Urban drainage integrated modelling: uncertainty propagation and parameters identifiability

Modélisation intégrée des réseaux d’assainissement pluvial urbain : propagation des incertitudes et possibilité d’identification des paramètres

Freni Gabriele, Mannina Giorgio, Viviani Gaspare
Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali
Università degli Studi di Palermo
Viale delle Scienze, 90128 Palermo, ITALY
freni@idra.unipa.it, mannina@idra.unipa.it, gviv@idra.unipa.it

RESUME
Dans le drainage urbain, les nouvelles possibilités informatiques ont également soutenu le développement de nouvelles approches intégrées ont visé la quantité de l'eau et l'analyse communes de qualité de la canalisation urbaine de totalité. Bien que l'avantage d'une approche intégrée ait été largement démontré, jusqu'au jour plusieurs aspects empêchent son applicabilité telle que la disponibilité rare des données de champ si comparé à la complexité modèle. Ces aspects empêchent parfois l'évaluation correcte des paramètres menant de ce fait à la grande incertitude en modelant la réponse. C'est un problème typique d'identifiabilité de paramètre qui sera discuté dans le papier actuel évaluant l'effet du procédé d'identifiabilité dans la confiance croissante d'opérateur en modelant des résultats. La méthodologie d'étude a été appliquée à un modèle urbain intégré fait maison de drainage qui a été calibré/validé considérant des données de champ rassemblées en captation expérimentale de Savena (Bologna - Italie).

ABSTRACT
In urban drainage, new computational possibilities also supported the development of new integrated approaches aimed the joint water quantity and quality analysis of the whole urban drainage system. Although the benefit of an integrated approach has been widely demonstrated, up to day several aspects prevent its applicability such as scarce availability of field data if compared with model complexity. These aspects prevent sometimes the correct estimation of parameters thus leading to large uncertainty in modelling response. This is a typical parameter identifiability problem that will be discussed in the present paper evaluating the effect of identifiability procedure in increasing operator confidence in modelling results. The presented methodology has been applied to a home-made integrated urban drainage model that has been calibrated/validated considering field data collected in Savena experimental catchment (Bologna – Italy).

KEYWORDS
Urban drainage integrated modelling, uncertainty analysis, identificability analysys
1 INTRODUCTION

In urban drainage, the integrated approaches aim at jointly analyse water quantity and quality characteristics of the whole urban drainage system, i.e. sewer system (SS), wastewater treatment plant (WWTP) and receiving water body (RWB). However, although the benefit of an integrated approach has been widely demonstrated, up to day several aspects prevent its applicability (Rauch and Harremoës, 1996) such as scarce availability of field data. These factors contiguously with the high complexity level of the adopted approaches introduce uncertainties in the modelling process that, as a matter of the fact, can be not clearly identifiable and assessable.

Three main uncertainty sources are generally classified: uncertainty of the model input variables, uncertainty of the model parameters values and uncertainty originating from the imperfect description of the physical reality by a limited number of mathematical relations (Harremoës, 1988; Willems, 2000). When dealing with complex modelling approaches, such as integrated approaches, in a context with insufficient field data, classical calibration approaches may lead to several equally consistent parameters sets and it is difficult to have sufficient confidence about the obtained results (Sorooshian and Gupta, 1983).

These considerations lead to the equifinality concept described by Beven and Binley (1992) accepting the fact that more parameter sets may exist able to provide a good fit between simulated and measured data. The non-uniqueness of parameter calibration set and the uncertainty connected to their estimation take the primary consequence that, for a given model structure and a given experimental layout, some modelling parameters cannot be reliably calibrated because the available information is not adequate to identify their specific effect on modelling output. This is a common parameter identifiability problem and its resolution can provide some interesting advantages for an adopted model such as:

- the definition of a subset of identifiable parameters that can be effectively calibrated with the available data;
- the identification of reducible (connected to identifiable parameters) and irreducible (connected with non identifiable parameters) uncertainty; the first can be reduced by the elongation of monitoring campaigns while the latter would require the monitoring of new variables that are not presently identifiable;
- the definition of "necessary and sufficient" modelling complexity level by reducing the number of non identifiable parameter maintaining an adequate model adaptation to the monitored variables thus increasing the operator confidence in modelling results.

The main objective of the present study is to apply parameters identification analyses to pin down the most sensitive model parameters in order to provide information on their calibration as well as the most efficient modelling approach complexity to be used. This approach will also allow identifying irreducible and reducible uncertainty.

