Global predictive real-time control of Quebec Urban Community’s westerly sewer network

M. Pleau*, G. Pelletier*, H. Colas*, P. Lavallée* and R. Bonin**
*BPR Groupe-conseil, 4655, blvd Wilfrid-Hamel, Quebec, Quebec, G1P 2J7, Canada
**Quebec Urban Community, 399, St-Joseph E., Quebec, Quebec, G1K 8E2, Canada

Abstract Quebec Urban Community (QUC) has selected Global Predictive Real-Time Control (GP-RTC) as the most efficient approach to achieve environmental objectives defined by the Ministry of Environment. QUC wants to reduce combined sewer overflows (CSOs) frequency to the St Lawrence river to two events per summer period in order to reclaim the use of Jacques-Cartier Beach for recreational activities and sports of primary contact. QUC’s control scheme is based on the Certainty Equivalent Control Open Loop Feedback (CEOLF) strategy which permits one to introduce, at each control period, updated measurements and meteorological predictions. A non-linear programming package is used to find the flow set points that minimise a multi-objective (cost) function, subjected to linear equality and inequality constraints representing the physical and operational constraints on the sewer network. Implementation of GP-RTC on QUC’s westerly network was performed in the summer of 1999 and was operational by mid-August. Reductions in overflow volumes with GP-RTC compared to static control are attributed to the optimal use of two existing tunnels as retention facilities as well as the maximal use of the wastewater treatment plant (WWTP) capacity.

Keywords Combined sewer overflows; optimal performance; real-time control; sewer networks

Introduction

Deterioration of water quality in receiving water bodies due to combined and sanitary sewer overflows (CSOs and SSOs) is an issue that needs to be addressed by sewer agencies to comply with both environmental regulations and citizens’ increasing desire for healthy river environments. Environmental regulations are usually expressed in terms of frequency or total volume of overflows for the period when recreational activities, such as sports of primary contact (e.g. swimming) and secondary contact (e.g. canoeing) can be practised. Technical solutions to reduce CSOs and SSOs include the increase of retention volume and wastewater treatment plant (WWTP) capacity, installation of separate sewer systems and stormwater retention ponds, and implementation of Real-Time Control (RTC). These high impact solutions can be paired with lower impact solutions such as reducing water use, decreasing impermeable surface areas, controlling the sewer system inlet flows and treating wastewater on-site.

For the Quebec Urban Community’s (QUC) sewer network, environmental objectives defined by the Ministry of Environment are as follows: two overflows to the St Lawrence river are permitted during a summer period (May 15th to September 15th), while four are permitted to the St Charles river. A study realised in the early stages of QUC’s CSO plan project has shown that, in order to comply with these objectives, 150 M$ would need to be invested in the construction of retention facilities and implementation of Global Predictive Real-Time Control (GP-RTC). However, it was decided that only GP-RTC would be implemented on the westerly network in Phase I (summer of 1999). The cost of this phase was estimated to 4.6 M$ and was expected to reduce CSO volumes by 50% and frequency by 60% on an annual basis. This project was partially funded by governmental programs PAEQ (Programme d’assainissement des eaux du Québec) and TICQ-1997 (Travaux d’infrastructures Canada-Québec).
The GP-RTC scheme implemented on QUC’s westerly network aims at reducing CSOs by using the sewer system transport, retention and treatment capacities to their full potential. A sub-optimal control strategy referred as Certainty Equivalent Open Loop Feedback (CEOLF) scheme in the literature (Papageorgiou, 1988) is used at the central station to compute flow set points needed to achieve the desired performance. Flow set points are applied at local stations located at strategic points on the sewer network. All major difficulties related to GP-RTC, which include computing time, robustness, applicability and adaptability, were addressed in the control scheme implemented at the central station and local stations.

QUC’s GP-RTC scheme

QUC’s westerly network GP-RTC scheme is based on a two-level architecture (Figure 1). The programmable logical controllers (PLCs), located at the local stations are responsible for two-way communications with the Central Station and the personal computers (PCs), and one-way communications from measurement devices and to the actuators controlling the sluice gates. Measurements taken from sensors are read by the PLCs, which send them to the PCs for validation. The PCs return the validation status of the data (valid or invalid) to the PLCs, which send both measurements and validation status to the Central Station. The PLCs compute the maximal permissible flows to be used by the optimisation routine and manage alarms (intrusion, maintenance, malfunction, etc.), which are also sent to the Central Station. The Central Station sends optimal flow set points for the next hour, date and time, and rainfall status to the PLCs. The PLCs compute gate openings needed to respect the flow set points and send them to the actuators. All data needed to manage the system are stored in a relational database located at the Central Station.

