality processes and their related models.

- The uncertainty contribution of water quantity modules to water quality ones is not negligible, and specific efforts should be provided by the modeller in order to adopt robust water quantity models, as their contribution to uncertainty affects both water quantity and water quality variables.

- The comparison of total variance computation by means of a lumped approach and by variance decomposition demonstrated the relevance of modelling error correlation when moving from upstream to downstream sub-models.

### REFERENCES