2 MODELLING PARAMETERS IDENTIFICATION: APPROACHES AND METHODS

Most of the techniques designed to find practically identifiable subsets of model parameters are based on an investigation of sensitivity functions. The present study will concentrate on numeric criteria for the evaluation of mutual dependence of sensitivity functions (Belsley, 1991; Brun et al., 2001) and on techniques based on correlation studies (Weijers and Vanrolleghem, 1997). In this section, we briefly
review and compare some of the recently proposed techniques for identifying parameter subset that can be efficiently calibrated for large environmental simulation models. Let assume that a deterministic model can be described by a general set of equations: \( y = f(\theta) \) where the vector \( y = (y_1, y_2, ..., y_n) \) represents the \( n \) modelling output variables corresponding to the available measures \( y^* = (y_1^*, y_2^*, ..., y_n^*) \) and the vector \( \theta = (\theta_1, \theta_2, ..., \theta_m) \) represents the \( m \) independent parameters of the model. The modelling equations should be integrated with observation equations linking and comparing modelling outputs \( y \) and observations \( y^* \) but this aspect is not relevant for the present discussion and thus omitted. Independently from the nature of modelling equations, sensitivity functions can be defined stating the relevance of dependency between modelling outputs \( y \) and parameters \( \theta \):

\[
\sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\partial y_i}{\partial \theta_j} \Delta \theta_j = s_{ij},
\]

where \( s_{ij} \) is the variability range of \( j^{th} \) parameter depending on prior knowledge and \( y_{s_i} \) is a reference (or scaling) value for variable \( y_i \) used for preserving the non-dimensional nature of the sensitivity function. The function \( s_{ij} \) is useful because it gives information of the raw dependency of the modelling output from the parameters; the function \( s_{ij} \) states the relative impact of different parameters on modelling outputs. Parameters \( \Delta \theta_j \) and \( y_{s_i} \) can have a great influence on the sensitivity analysis results being the magnitude and the scaling parameter of the sensitivity function (Reichert and Vanrolleghem, 2001). In the present paper, \( y_{s_i} \) has been defined as the average measured value of the \( i^{th} \) variable and \( \Delta \theta_j \) has been taken by literature variability of \( j^{th} \) parameter. Scarcely identifiable model parameters can generate small sensitivity function values or an approximate linear dependence of sensitivity functions with respect to the parameters. In the first case, the parameter does not greatly affect the modelling output; in the latter case, the parameter variability does not clearly affect the modelling output and it can be perceived as a sort of underlying noise increasing the uncertainty transferred to the output variable without providing information on parameter calibration. Two techniques will be discussed in the present paper based on the previous considerations. The first technique was proposed and illustrated by Brun et al. (2001). The second technique was proposed and applied to the WWTP models by Weijers and Vanrolleghem (1997). The first technique is based on two steps in order to analyse the two possible causes of non identifiability previously discussed. First, a sensitivity ranking of parameters is done by averaging the sensitivity of different modelling outputs to the parameter:

\[
\bar{s}_j = \frac{1}{n} \sum_{i=1}^{n} s_{ij},
\]

Then the collinearity index is quantified which quantifies the minimum achievable norm of a linear combination of the normalised sensitivity functions with normalised coefficients (Brun et al., 2000):
where $\beta_j$ are the coefficients of a possible linear regression and the matrices $S_j$ represents the relative sensitivity $(\tilde{s}_{i,j})$.

$\gamma(\theta) = \sqrt{\min \left( \frac{1}{\min \left( \sum_{j=1}^{m} \tilde{s}_{i,j} \beta_j \right)} \right)}$ \hspace{1cm} (3)

$\gamma(\theta)$ is unity for linearly independent sensitivity functions and approaches infinity with increasing degree of dependence. Note that $\gamma(\theta)$ gives information on problems due to “compensability” within the subset of parameters thus allowing for analysing the behaviour of the model when a parameter is neglected assuming a fixed value. A first set of identifiable parameters is constructed from the sensitivity ranking neglecting the parameters with lower average sensitivity function (eq. 2). This phase allows for dropping those parameters that are characterised by a low model sensitivity and, for this reason, they can be considered non identifiable. After the first selection, parameter subsets can be obtained starting with the most sensitive parameter and adding less sensitive ones unless the collinearity index becomes too large (larger than a threshold of the order of 10 to 15). Parameters that lead to a too severe increase in the collinearity index are omitted and the procedure continues until either no additional parameter is found leading to a moderate increase in the collinearity index value.