The Central Station

The Central Station performs two main tasks: computation of flow set points and rainfall status. There are two possible rainfall statuses, “Wet” or “Dry”. Rainfall status is determined from five criteria that depend on the meteorological forecast, rain gauge measurements, volumes accumulated in retention facilities, flow rates at control sites and WWTP capacity. Rainfall status is used by the local stations to determine the local mode of opera-
tion and by the Central Station to manage simulation scenarios. Flow set points are the main output from the optimisation routine. Figure 2 presents a general overview of the components of QUC’s CEOLF control scheme used to compute these flow set points. The approach consists of dividing the sewer network into two parts: the uncontrolled and the controlled sections. The uncontrolled section is independent from the manipulated flows and, therefore, it is not defined in the optimisation problem. The CEOLF scheme solves, at each control period, a deterministic optimisation problem over a control horizon (i.e. a forecasting window), with updated measurements and rainfall predictions. For QUC’s westerly network, a 5-minute control period and a 2-hour control horizon were chosen to reduce annual CSO volumes by 50% and frequency by 60%.

The CEOLF inputs. There are eight types of inputs in the CEOLF control scheme designed for QUC (shadowed boxes in Figure 2), of which three are direct inputs to the optimisation algorithm. Present and future rainfall intensities are obtained from a meteorological forecasting model using radar images (1-hour horizon), from rain gauge measurements and from extrapolation techniques. Since predictions are needed for the entire control horizon, rainfall intensities from the forecasting model are extrapolated for the second hour by fitting future intensities on a Gaussian curve based on past rainfall intensities. If the forecasting model is inactive, present rainfall intensities are obtained from 15 rain gauges spread on QUC’s territory. Flows and volumes are supplied to the optimisation algorithm either by the local stations or by a reference hydraulic model. Flow and volume data are therefore always available at any point on the network.

Gate control status refers to the type of control at a local station, namely “Global”, “Local” or “Static”. Global control means that flow set points are calculated by the optimisation routine, local control means that flow set points are determined locally by a set of rules imbedded in the PLC and static control means that the control device has a fixed opening. An important characteristic of the control scheme developed for QUC is that all three types of control are taken into account in the global evaluation of flow set points, since they all have an influence on the hydraulic behaviour of the controlled section of the sewer network. Maximal flow rates are calculated to impose a realistic upper limit on the flow set points. All gate control status and maximal flow set points are determined at the local stations. If the status of a gate changes during a rainfall event, the CEOLF scheme reacts immediately to the new system behaviour and computes flow set points according to the modified control status.

The optimisation routine needs to know the primary, secondary and hydraulic WWTP capacities. Primary treatment capacity and secondary treatment capacity are supplied by the WWTP itself, while the hydraulic capacity is a function of the St Lawrence river tide where the diffuser is located. The optimisation routine uses the secondary treatment capacity (usually 3.66 m$^3$/s) until overflow becomes inevitable, at what point it switches to the lowest value between the primary treatment capacity (usually 5.85 m$^3$/s) and the hydraulic capacity. Local control objective weights are used to favour some overflow sites and to favour the dewatering of some retention facilities. Downstream overflow sites are favoured since there is less uncertainty on downstream flow rates due to the presence of measurement devices. At QUC, there are two existing retention tunnels, with total volumes (including the volume occupied by flow) of 15,600 m$^3$ and 16,100 m$^3$. Their combined retention capacity is approximately 15,000 m$^3$. Dewatering of the upstream tunnel (“Versant-Sud” tunnel) is favoured since its maximal dewatering rate is low. Moreover, this tunnel is the first to be filled-up when back-to-back events occur. Finally, the last input is concerned with flow constraints. At QUC, no surcharge flow is allowed in the controlled section of the network.
The CEOLF hydrologic and hydraulic models. There are four simulation models used in QUC’s GP-RTC scheme: a non-linear hydrologic model and a non-linear hydraulic model for the uncontrolled section, and a non-linear hydraulic model and a linear hydraulic model for the controlled section (Figure 2). The uncontrolled section of the network is simulated over the entire control horizon in order to provide the present and future inlet flows to the controlled network. Hydraulic simulations in the controlled section are performed by a non-linear reference model (Mailhot et al., 1998) and a linear Auto-Regressive Moving-Average (ARMA) model. The reference model is a non-linear hydraulic model used to predict future flow rates in the sewer network and provide estimates of flows where there are no measurement devices or no valid measurements. The ARMA model is linked with a Kalman filter to increase the robustness of the CEOLF scheme, with respect to modelling uncertainties and unknown perturbations (Gelormino and Ricker, 1994).

