Modelling real-time control of WWTP influent flow under data scarcity
Stefan Kroll, Geert Dirckx, Brecht M. R. Donckels, Mieke Van Dorpe, Marjoleine Weemaes and Patrick Willems

ABSTRACT

In order to comply with effluent standards, wastewater operators need to avoid hydraulic overloading of the wastewater treatment plant (WWTP), as this can result in the washout of activated sludge from secondary settling tanks. Hydraulic overloading can occur in a systematic way, for instance when sewer network connections are extended without increasing the WWTP’s capacity accordingly. This study demonstrates the use of rule-based real-time control (RTC) to reduce the load to the WWTP while restricting the overall overflow volume of the sewer system to a minimum. Further, it shows the added value of RTC despite the limited availability of monitoring data and information on the catchment through a parsimonious simulation approach, using relocation of spatial system boundaries and creating required input data through reverse modelling. Focus was hereby on the accurate modelling of pump hydraulics and control. Finally, two different methods of global sensitivity analysis were employed to verify the influence of parameters of both the model and the implemented control algorithm. Both methods show the importance of good knowledge of the system properties, but that monitoring errors play a minor role.

Key words | data scarcity, global sensitivity analysis, modelling, real-time control, system reduction, waste water treatment plant influent

INTRODUCTION

Real-time control (RTC) is known to be a solution to a variety of problems that can occur in existing urban drainage systems. Several applications of RTC have been described in literature (e.g. Seggelke et al. 2005, 2013; Tränckner et al. 2007), where the operation of existing systems was optimised by developing integrated control strategies for the sewer system and the wastewater treatment plant (WWTP). The case study presented in this paper, however, makes use of rule-based RTC as a means to mitigate the impact of additional catchment area being connected to the WWTP before the treatment plant could be upgraded accordingly. More specifically, several scenarios were evaluated for controlling existing pumping stations (PSs) in the sewer network, where the objective was to minimise the environmental impact characterised by total combined sewer overflow (CSO) volume while restricting the total influent flow rate to the WWTP to its maximal capacity.

Since the availability of information on the catchment was limited, a parsimonious simulation approach was adopted. While the modelling of sewer system storages and flows was done by means of storage units and empirical relations, the simulation of pump flows required a more detailed approach, explicitly considering the switch-on and switch-off delay and pumping rates depending on the water level in the pump pit. As no detailed data on water quality were available, the requirements of the study focussed on CSO volume as the performance criterion. To gain deeper insights into the behaviour of the model and the relevance of the tested control scenario and to determine the most influential parameters, two established methods for global sensitivity analysis (GSA) were applied, i.e. Morris screening (Morris 1991) and extended FAST (Fourier amplitude sensitivity testing as described by Donckels et al. (2014)).
METHODS AND MATERIALS

Case study

The case study has been carried out for the catchment of Olsene in western Flanders, Belgium. Figure 1 gives an overview of the study’s layout and key data.

The WWTP of Olsene receives combined wastewater from five main trunk sewers that collect the wastewater of four different municipalities. Each of these collectors ends in a PS that directly feeds the WWTP. Two of the branches were connected only recently, thereby increasing the number of connected person equivalents (PE) from 13,000 to 23,500. As the WWTP was designed to treat the wastewater of 21,000 PE, it is by design hydraulically overloaded during times of heavy rainfall when all the PSs are expected to work at their maximum pump capacity. Sludge concentrations in the WWTP had to be kept low in order to avoid washout from secondary clarifiers during peak loading. As a result, despite its sufficient aeration capacity, the WWTP effluent did not comply with the Flemish standard (Heyman & Smout 2010), demanding 75% nitrogen removal.

