

## Une méthodologie pour concevoir des règles pour le contrôle en temps réel des systèmes de drainage urbain à l'aide d'arbres de décision par régression

### A methodology for designing rules for real-time control of urban drainage systems using regression decision trees

Juan Esteban Ossa Ossa<sup>1</sup>, Sophie Duchesne<sup>1</sup>, Geneviève Pelletier<sup>2</sup>

1 Institut National de la recherche scientifique, [Juan\\_Esteban.Ossa\\_Ossa@inrs.ca](mailto:Juan_Esteban.Ossa_Ossa@inrs.ca) et [Sophie.Duchesne@inrs.ca](mailto:Sophie.Duchesne@inrs.ca)

2 Université Laval, [genevieve.pelletier@gci.ulaval.ca](mailto:genevieve.pelletier@gci.ulaval.ca)

#### RÉSUMÉ

Les stratégies de contrôle en temps réel peuvent être classées en deux catégories principales, soit les approches heuristiques et les stratégies d'optimisation. Ces dernières sont très exigeantes en termes de calculs et peuvent être très complexes à appliquer, ce qui pose des défis pour leur mise en œuvre opérationnelle. C'est pourquoi les approches heuristiques sont les plus souvent appliquées, bien qu'il n'existe pas de méthodologie formelle permettant de formuler de façon optimale leurs règles. Cette étude présente un cadre permettant de formuler de règles pour contrôler les niveaux d'eau dans un bassin de rétention d'eaux à l'aide d'une approche de contrôle basée sur des règles intégrant des prévisions de précipitations. La méthode utilise l'apprentissage automatique par arbres de décision par régression pour apprendre la logique de contrôle à partir du comportement optimal du système, tel que défini par une approche prédictive basée sur l'optimisation. Le modèle a été entraîné en appliquant un contrôle prédictif optimal sur 10 événements pluvieux. L'application de la méthode à une étude de cas où l'objectif de contrôle était de maintenir le niveau d'eau dans le bassin entre 15,1 m et 15,4 m, afin d'éviter les débordements et les dommages aux infrastructures, a démontré que les règles dérivées d'arbres de décision surpassaient le contrôle statique et atteignaient des performances comparables à celles obtenues grâce au contrôle prédictif. En outre, cette méthode permet d'extraire des règles de décisions dans des formats facilement compréhensibles, tels que des règles SI-ALORS, à partir des résultats du contrôle prédictif optimal.

#### ABSTRACT

Real time control strategies can be classified into heuristic and optimization-based approaches, with the latter being very demanding in terms of computational requirements and being operationally complex, posing challenges for their field implementation. As a result, heuristic approaches are being applied more widely at the operational level, even if there is a lack of methodologies for formulating optimal rule sets for these approaches. This study presents a framework for formulating a set of rules to control water levels in a stormwater retention basin with a rule-based control approach based on rainfall forecasts. The method uses machine learning regression decision trees to learn the control logic from the optimal behavior of the system, as defined using optimization-based model predictive control. The model was trained by applying model predictive control over 10 historical rainfall events that occurred between 2022 and 2024. Applied to a case study where the control objective was to maintain water elevation into the basin between 15.1 m and 15.4 m, to avoid overflows and damage to infrastructure, the method demonstrated that the derived rules outperformed static control and achieved performance comparable to that obtained through the direct application of model predictive control. Furthermore, the method was effective in extracting control decisions in understandable formats, such as IF-THEN rules, from the behavior of the predictive control model.

#### MOTS CLÉS

Real-time control, Urban drainage systems, Stormwater retention basin, Rule based control, Machine learning.

---

## 1 INTRODUCTION

Many studies showed the effectiveness of real-time control (RTC) to improve urban drainage systems (UDS) performance under both extreme and normal operating conditions (Shishegar et al., 2019; Jean et al., 2024). RTC strategies in UDS can be classified into heuristic control and optimization-based control. Heuristic control, particularly rule-based control (RBC), is based on predefined offline rules obtained from expert knowledge, operational experience, or historical data. In contrast, optimization-based control, such as model predictive control (MPC), employs mathematical models of system dynamics to identify optimal control actions in real-time. Research has shown that, while heuristic and optimization-based methods may perform similarly for small or simplified systems, MPC outperforms RBC in larger or complex UDS (Kroll et al., 2018; van der Werf et al., 2022). However, the practical implementation of MPC remains limited due to its computational demands and the complexity it introduces for system operators (Pei et al., 2024). As a result, RBC continues to lead the operational management of UDS due to its simplicity, transparency, low computational requirements, and ease of implementation in real-world settings.

