Development of a semi-automated model identification and calibration tool for conceptual modelling of sewer systems

Vincent Wolfs¹, Mauricio Florencio Villazon²,³, Patrick Willems⁴

¹ University of Leuven, Hydraulics Section, Kasteelpark Arenberg 40, BE-3001 Leuven, Belgium
vincent.wolfs@bwk.kuleuven.be
² University of Leuven, Hydraulics Section, Kasteelpark Arenberg 40, BE-3001 Leuven, Belgium,
mauricio.villazon@bwk.kuleuven.be
³ Universidad Mayor de San Simón, Hydraulics Laboratory, Petrolera Avenue Km 4.2 Cochabamba, Bolivia
⁴ University of Leuven, Hydraulics Section, Kasteelpark Arenberg 40, BE-3001 Leuven, Belgium,
patrick.willems@bwk.kuleuven.be

ABSTRACT
Applications such as real-time control, uncertainty analysis and optimization require an extensive number of model iterations. Full hydrodynamic sewer models do not suffice for these applications due to the overlong computation time. Simplifications are consequently required. A lumped conceptual modelling approach results in a much faster calculation. The process of identifying and calibrating the conceptual model structure could, however, be time consuming. Moreover, many conceptual models lack accuracy, or do not account for backwater effects. To overcome these problems, a modelling methodology was developed which is suited for semi-automatic calibration. The methodology is tested for the sewer system of the city of Geel in the Grote Nete river basin in Belgium, using both synthetic design storm events and long time series of rainfall input. A MATLAB/Simulink® tool was developed to guide the modeller through the step-wise model construction, reducing significantly the time required for the conceptual modelling process.

KEYWORDS
Conceptual modelling; CSO; Semi-automatic calibration; Sewer system

1 INTRODUCTION
Many applications, such as long term simulations, uncertainty analysis and real time control of structures, require hydraulic models of sewer systems with very little computation time. At the same time, the results must be accurate to meet the purposes in mind. This contradiction makes it difficult to find a generally applicable method. The current methods can be classified in two groups.
The first group consists of the physically based models, emanating from de Saint-Venant equations. Full numerical solution of de Saint-Venant equations, denoted as dynamic routing, can be done using several numerical schemes (Cunge & Verwey, 1980). Implicit schemes offer the most accurate and stable solution, and are therefore used by the majority of commercial software, such as InfoWorks (Innovyze, UK), SWMM (EPA, US) and MIKE URBAN (DHI, DK). Different levels of simplifications of de Saint-Venant equations exist for describing unsteady flow. The simplest approximation, namely the kinematic wave model, is easier and significantly faster to solve. However, this approach is inapplicable when backwater conditions predominate. Therefore, some models, like the SURKNET model of Pansic and Yen (1982) and the SERAIL model of Chocat et al. (1983), use the kinematic wave approximation to compute the flow and switch to a simple pressurized algorithm when backwater effects prevail. The significant simplifications result in a limited usability of the model and reduced accuracy (Duchesne, 2001). Moreover, the computation time is still too large for the intended applications. Motiee et al. (1997) used a storage model with a combination of the kinematic and diffusion wave models. Finding a solution is an iterative process, thus augmenting computation time.

The second group consists of conceptual models. These use conceptual relations instead of momentum equations, while respecting conservation of mass. The far smaller computation effort puts them in favour to the above-mentioned models. Well known examples are the unit hydrograph method and Muskingum-Cunge routing (Cunge, 1969), where the typically calibrated Muskingum parameters are related to hydraulic and physical characteristics such as reach length and channel bed slope (Ponce, 1979). Many elaborations exist originating from the basic Muskingum method (e.g. Achleitner et al., 2007; Birkhead and James, 2002; Franchini, 2011; Kumar 2011). However, most of these focus on river modelling and cannot deal with reverse flows or significant backwater effects. Therefore, problems can arise when calculating flows in flat looped networks for example.

The aim of the presented research is to establish a general methodology to generate models that comply with the following requirements:

- be able to describe whatever type of flow (free surface and surcharged), thus including backwater effects;
- calculate overflow discharges accurately;
- have very short computation times;
- be stable and robust;
- allow semi-automatic model structure identification and calibration.

In order to be generally applicable in practice, this methodology must be as simple as possible. This can only be achieved using a physically based approach.