Although an upgrade of the WWTP to a capacity of 28,000 PE is planned for the mid-term, a short term solution was required to reduce the input to the WWTP during wet weather periods, allowing the WWTP to cope with the new loading scheme without decreasing operational performance to an unacceptably low level. Since all PSs were already equipped with the required hardware for data transmission to the WWTP, the application of RTC seemed a logical and cost-efficient choice. This choice was also supported by the specific characteristics of the sewer network, such as oversized collector pipes (creating considerable inline storage) and overall low invert slopes. A direct evaluation of the studied area with tools for the estimation of RTC potential, such as suggested by Schütze et al. (2004) or Zacharof et al. (2004), was not possible due to the lack of detailed data. However, a previously conducted screening on a number of comparable catchments in Flanders revealed a high potential for RTC in catchments with comparable characteristics (Dirckx et al. 2007).

Modelling approach and data issues

Determining inputs for model simulations

The modelling approach, implemented in Matlab-Simulink, was limited to the very essential parts of the considered system. Detailed surface data of the catchment, required for rainfall-runoff modelling, were not available. However, the availability of monitoring data for four of the five PSs allowed for the relocation of the spatial boundaries of the system (Vanrolleghem et al. 2005) without necessitating assumptions on catchment and rain data. This means that the resulting model only describes those parts of the system that are directly influenced by the applied control strategy. For each sewer branch, the system is composed of
Modelling RTC of WWTP influent flow under data scarcity

S. Kroll et al. | Modelling RTC of WWTP influent flow under data scarcity | Water Science & Technology | 73.7 | 2016

(1) a volumetric storage unit that simulates the storage behaviour of the sewar main trunk that is drained by the PS and (2) the PS itself. For each such sub-system it is possible to set up the mass balance as follows:

\[
\frac{dV}{dt} = Q_{in} - Q_{out} - Q_{cso}
\]  

(1)

where \(Q_{in}\) is the sum of all inflows to the sub-system (m³/h), \(Q_{out}\) is the flow rate of the PS (m³/h), \(Q_{cso}\) represents the CSO discharge (m³/h) and \(V\) is the activated storage volume in the upstream sewer and the pump pit (m³). CSOs at the PSs were simulated for the different scenarios, but were neglected when generating the input \(Q_{in}\) data as their influence on the total volume balance is negligible.

As providing the data via rainfall-runoff simulations was not an option, \(Q_{in}\) (assumed to be not influenced by the operation of the PS) was generated by reverse modelling (as described by, for example, Leonhardt (2014)) from historical data sets of monitored pump flows \(Q_{out}\) and water levels that reflect the activated storage volume. The conversion from water level to volume was carried out using storage curves, i.e. curves based on the storage volume, assuming one (single) functional relationship between activated static storage and water level (further details on the concept are given by Dirckx et al. (2011)). This automatically implies that the dynamic storage is neglected, but due to the reverse modelling approach dynamic effects are accounted for in the input data. For one PS (PS Heuvelhoekstraat) no monitoring data were available. Here, \(Q_{in}\) was estimated from another PS with existing data after applying a correction factor that accounts for the difference in the number of connected PE.

**Modelled processes**

For the five collectors connected to the WWTP, five identical modules were created and individually parameterised according to the properties of the PSs and collectors. For each module, the mass balance was calculated according to Equation (1). Storage curves were applied to determine the water level in the pump pit, \(h\), based on the volume balance. To incorporate the dependence of pumping rates on water levels in the pump pit, all water levels and flows during wet weather were fitted using linear regression functions \(Q_{out} = f(h)\) limited by a maximum observed flow. Overflows \(Q_{cso}\) values were calculated using the Poleni formula (see, for example, Bollrich (2000)) based on the water level \(h\) in the pump pit:

\[
Q_{cso} = \frac{2}{3} \mu \sqrt{2g} \cdot b \cdot (\max(h - w, 0))^{3/2}
\]  

(2)

where \(\mu\) is the overflow coefficient (~), \(g\) is the gravitation acceleration (m/s²), \(b\) is the weir crest length (m), \(h\) is the water level in the pump pit and \(w\) is the weir crest level (m).