The technique proposed by Weijers and Vanrolleghem (1997) also starts with a sensitivity ranking based on the same averaged sensitivity indexes presented in Eq. 3 in order to preliminary select only sensitive parameters. The final subset selection from those parameters is done by a criterion based on the determinant or the condition number of the Fisher information matrix (FIM):

$$FIM = S^T S$$ \hspace{1cm} (4)

Based on FIM, two functions can be used for studying the shape and the size of the modelling confidence region, i.e. the region in the parameters space where the model outputs are more sensitive to parameters variation:

$$CN = \frac{\max \{EV[S^T S]\}}{\min \{EV[S^T S]\}}, \hspace{0.5cm} DET = 2^{\frac{m}{2}} \sqrt{\det S^T S}$$ \hspace{1cm} (5)

where the matrices $S$ represent the sensitivity $(\tilde{s}_{i,j})$ and $m$ is the number of parameters. The function CN is a measure of the shape, the function DET is a measure of the size of the confidence region (small size for large values of DET) if the parameters would have been estimated by the method of least squares. Both techniques described above are local analyses that use only derivatives of model results with respect to model parameters for a specified set of parameter values. Literature does not provide critical values for CN or DET thus the approach can be considered as a comparative tool between different modelling parameters subsets in order to understand the maximum obtainable information from measured data and provide modelling modifications or field campaign extension for obtaining a more robust calibration of the model (Weijers and Vanrolleghem, 1997).

### 3 THE IDENTIFICABILITY ANALYSIS APPLICATION

For the simulation of the whole system, a home-made model previously developed by the Authors has been adopted. 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 integrated model is made up mainly of three sub-models for the simulation of the components; each sub-model is divided in a quantity module and a quality one for the simulations, respectively, of the hydrographs and of the pollutographs. The analysis has been applied to an experimental catchment in Bologna (Italy) which consists of a part of the Bologna sewer network and a reach of the Savena river (Artina et al., 1999). The WWTP has been neglected in the present study because it is connected to the river downstream from the analysed area. For details on modelling approach refer to previous literature (Mannina et al., 2004; Mannina, 2005; Mannina et al. 2006). Available measured data consist of RWB flow (Q), BOD and oxygen demand (OD) at a specific river cross-section during 6 rainfall events (Artina et al., 1999). Nineteen parameters characterise the modelling modules considered in the study. Their brief explanation, their variation ranges and averaged sensitivity functions (eq. 2) are presented in Table 1. During the application of both identification methodologies presented in the previous paragraph, the parameters characterised by \( \tau_j < 0.2 \) have been neglected. The threshold value has been adopted by comparison with the other parameters; in fact, a gap is evident in Table 1 among scarcely sensitive parameters and the others. From the first screening process 5 parameters have been excluded by the following analyses as indicated in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \Delta \theta_j )</th>
<th>( \tau_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial hydrological abstraction [mm]</td>
<td>0.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Catchment runoff coefficient [-]</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Catchment linear channel constant [min]</td>
<td>7.0</td>
<td>14.0</td>
</tr>
<tr>
<td>Catchment linear reservoir constant [min]</td>
<td>6.0</td>
<td>26.0</td>
</tr>
<tr>
<td>Sewer linear reservoir constant [min]</td>
<td>6.0</td>
<td>18.0</td>
</tr>
<tr>
<td>Build-up rate in the Alley-Smith model [kg/(ha*day)]</td>
<td>4.0</td>
<td>12.0</td>
</tr>
<tr>
<td>Decay rate in the Alley-Smith model [day(^{-1})]</td>
<td>0.33</td>
<td>1.0</td>
</tr>
<tr>
<td>Wash-off coefficient in the Alley-Smith model [mm(^{-1})]</td>
<td>0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>Wash-off factor in the Alley-Smith model [-]</td>
<td>0.12</td>
<td>3.20</td>
</tr>
<tr>
<td>Erosion factor in sewer in the Skipworth model [kg]</td>
<td>0.03</td>
<td>15.00</td>
</tr>
<tr>
<td>Sewer suspended load linear reservoir constant [min]</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>Sewer bed load linear reservoir constant [min]</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>CSO activation dilution factor [-]</td>
<td>1.0</td>
<td>10.0</td>
</tr>
<tr>
<td>Max dilution factor allowed downstream the CSO [-]</td>
<td>10.0</td>
<td>22.0</td>
</tr>
<tr>
<td>Riverbed roughness (Gauckler–Strickler) [m(^{1/3})/s]</td>
<td>20.0</td>
<td>90.0</td>
</tr>
<tr>
<td>Longitudinal dispersion coefficient [m(^2)/s]</td>
<td>59.3</td>
<td>179.0</td>
</tr>
<tr>
<td>De-oxygenation coefficient [s(^{-1})]</td>
<td>55.0</td>
<td>166.0</td>
</tr>
<tr>
<td>Sediment oxygen demand coefficient [s(^{-1})]</td>
<td>21.0</td>
<td>63.0</td>
</tr>
<tr>
<td>Re-aeration coefficient [s(^{-1})]</td>
<td>1685.0</td>
<td>4995.0</td>
</tr>
</tbody>
</table>

Table 1. Model parameters variation range and average model sensitivity (parameters neglected after pre-selection in italic)
4 RESULTS DISCUSSION

After the first screening, parameters subsets have been randomly generated with different dimension between 13 and 5. In each subset, the selected number of parameters has been considered variable while the others have been blocked to the average value in the variation range. All the possible parameter combinations have been analysed with a total number of considered subsets equal to about 14,000. Collinearity and FIM analyses have been performed on each parameter subset in order to obtain some interesting consideration about the maximum number of identifiable parameters with available measured data and the amount of irreducible uncertainty that cannot be eliminated in the current measures layout (i.e. without introducing new measuring stations or new measured variables).