The CEOLF optimisation algorithm. The optimisation problem consists of finding the flow set points that minimise the value of a multi-objective (cost) function, with respect to physical and operational constraints. For QUC’s westerly network, all constraints are linear in order to reduce computing time. The system’s non-linear behaviours are defined instead in the multi-objective function. The control objectives are, in order of priority, to minimise
overflows, to maximise the use of WWTP capacity, to minimise accumulated volumes and to minimise set point variations. Along with these global control objectives and local weights defined in the objective function, an uncertainty factor is associated with the optimisation variables to take into account the fact that predictions in the far future are more uncertain than in the near future (Pleau et al., 1996). The linear equality constraints are used to define the relationships between the optimisation variables. In particular, the set of equality constraints include the linear hydraulic ARMA model. The inequality constraints are used to set physical and operating boundaries. They limit accumulated volumes in the tunnels and flow rates below the pipes’ hydraulic capacities. They also constrain flow set points below maximal values computed at the local sites and limit flows conveyed to the WWTP.

The optimisation routine is implemented with a non null solution space algorithm to guarantee that flow set points exist for each control period. The algorithm can relax some flow constraints when surcharge flows cannot be avoided and can add fictitious water in retention facilities when corrections from Kalman filtering would produce negative volumes of water. Overall, the optimisation problem defined for the controlled section of QUC’s westerly network comprises 1380 constraints and 1196 variables. The optimisation problem is solved by a non-linear programming algorithm (MINOS®). Three special GPR-RTC features are implemented into the CEOLF control scheme to increase its adaptability and its robustness. These are the synchronisation feature, the on-line calibration module of the ARMA model and the hot start-up module. The synchronisation feature guarantees that the CEOLF simulation clock is not desynchronised with the real-time clock during a long period of time. If for exceptional reasons, the execution of the control loop takes longer than the time allowed for real-time control, the CEOLF scheme is executed without calling the optimisation routine, as long as the control scheme time has not reach real time. On-line calibration of the ARMA model is made possible by using a robust recursive least squares algorithm. Finally, the hot start-up module ensures that simulated variables will reach real hydraulic conditions rapidly when the system is restarted after a shutdown.

Local stations
There are three types of local stations: (1) rain gauge stations to obtain rainfall intensities, (2) monitoring stations to obtain flow, velocity, water level and gate opening measurements, and (3) control stations to compute gate positions, control sluice gates and collect various field data. Local stations communicate between themselves and with the Central Station via either radio links or telephone lines. The operating mode of a local control station depends on several variables including rainfall status, validity status of measurements, communication status and control mode imposed by the sewer network operator. It should be noted that, when the Central Station loses communication with a local station, the rainfall status is automatically switched to “Wet” for security.

Normal operating mode during dry weather. This operating mode is designed to avoid overflows during dry weather, to maximise the useful life of all mechanical devices and to minimise energy costs. In this mode, gates are fully opened at control sites with retention facilities and positioned to intercept the maximal dry flow at control sites without retention facilities. For the latter, gates are fully opened 1 minute every 12 hours in order to flush debris and floatable materials. A local station must satisfy three criteria in order to operate under this mode: (1) rainfall status must be “Dry”, (2) control decisions must be taken by the CEOLF scheme (as opposed to manual control), and (3) gates must be operational.