Switch-on and -off delays of pumps due to the inertia of the moving parts of the pump and the mass of the accelerated water column were modelled by linear functions in the form of

\[
Q_{out} = \frac{Q_{nom}}{T_{del}} t_{sim} \text{ for the interval } [T_{on}; T_{on} + T_{del}]
\]  

(3)

and

\[
Q_{out} = \frac{Q_{nom}}{T_{del}} t_{sim} \text{ for the interval } [T_{off}; T_{off} + T_{del}]
\]  

(4)

where \(Q_{out}\) is the actual pump flow (m³/h), \(Q_{nom}\) is the pump capacity (nominal flow) (m³/h), \(T_{del}\) is the pump delay (s) and \(t_{sim}\) is the time of the simulation environment (s), starting each time from 0 when the pump switch is being triggered.

**Control algorithm**

Two control algorithms were considered for the PSs, aiming at limiting the total influent flow to the WWTP in order to avoid its overloading (and potential non-compliance), i.e. one local and one global control scenario. For both scenarios, it was assumed that the WWTP is solely limited by its hydraulic capacity, which is reached after 1 to 2 hours, and that the aeration system can cope with the resulting peak loads at all times.

**Local control by time slots**

A simple, yet frequently used, approach to flow limitation problems is the application of time slots; i.e. each PS is operated in a cycle with a fixed length of time, where it is switched on and off for predefined time intervals. By varying the duration of the time intervals relative to the duration of the total cycle for each PS, the resulting flow rate to the WWTP can be tuned. In addition, if equal cycle times are chosen for all PSs, it is possible to equalise the total flow by synchronising all individual PSs. A potential downside of
Global control by ranked filling degree

This type of control necessitates continuous communication between the individual PSs and a central control unit. It requires the following input parameters: the maximum inflow capacity of the WWTP ($Q_{\text{max}}$ – this may be adapted by the WWTP operating staff at any time), the individual capacities of the PSs ($Q_i$) and the individual reduced capacities of the PSs ($Q_{i,\text{red}}$). These reduced flows can either be the result of working with only a part of the available pumps in the PS or frequency control. As shown in Figure 2, filling degrees (as detailed by Dirckx et al. (2011)) of each sewer branch are used to prioritise the decision whether or not a pump has to be switched on (and at what flow) following a ranking from high to low.

If the PS with the highest filling degree can be switched on at full capacity without exceeding the maximum allowed inflow, the setpoint will be adjusted accordingly. If the maximum allowed inflow would be exceeded, the algorithm checks if it can switch on the considered PS at its reduced capacity without exceeding the allowed inflow. If not, the PS will be switched off (setpoints for pump flow per PS, $sp_i = 0$). In the following step, the secondly ranked PS is considered and the original maximum allowed flow is now decreased with the flow assigned in the previous setpoint. This control loop is repeated until all setpoints are assigned.

Sensitivity analysis

A GSA has been carried out on the modelling approach in combination with the global RTC algorithm, in order to explore the influence of the model parameters on the simulated total CSO volume, which could not be validated based on monitoring data.

As Vanrolleghem et al. (2015) reported after a literature review, there are inconsistencies in the described parameterisation of the GSA methods (for instance, minimal required number of simulations) and the interpretation of the results. Consequently, two different GSA methods were applied and their results compared.

- The Morris screening method (Morris (1991), enhanced by Campolongo et al. (2007), creates multiple one-at-a-time experiments at different, randomly chosen points of the parameter space and evaluates the so-called elementary effects of the altered parameters on the simulation results. This method is focussed on low computational effort, but results of previous applications (Cosenza et al. 2016, Vanrolleghem et al. 2015) revealed problems in the convergence of the method. Here, a total of 8,800 simulations were performed for the 43 model input factors listed in Table 1.

- Extended Fourier amplitude sensitivity testing (eFAST) imposes individual sinusoidal frequencies on the model parameters throughout different scenarios and evaluates which frequencies can be detected in the simulation results. In this way, first and total order effects of the parameters can be determined (Donckels et al. 2014). For applying eFAST to this model, 55,900 simulations were performed.