2 MODEL PRESENTATION

2.1 Concept

The sewer system is represented as a reservoir model, with an inflow, throughflow and overflow. The continuity equation forms the base of the model. In order to identify and calibrate the model structure that characterizes the system, knowledge is needed about the incoming and outgoing flows. Since very limited accurate measurement data on these aspects is available, the characterization is done using the simulations of a detailed hydrodynamic model. The used models are assumed to mimic reality accurately.

This research is limited to the hydraulic aspect of the sewer system. Therefore, the runoff flow that enters the sewer system in the hydrodynamic simulations is used as inflow in the simplified model.
2.2 Characterization of the sewer system

2.2.1 Concentration time

The concentration time is the time the rainfall needs in order to travel from the remotest place in the catchment (in time, not necessarily in distance) to the point in the sewer system where the design calculation is made (Chow, 1988). Therefore, the concentration time of a sewer system is the sum of the inlet time (time the water flows over the surface before entering the sewer system) and the flow time (time the water flows through the sewer system from the inlet point to the design point). Since a single reservoir model has an instantaneous (exponential) response, it cannot reproduce the peak shift between inflow and outflow. Therefore, smoothing of the input is necessary before entering the reservoir system.

The smoothing of the inflow hydrograph is done by averaging the runoff over the concentration time. This way, part of the smoothing due to the flow in the sewer system will already be included in the simplified model. The concentration time is determined as the center of gravity difference between rainfall input and downstream flow (sum of the throughflow and overflow discharges). Due to the lumped modelling approach, insight into the flow in every individual pipe is unnecessary. Therefore, the concentration time is determined for the total downstream flow.

Note that as the flow time and inlet time are dependent on the flow velocity, and the velocity is in some extent related to the rainfall input, the concentration time will not be a constant value for a given location in the system. Good results are, however, often obtained when a constant $T_c$ is applied. To keep the model as simple as possible, the variability of the concentration time is neglected.

2.2.2 Storage/throughflow-relationships

The relationships between the state parameters (i.e. the storage) and the flow parameters (i.e. throughflow and overflow discharge) form the core of the model. To clarify the characterization of a system, a small fictitious gravity sewer system was built in InfoWorks CS in accordance with the conventional design rules. Figure 1 visualizes the relationship between storage and throughflow for this system. Several hydrodynamic simulations were performed using a wide range of composite storm events to obtain the throughflow and overflow discharge. By applying the continuity equation to the inflow and outflow, the storage volume can be obtained at every time step.

![Storage/throughflow-relationship](image)

Figure 1: Storage/throughflow-relationship for a small gravity sewer system for a range of composite storms with different frequencies $f$ [p.a.] and return periods $T$ [years].
One can clearly notice a much larger storage during the rising part of the hydrograph (i.e. increasing rainfall intensity), than during the falling branch. These hysteresis effects are typical for sewer systems where backwater effects occur. If the backwater effect is not very significant, one can use an average curve to describe the storage/throughflow-relationship. However, this paper proposes a different approach by dividing the storage in several (artificial) components. This way, the model will be able to account for backwater effects.

One part of the storage, denoted as static storage, can unambiguously be related to the throughflow by one or more linear relationships. The maximum static storage is defined as the storage in the sewer system when the system is filled up to the crest of the overflow, the water is stagnant and there is no throughflow (Vaes, 1999). This situation will often be reached at the end of the overflow event, when the incoming flow is negligible.

The remaining part of the storage is called dynamic storage. Vaes (1999) showed that this storage is correlated to the maximum overflow discharge. Moreover, it is likely there is a connection to the incoming flow as well. Hence, the dynamic storage is probably described best by a relationship that takes the inflow and total outflow into account. For practical reasons however, the dynamic storage is often only linked to the inflow (e.g. Muskingum method). Figure 2 shows an example where the dynamic storage is included as a function of solely the inflow, after smoothing the inflow over a concentration time. The static volume was calculated using a bi-linear relationship with the throughflow. The agreement between the total storage obtained by a full hydrodynamic simulation and the one obtained using the above mentioned relationships is close.

![Figure 2: Piecewise linear “dynamic” approach (with slopes k_{stat} and k_{dyn}) for the storage/throughflow-relationship of a small gravity sewer system for the composite storm with frequency 7 p.a.](image)

When the overflow starts spilling, significant additional backwater effects can occur, resulting in an increased storage. Therefore, it can be difficult to relate the dynamic storage unambiguously to the incoming flow. This can be resolved by defining a third kind of storage during an overflow event, namely an extra static storage. This is the storage enclosed by the horizontal planes through the crest level on the one hand and through the maximum water level above the crest of the overflow on the other hand (Fig. 3). This extra static storage can be related to the overflow.
2.3 Transfer Function (TF) models

In order to obtain a methodology suitable for the semi-automatic identification and calibration of the conceptual model structure, based on the system characterization principles outlined above, use is made of general transfer function (TF) models.