Normal operating mode during wet weather. The normal operating mode during wet weather is designed to achieve QUC’s control objectives. Gates are positioned according to
the flow set points computed by the optimisation algorithm. Flow set points are converted to gate positions by an adaptive integrative regulator with dead band and clamps. The dead band is defined in terms of the minimal deviation between set points and intercepted flows that justifies gate movement. This feature is designed to avoid excessive start-up/shutdown of motors resulting from measurement noise. The lower clamp is static and defined as the minimal gate position. By opposition, the upper clamp is dynamic and ensures that the difference between the upstream water level and the gate opening remains larger than a specified positive value. This minimal difference guarantees that the intercepted flow used by the adaptive integrative regulator is accurately assessed using an energy balance equation. When the water head in the vicinity of a gate is significant, the kinetic term of the energy balance equation becomes small compared to the potential energy term and a sound estimate of the intercepted flow can be obtained. If the intercepted flow cannot be computed using the energy balance equation due to measurement problems, the intercepted flow is estimated using a mass balance equation for sites without retention. Upstream flow is measured by a flowmeter, while overflow is determined from an overflow structure standard equation. A local station must satisfy five criteria in order to operate under this mode: (1) rainfall status must be “Wet”, (2) control decisions must be taken by the CEOLF scheme, (3) gates must be operational, (4) measurements needed to calculate the intercepted flow must be available and valid, and (5) the last successful communication with the Central Station must have been made less than 1 hour ago.

**Downgraded operating mode.** When conditions for normal operating mode are not met, local stations switch to the downgraded operating mode. The primary goal of the downgraded mode is to protect citizens from flooding (basement or street) and the network from surcharges. A secondary goal is to achieve the maximal system performance with the available information with regard to the control objectives. Gate positioning depends on the nature of the problem. For example, when there is no communication between the Central Station and a local station for more than 1 hour, some local stations take advantage of their ability to communicate with other local stations to exchange information and command proportional filling of the two tunnels.

**Specific tasks.** Local control stations are responsible for the determination of the gate control status, the computation of maximal flow rates and data validation. Gate control status and maximal flow rates are inputs to the CEOLF control scheme and were discussed earlier.

**Table 1** Overflow volumes (m³) for four rainfall events under GP-RTC and static control

<table>
<thead>
<tr>
<th>Site</th>
<th>13/08/99</th>
<th>17/08/99</th>
<th>06/10/99</th>
<th>09/10/99</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global</td>
<td>Static</td>
<td>Global</td>
<td>Static</td>
</tr>
<tr>
<td>Suète</td>
<td>3136</td>
<td>5278</td>
<td>1066</td>
<td>1490</td>
</tr>
<tr>
<td>Jones</td>
<td>2128</td>
<td>3276</td>
<td>264</td>
<td>802</td>
</tr>
<tr>
<td>Dijon</td>
<td>1190</td>
<td>2832</td>
<td>214</td>
<td>567</td>
</tr>
<tr>
<td>Junction</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WWTP</td>
<td>0</td>
<td>3657</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>6454</td>
<td>15,043</td>
<td>1544</td>
<td>2858</td>
</tr>
<tr>
<td>% red. w</td>
<td>57%</td>
<td>46%</td>
<td>93%</td>
<td>89%</td>
</tr>
<tr>
<td>% red. w/o</td>
<td>43%</td>
<td>46%</td>
<td>95%</td>
<td>79%</td>
</tr>
</tbody>
</table>
Data validation is used to tag data with a validation status depending on predetermined criteria. Validation can be performed using threshold values on raw measurements or their variance and by using a statistical validation algorithm. This last method is based on statistical process control and probability theory. The idea is to take advantage of the redundancy of some measurement devices to validate by comparison. Validation status is used at local stations in the computation of intercepted flow, overflow, accumulated volumes and in the determination of the operating mode. At the Central Station, only valid data are inputted in the control scheme algorithm.

QUC’s GP-RTC performance

QUC’s control scheme was implemented during the summer of 1999. GP-RTC of wet-weather events became effective in mid-August. The control scheme was designed to be in operation during the summer period (May 15th to September 15th), but since it did not rain much during the summer, the system was kept in operation until November for testing and fine-tuning. Table 1 presents overflow volumes for four rainfall events under GP-RTC and static control at 5 major overflow sites located in the controlled section of the westerly sewer network. Three sites with overflow structures are controlled: Jones, Dijon and the Junction chamber. Suète is a site with an orifice, which was modified during the GP-RTC implementation to intercept a higher flow (from 0.118 m³/s to 0.31 m³/s) as part of the global control scheme. Total water depth \(H\), maximal 5-minute rainfall intensity \(I\) and duration of event \(D\) at a specific rain gauge (Ste-Geneviève) are indicated for each rainfall event.