Both GSA methods have been used with the set of parameters listed in Table 1. All parameter distributions were considered to be uniform. Convergence analysis was performed by the evaluation of the variability of the sum over all sensitivity indices normalised by the number of factors, as suggested by Vanrolleghem et al. (2015).
RESULTS AND DISCUSSION

Modelling

As an illustration, Figure 3(a) shows a comparison between measured data and simulation results for one of the pumping stations (PS Leiestraat) in the reference scenario (i.e. no control).

This example shows that the duration of the events at high pump flow resulting from simulations based on the synthetic inflow rate (\text{sim}(f(Q_{in}))) can deviate significantly from the monitored pump flow. Conversely, if the simulation is based directly on the monitored water level (\text{sim}(f(h))) the same events could be simulated with high accuracy. This shows that the use of synthetic inflow data does not lead to accurate results for individual events due to the non-linearity of the water level–storage relation in the real system. However, averaged over the total amount of considered events, the accuracy is sufficient because both over- and underestimations occur over the series of events, leading to a total volumetric error of 7%.

Apart from the switching frequency during dry weather (not relevant for this study) the pump behaviour is accurately simulated. Both switch-on time (maximum observed deviation over all PSs is 22 min) and maximum pumping rate (maximum observed deviation over all PSs is 6%) during wet weather match well with reality. In conclusion, it can be stated that using the synthetic inflow rate \( Q_{in} \) leads to a realistic characterisation of wet weather events. The approach is therefore deemed suitable for the objectives of this study as it allows for the comparison of scenarios (such as modified pump operation) that affect the water level \( h \). It might be an interesting alternative for WWTP modellers when working on integrated systems where system modifications do not affect the further upstream parts of the catchment and rainfall-runoff processes do not need to be considered explicitly.

### Table 1

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial value</th>
<th>Relative uncertainty range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weir crest level (m)</td>
<td>7.68, 8.35, 8.7, 7.99, 8.5</td>
<td>10</td>
</tr>
<tr>
<td>Weir crest length (m)</td>
<td>5, 3, 4, 6, 4</td>
<td>10</td>
</tr>
<tr>
<td>Weir discharge coefficient (−)</td>
<td>0.66, 0.66, 0.66, 0.66, 0.66</td>
<td>20</td>
</tr>
<tr>
<td>Pump capacity in (l/s)</td>
<td>67, 33, 79, 59, 189</td>
<td>20</td>
</tr>
<tr>
<td>Switch-on delay (s)</td>
<td>30, 30, 120, 30, 30</td>
<td>70</td>
</tr>
<tr>
<td>Switch-off delay (s)</td>
<td>30, 30, 30, 30, 30</td>
<td>70</td>
</tr>
<tr>
<td>Inflow estimation ratio PS Hvl (−)</td>
<td>2.5</td>
<td>20</td>
</tr>
<tr>
<td>Level sensor error factor (−)</td>
<td>1, 1, 1, 1, 1</td>
<td>20</td>
</tr>
<tr>
<td>Storage curve error factor (−)</td>
<td>1, 1, 1, 1, 1</td>
<td>20</td>
</tr>
<tr>
<td>Max WWTP capacity (l/s)</td>
<td>375</td>
<td>20</td>
</tr>
</tbody>
</table>

Figure 3 | (a) Monitored and modelled pump flows for one event of the base scenario; \( \text{sim}(f(h)) \): modelled based on water level, \( \text{sim}(f(Q_{in})) \): modelled based on synthetic inflow rate; (b) modelled total WWTP influent flow for local control (top) and global control (bottom); raw data and 1 h moving average.
Pros and cons of the control strategies

As shown in Figure 3(b), both local and global control strategies can reliably ensure the flow limitation to the WWTP, but the global control resulted in a more stable influent flow rate. The local control can ensure an average maximum inflow, but the influent flow during one cycle can vary strongly. A second advantage of the global control algorithm is its possibility to directly alter the desired maximum capacity (now set to 375 l/s). This enables operators to dynamically adapt the setpoint according to the operational state of the WWTP as, for example, suggested by Seggelke et al. (2003). Adaptation of this setpoint is less straightforward for the local control algorithm as all time slots have to be recalculated.