The adoption of collinearity criterion shows that collinearity index grows rapidly with increase of parameters number in the analysed subset. Figure 1 shows the minimum reachable collinearity index value for each parameter subset dimension (Np). The threshold value, proposed by Brun et al. (2000) and equal to 15, is reached considering subsets dimension equal to 7 (Figure 1). Thus, according to this criterion the maximum number of identifiable parameters should be 7. Lower parameter subset dimension should lead to imprecise modelling response and inefficient use of available data during calibration. The increase of parameters subsets dimension leads to a rapid increase in the collinearity index demonstrating a progressively wider capacity of the parameters to “compensate” each other variation. This effect should lead to increasing equifinality between parameters thus reducing the confidence of the operator in modelling calibration and results robustness.

The FIM analysis has been performed adopting both criteria presented in equation 5. Both for CN and for DET, maxima have been taken for each parameter subset dimension (Figure 1). FIM criteria showed a maximum analysing subsets dimension equal to 7 thus confirming the results given by collinearity criterion. The confidence region shape criterion (CN) showed lower variability with the dimension of parameters subset while the size criterion (DET) demonstrated higher variability. Especially the DET criterion allows for understateing that, for lower parameter subset dimension, the confidence region is large depending on the fact that the model, simplified by the fact that several parameters are blocked to constant values, requires large parameters variability to adapt the measured data; increasing the number of variable parameters in the subset above 7, the equifinality among parameters thus increasing the extension of the confidence region. Analogous considerations can be made for the CN criterion although doubts may rise in considering the shape of the confidence region as a possible index of parameter identifiability.

All the proposed criteria converge also on the list of identifiable parameters: catchment runoff coefficient, sewer linear reservoir constant, build-up rate in the
Alley-Smith model, wash-off coefficient in the Alley-Smith model, CSO activation
dilution factor, riverbed roughness (Gauckler–Strickler) and de-oxygenation
coefficient. Differently from collinearity approach, during the FIM analysis, the best
parameter grouping has been selected on statistical basis by selected the 7 most
frequent parameters present in the best 5% of the analysed subsets. This approach
has been used after understanding that several 7 parameters subsets had similar CN
and DET values so that the best combination of parameters is not really
representative of the whole model behaviour.

In order to understand the level of uncertainty connected with the identifiable and non
identifiable parameters, an uncertainty analysis have been performed using GLUE
approach (Beven and Binley, 1992) in the case where all parameters are considered
variable (Fig. 2a, 2b and 2c) and uncertain and in the case where non identifiable
parameters are blocked to an average value in the variation range (Fig. 2d, 2e and
2f). This comparison allows understanding the amount of uncertainty that can be
reduced by continuing the presently available monitoring campaign and the amount of
uncertainty that can be only reduced by extending the monitoring campaign to other
variables and monitoring points.

The same approach is also a measure of the increasing confidence of the operator in
modelling results after the identifiability analysis. In fact, the identifiability analysis is
able to define the limits of calibration and it can address the operator to those
parameters that can be blocked to a constant value without losing in model
applicability and gaining the maximum from available data. The comparison between
figure 2 shows that the uncertainty band both for the Q and for DO can be reduced of
about 40% by blocking non identifiable parameters.

5 CONCLUSIONS
The present paper applied different parameter identifiability approaches to an urban
drainage integrated model. Integrated models are often characterised by high
complexity and calibration problems connected with the discrepancy between
available data and model parameterization. The study allowed for some interesting considerations:

- all the adopted criteria are efficient and unambiguous in determining the number of identifiable parameters;
- collinearity approach is surely simpler but doubts may rise during the application for defining the threshold to define excessively compensable parameters; in the present study, the growth of the collinearity index was evident between 7 and 8 parameters subsets, but this condition may be sometimes weak if the approach is applied without comparison with other criteria;
- FIM analysis showed to be a powerful method for analysing model sensitivity and parameter identifiability; nevertheless, the approach can gain in robustness if coupled with a statistical analysis of the occurrence of different parameters in the better identifiable subsets;
- the identifiability analysis is also a powerful tool for increasing the operator confidence in modelling results; the approach can be used for blocking some non identifiable parameters thus wisely modifying the structure of the model and reducing the correspondent uncertainty.

LIST OF REFERENCES