The WWTP can fully treat (primary and secondary treatment) wastewater reaching the plant at a flow rate up to 3.66 m³/s. The fraction of flow above 3.66 m³/s up to the hydraulic capacity of the diffuser undergoes primary treatment only. The fraction of flow above the hydraulic capacity of the diffuser is overflowed directly to the St Lawrence river. All “overflows” at the WWTP shown in Table 1 are actually wastewater volumes that have undergone primary treatment. Overflows under “Global” are obtained using input data collected by measurement devices, whereas overflows under “Static” are obtained from simulation results made with a configuration of the network before the implementation of GP-RTC. The last two rows present the percentage of reduction in overflow volumes obtained with GP-RTC compared to static control, including and excluding the WWTP.

In general, overflows occur with high intensity rainfall events rather than with large water depth events. As seen in Table 1, the lowest reduction in overflow volume from GP-RTC occurs for the event with the largest maximal 5-minute rainfall intensity. Overflows occur at Suète with both GP-RTC and static control due to the fact that the permissible intercepted flow is lower than the upstream flow reaching the control site. The permissible intercepted flow in static control is equal to the maximal dry weather flow rate, the fraction of flow exceeding this value is overflowed. Overflows at Jones with GP-RTC are due, in part, to the fact that avoiding surcharges has a higher priority than minimising overflows. During the two August events, the control scheme predicts that the lowest pipe capacity in the interceptor is going to be surpassed, therefore the gate is moved so that the intercepted flow at the control site, when combined with appropriate downstream flows, does not create a surcharge in this pipe. Overflows at Jones could be reduced (even eliminated) if a recalibration of the hydrologic model for this watershed was done with recent data.

The two overflows at Dijon were generated to avoid surcharge flow in a pipe section located between the control gate and the Versant-Sud tunnel. This flow constraint was later found to be unjustified and was removed. Overflows occur at Dijon only when the Versant-Sud tunnel is full. Both October events generate wastewater volumes that do not undergo secondary treatment at the WWTP. This is due to the fact that the WWTP capacity is not
known in advance. Since one of the control objectives is to maximise the use of the WWTP capacity, when flow entering the WWTP matches its capacity and the capacity drops abruptly, overflow can occur during a brief period (less than 5 minutes).

Summary and conclusion

QUC wants to reduce the frequency of CSOs to two overflows to the St Lawrence river per summer period. With such objectives, citizens could enjoy Jacques-Cartier Beach for recreational activities during the summer. To control CSOs on QUC’s westerly network, preliminary engineering studies concluded that global predictive RTC was a cost effective approach. QUC’S GP-RTC scheme is based on a CEOLF scheme in which the optimisation problem is solved using a non-linear programming algorithm (MINOS®). Implementation was planned cautiously in order to obtain real performances as close as possible to those predicted during simulation studies. In particular, robustness was addressed at the local stations by implementing data validation, by designing a downgraded operating mode and by using an adaptive integrative regulator. At the Central Station, robustness of the CEOLF scheme was increased by the use of Kalman filtering, by programming a non null solution space algorithm and by designing a hot start-up module and a synchronisation module.

A linear hydraulic ARMA model is used in the definition of the equality constraints of the optimisation problem to reduce computing time. Non-linear hydraulic behaviours were defined in the multi-objective function in order to reduce the impact of the use of a linear hydraulic model on the system performance. The dynamic nature of sewer networks raises concern about the GP-RTC scheme’s adaptability. To guarantee that results obtained for the 1999 summer events would be reproducible in the future, a module that automatically rebuilds the non-linear and linear models used in the CEOLF scheme was implemented. Finally, an on-line calibration algorithm that computes, at each control period, a new set of parameters for the ARMA model was implemented. Although it is not possible to analyse the system performance in terms of overflow frequency and volume over the entire 1999 summer period, it is clear that the control scheme is taking the correct decisions considering the inputs provided. Reductions in overflow volumes from GP-RTC compared to static control are very high for autumn rainfall events and are expected to be around 60% for high-intensity rainfall events.

References