2.3.1 General system

Chow et al. (1988) suggested the following storage function for linear systems:

\[
S = a_1 Q + a_2 \frac{dQ}{dt} + a_3 \frac{d^2Q}{dt^2} + \ldots + a_n \frac{d^{n-1}Q}{dt^{n-1}} + b_1 I + b_2 \frac{dI}{dt} + b_3 \frac{d^2I}{dt^2} + \ldots + b_m \frac{d^{m-1}I}{dt^{m-1}}
\]  

in which \(a_1, \ldots, a_n\) and \(b_1, \ldots, b_m\) are time invariant coefficients and \(I\) and \(Q\) are the inflow and outflow respectively. Differentiating the reservoir storage \(S\), substituting in the continuity equation and rearranging yields:

\[
Q_t = \frac{b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}}{1 - a_1 z^{-1} - \ldots - a_n z^{-n}} z^{-\delta} I_t
\]  

The backwards shift operator \(z\) is defined as \(z^{-1} Q_t = Q_{t-1}\). Parameters \(n\) and \(m\) are called the transfer function orders, and \(\delta\) is a pure time delay for the input series. It is important to note that a transfer function can represent an arrangement of linear reservoirs (Beven, 2001). This makes the use of TF’s particularly useful in hydrology and other modelling applications.

2.3.2 Mathematical approach of the static-dynamic theory

The continuity equation for discrete time steps is given by:

\[
S_t = S_{t-1} + (I_{t-1} - Q_{t-1}) \cdot \Delta t
\]  

The concept presented in Section 2.2.2 states that the total volume can be divided in a static and dynamic part:

\[
S_t = S_{\text{dyn},t} + S_{\text{stat},t}
\]  

Relating the static storage \(S_{\text{stat}}\) to the outflow \(Q\) and the dynamic storage \(S_{\text{dyn}}\) to the inflow \(I\), both using simple time-independent linear relations, yields:
\begin{align*}
S_{\text{stat},t} & = r_s \cdot Q_t , \quad (5) \\
S_{\text{dyn},t} & = r_d \cdot I_t , \quad (6)
\end{align*}

Substituting (4), (5) and (6) in (3), and rearranging leads to:

\begin{align*}
Q_t \left( 1 - \left( 1 - \frac{\Delta t}{r_s} \right) \cdot z^{-1} \right) = I_t \left( -\frac{r_d}{r_s} + \left( \frac{r_d + \Delta t}{r_s} \right) \cdot z^{-1} \right) \quad \text{(7)}
\end{align*}

One can easily notice the resemblance with the more general formulation Eq. (2). The static-dynamic theory in its simplest form thus represents two linear reservoirs in parallel, where the second one has a time delay of one time step. This approach is highly similar to the Muskingum method, but provides greater flexibility. For instance, one can make the dynamic volume dependent on the inflow on time step \( t \) and \( t-1 \):

\begin{align*}
S_{\text{dyn},t} = r_{d,1} \cdot I_t + r_{d,2} \cdot I_{t-1} \quad (8)
\end{align*}

Substituting (4), (5) and (8) in (3) yields:

\begin{align*}
Q_t \left( 1 + \frac{\Delta t - r_s}{r_s} \cdot z^{-1} \right) = I_t \left( -\frac{r_{d,1}}{r_s} + \left( \frac{r_{d,1} - r_{d,2} + \Delta t}{r_s} \right) \cdot z^{-1} + \left( \frac{r_{d,2}}{r_s} \right) \cdot z^{-2} \right) \quad (9)
\end{align*}

It is clear that TF’s will play a crucial role in the conceptual modelling process. However, the proposed relationships (e.g. Eq. (5), (6) and (8)) will often be too elementary, and more complex model structures are needed to provide accurate results. This can be achieved using piecewise linear relationships. Multiple successive linear segments can approximate almost every possible relationship. This way, the extra storage during overflow events can be incorporated as well. The proposed solution above, and the associated advantages, can be retained by translating the coordinate system: the origin of the new coordinate system is located at the transition from one linear relationship to another. Nevertheless, calibration of such model structures is cumbersome. Therefore, the calculation of the static and dynamic storage is uncoupled. In addition to the analytical solution to find the outgoing flow (e.g. Eq. (7) and (9)), this is a second methodology to conceptually model sewer systems. The static storage is calculated based on piecewise linear relationships, whereas the dynamic storage is modelled as an arrangement of linear reservoirs depending on the incoming flow. Hence, the dynamic volume is calculated using a TF.