The main advantage of global control over local control can be found in the total CSO volume calculated for the different scenarios. While the application of global control increases the expected CSO volume by 23% with regard to the uncontrolled reference scenario (leading to non-compliance with the WWTP effluent standard), the best solution found for local control based on time slots resulted in a CSO volume that was almost twice as high, i.e. 40%. None of the proposed control strategies was able to improve the current situation, which represents a scenario that exceeds the maximum WWTP capacity, without measures taken upstream to reduce the CSO volumes at the PS.

Based on the modelling results, the developed RTC algorithm has been implemented and successfully operated on full scale since 2013. In its design, the control algorithm is applicable to any WWTP independent of the number of PSs (or regular screw pumps), or any combination of parallel PSs with a downstream flow limitation. It offers a solution for quick implementation and robust operation to any wastewater operator confronted with comparable challenges. Investigations for the deployment in another case study are ongoing.

Model and control parameter sensitivity

Morris screening and eFAST show very similar results for the parameter ranking: the eight most influential parameters are ranked identically by both methods. The parameters with the highest influence on total CSO volume are the WWTP capacity and the flow estimation factor for PS Heuvelhoekstraat. This finding could be anticipated, as these two parameters are intuitively the most relevant boundaries for the systems volume balance. Inaccuracies in the monitored water levels or the storage curves did not show a high impact on the model result.

Structural information of the PSs, on the other hand, showed to be of high importance: the pump capacities of PS Leihoeckstraat and Heuvelhoekstraat are ranked third and fourth, while PS Neerhoek and Leiestraat only rank 11th and 12th, respectively. With Leihoeckstraat being the second smallest PS, these results were unexpected. Weir crest levels showed an even higher spread again, with PS Leihoeckstraat and Heuvelhoekstraat ranked much higher than the other locations. These results show that it is difficult to draw generic conclusions about parameter sensitivities, especially for larger models. However, the overall results indicate that the efficiency of the control algorithms is dominated by its parameterisation and available pump capacity, while monitoring errors play a minor role.

Interestingly, the Morris screening showed considerably better convergence after 8,800 simulation runs than eFAST after 55,900 runs. These outcomes are in line with findings from Gamerith et al. (2013), who report that Morris screening returned stable results for comparably small numbers of simulations. Vanrolleghem et al. (2015), however, found Morris not to show satisfying convergence. This might indicate that Morris screening is applicable for models of simple structure, as the one considered here, while eFAST appears to be the apt choice for more complex model structures.

CONCLUSIONS

A simple and robust RTC algorithm has been developed based on a limited amount of data for five PSs that drain the rainfall-runoff from a rather small catchment directly into a WWTP. As the aim of the study was to limit the inflow to the WWTP during rain events, focus was on control of the influent pumps. The developed algorithm is applicable to other systems where the WWTP influent (or other flow limiting locations) is controlled by a number of pumps and where system performance evaluation based on CSO volume is possible.

Simulation results showed that, compared to the reference scenario where no control is implemented, the increase of spill flows could be kept at an acceptable level while limiting the total influent flow to the WWTP to its capacity, without necessitating further mitigation measures in the sewer network. Global control using a ranked filling degree performed considerably better than time-slot-based local control.
Morris screening and eFAST ranked the eight most influential parameters identically. For this specific model, Morris screening shows to be advantageous as it delivers better convergence after a smaller amount of simulations than eFAST. The results, however, also indicated that generic conclusions on the influence of parameters cannot be drawn, but that knowledge of the model structure plays a very important role.

REFERENCES


First received 11 June 2015; accepted in revised form 8 December 2015. Available online 23 December 2015.