### 2.4 Automatization of the model identification and calibration process

To facilitate the calibration process, a MATLAB/Simulink© tool was developed incorporating the two methodologies above. To improve user-friendliness, a GUI was added. The modeller must specify the incoming (rainfall runoff and wastewater flow) and outgoing flow, as well as the overflow discharge. Based on the prompted hydrographs, the tool suggests the most appropriate model structure and relationships. The user can adopt these, or choose others based on interactive plots and statistical criteria.

Identification of the most appropriate TF model structure is done using the CAPTAIN Toolbox (Taylor et al., 2007). The Cross-Correlation function (Box and Jenkins, 1970) and various statistical criteria, like the ubiquitous Coefficient of Determination ($R^2$) and the “Young Information Criterion” (YIC, Young, 1984) are used. YIC is a criterion that combines elements of goodness of fit and standard errors on the coefficients. The higher the model order, the larger the standard errors will tend
to be. Large standard errors on the coefficients are an indication that the model is over-parameterised (Taylor et al., 2007). The same toolbox is applied for the estimation of the parameters. Well-known methods of least-squares and maximum-likelihood are implemented in the modelling tool, but also the far more complex Refined Instrumental Variable algorithm (RIV) (Pedregal, 2007). When modelling basic models such as Eq. (7), the modeller can use self-chosen, physically meaningful values for \( r_s \), \( r_d \) and \( \Delta t \) as well.

After the identification and calibration process, the model structure and parameters are transferred to Simulink®. The Simulink platform is tailored for time dependent and dynamic simulations. The included user interface is block oriented allowing straightforward coupling of the modelled subsystems. Besides the pre-existing blocks implemented in Simulink, users can define their own blocks.

3 APPLICATION EXAMPLE

The developed tool was used for modelling the combined sewer system of Geel, located in the relatively flat Grote Nete catchment in Flanders. An accurate InfoWorks CS model of this sewer system was made available by the water company Aquafin, which was calibrated using measurements. The sewer system was divided in multiple subsystems, based on the presence of throttle pipes. The following paragraph discusses the simulation results of one such subsystem. The system has a reduced (impervious) area of 12.33 ha and a population equivalent (pe) of 430. It counts three CSOs, which spill into the river Grote Nete.

Composite storms are used for hydrodynamic simulations with frequencies ranging from 20 (‘f20’) to 1 per year (‘f1’), and return periods from 2 (‘T2’) to 20 years (‘T20’). All storms were developed by Vaes et al. (1996). For every frequency or return period, all Intensity/Duration-relationships of Uccle are included in a design storm. The Intensity/Duration/ Frequency – relationships are based on a historical rainfall series of 27 years (for the period 1967-1993) with a time step of 10 minutes recorded by the rain gauge at Uccle, the location of the main meteorological station in Belgium. The maximum rainfall intensities of the storms vary from 7.51 mm/h to 65.21 mm/h. All storms have a time step of 10 minutes and a duration of 48 hours.

The sewer system was conceptually modelled using the two previously described methodologies: the analytical solution of the derived TF (Eq. (7)) (denoted as “CM: method 1”), and the more complex system with uncoupled static and dynamic storage (“CM: method 2”). To present the results more comprehensible, the flow over the three CSOs is cumulated. CSO structures can be modelled separately though. The following criteria (Grecu and Krajewski, 2000) were used to evaluate the quality of the calibration, with \( S \) and \( C \) the simulated values of the hydrodynamic and the conceptual model respectively:

\[
IA = 1 - \frac{\sum_{i=1}^{N}(S_i - C_i)^2}{\sum_{i=1}^{N}(S_i - \bar{S})^2 + (C_i - \bar{C})^2} \quad [0, 1] \quad \text{Index of agreement}
\]

\[
NSE = 1 - \frac{\sum_{i=1}^{N}(S_i - C_i)^2}{\sum_{i=1}^{N}(S_i - \bar{S})^2} \quad [-\infty, 1] \quad \text{Nash-Sutcliff Efficiency}
\]

\[
B = \frac{\bar{C}}{\bar{S}} \quad [-\infty, +\infty] \quad \text{Bias}
\]
Comparative plots of simulated throughflow and overflow by both the hydrodynamic and conceptual model are shown in Fig. 4.

Both methods manage to mimic the hydrodynamic simulations relatively well, although the falling part of the hydrograph describing the throughflow obtained by the CM Method 1 is somewhat too smoothed. The sudden decline of the throughflow in the rising branch due to backwater effects can be approximated satisfactory by the two methods as well. The bias B (Table 1) is close to unity, which indicates a good approximation of the outgoing volume. The Index of Agreement (IA) and Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970) confirm the adequate fit. CM Method 2 surpasses the simpler model in every aspect. The advantages of the more complex modelling method become even clearer when simulating the overflow. Although the first method succeeds in predicting the overflow volumes well, the shape of the hydrograph deviates significant from the hydrodynamic simulation. The second modelling method results in a far better fit.

Table 1: Summary of quality indexes with CM1 and CM2 denoting respectively modelling methods 1 and 2.

<table>
<thead>
<tr>
<th>Event</th>
<th>IA</th>
<th>Throughflow</th>
<th>Cumulated CSO discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CM1</td>
<td>CM2</td>
<td>IA</td>
</tr>
<tr>
<td>f20</td>
<td>0.971</td>
<td>0.992</td>
<td>0.905</td>
</tr>
<tr>
<td>f10</td>
<td>0.990</td>
<td>0.996</td>
<td>0.962</td>
</tr>
<tr>
<td>f7</td>
<td>0.991</td>
<td>0.997</td>
<td>0.964</td>
</tr>
<tr>
<td>f1</td>
<td>0.992</td>
<td>0.998</td>
<td>0.968</td>
</tr>
<tr>
<td>T2</td>
<td>0.988</td>
<td>0.995</td>
<td>0.954</td>
</tr>
<tr>
<td>T5</td>
<td>0.983</td>
<td>0.993</td>
<td>0.932</td>
</tr>
<tr>
<td>T10</td>
<td>0.983</td>
<td>0.992</td>
<td>0.932</td>
</tr>
<tr>
<td>T20</td>
<td>0.981</td>
<td>0.988</td>
<td>0.920</td>
</tr>
</tbody>
</table>
Next, a long term simulation was performed using 15-minute rainfall measurements from a nearby rain gauge for the period 13/09/2001 to 11/01/2010. Even if the conceptual model uses a time step of one minute, the computation time remains less than 3 minutes. Fig. 5 shows the simulation results of the hydrodynamic and conceptual model. Although the error on some individual emission events can be quite extensive, the general shape and volume are predicted accurately.

![Figure 5: Compressed time series of the overflow discharges obtained by simulations with the full hydrodynamic and conceptual model, for one of the overflow locations in the system of Geel](image)

4 CONCLUSIONS

Hydraulic models with very short calculation time are indispensable for applications such as real time control and uncertainty analysis. However, current conceptual models often lack accuracy in predicting throughflow and CSO discharges. In addition, the calibration process is often very time-consuming. The presented research aims to overcome these limitations by the development of a new conceptual modelling approach. The basic idea behind the methodology is the division of the sewer system’s storage in several artificial components. Elaboration of the theory resulted in two methods with great flexibility, which were incorporated in a MATLAB/Simulink® tool. The tool enables semi-automatic model structure identification and calibration to speed up the modelling process. Afterwards, the model is built in Simulink. The methodology and tool were used to simulate throughflow and overflow discharges of the sewer system of the city of Geel using composite storm events and long term series of rainfall input. Calibration was done based on simulation results obtained by a full hydrodynamic InfoWorks CS model. The conceptual model produced highly similar results with a significant increase in computational speed. Furthermore, results remained stable regardless of the time step used.

5 ACKNOWLEDGEMENTS

This research was supported by the Agency for Innovation by Science and Technology in Flanders (IWT). The author would like to thank the Flemish water company Aquafin, who provided the InfoWorks CS model of the system of Geel, and Innovyze for the software support.

6 REFERENCES


Vaes G., Willems P. and Berlamont J. (1996). IDF-relationships and composite storms for the design of sewer systems (in Dutch), Hydraulics Laboratory, University of Leuven, Belgium.


7 QUESTIONARY

1. Name of the author that will present the paper: **Vincent Wolfs**

2. Is the first author an young researcher (according to IWA, under the age of 35):  **Yes**

3. Have you submitted the extended abstract yet:  **Yes**

4. Have you previously published this paper at another conference or Journal:  **No**

5. Do you want to have the full paper offered for the Journal:
   - Water Science and Technology:  **Yes**
   - Journal of Hydraulic Research:  **Yes**

6. Have you checked your paper for style, formatting and English language:  **Yes**