The development of RBC strategies is still difficult despite their widespread use. The design of heuristic rules depends on understanding system behavior, yet this knowledge is difficult to obtain and poorly documented (Garcia et al., 2015; Kroll et al., 2018). Existing studies employing RBC rarely describe their rule formulation process explicitly, instead citing trial-and-error methods (Meneses et al., 2018), industry standards (Bowes et al., 2022), or engineering knowledge (Bilodeau et al., 2018). For this reason, some authors (Garcia et al., 2015; Kroll et al., 2018) have identified the lack of a standardized methodology for the formulation of standards as a critical gap in this field, which persists today. Kroll et al. (2018) proposed a procedure for designing RBC using equal-filling degree concepts for ‘on/off’ control and showed comparable performance to MPC for combined sewer networks. More recent work has explored data-driven approaches to improve heuristic performance in other domains since advances in machine learning (ML) and hyper heuristics offer new opportunities for automated heuristic design (Bouazza, 2024). In particular, decision trees (DT) can extract interpretable rules from optimal decision patterns. Building on these advances, this article proposes a framework that uses ML regression DTs to derive refined RBC rules (i.e., different variations in gate opening between 0% and 100%, not just “on/off”) from optimal decision control settings. The methodology extracts control logic based on rainfall forecasts and optimal system responses, resulting in transparent and implementable rules that retain the practical advantages of RBC.

## 2 METHODOLOGY

### 2.1 Case study

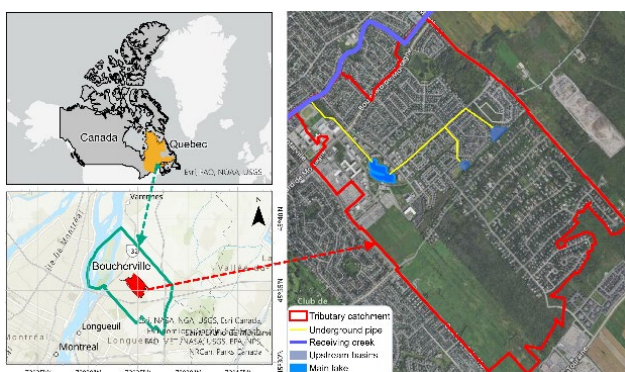


Figure 1. Case study location

The case study is a stormwater retention basin (urban lake) located in the city center of Boucherville, Quebec, Canada. Figure 1 shows the study site location and the contributing drainage area. The lake collects runoff from a 360-ha urban catchment area and serves as a recreational site and a flood management facility, requiring water level management to prevent overflow and protect surrounding infrastructure. Operational and physical constraints require water levels to be maintained between 14.8 m and 15.4 m, with a target operational level of 15.1 m. Exceeding the upper limit poses a risk of flooding public spaces and adjacent urban infrastructure, while

maintaining levels below or close to the operational target optimizes storage capacity for incoming runoff. To achieve these objectives, a downward opening gate is simulated at the weir located in the lake outlet control chamber, allowing water levels to be regulated through RTC. The simulation of the sluice gate and the implementation of RTC aim to maintain water levels within the specified range, while reducing peak discharge flows to the downstream watercourse.

## 2.2 RBC strategy development

The methodology for developing the control rules for the RBC strategy consists of the two steps briefly described below.

### 2.2.1 Hydrological model and rainfall data analysis

The first step consists of characterizing lake behavior under different hydrological conditions using a hydraulic-hydrological model of the study area, calibrated and validated with an average Nash-Sutcliffe Efficiency (NSE) of 0.75 for inflows and outflows. The model, developed in PCSWMM (CHI, 2025) and runs through PySWMM (McDonnell et al., 2020), provides information on the system behavior. Historical rainfall data from 2007 to 2025 were obtained from the Montreal's rain gauge network, Environment Canada, and supplemented with measurements from a rain gauge installed next to the lake in 2024 and 2025. This dataset was processed to identify individual rainfall events using a 6-hour dry period criterion between events, excluding events with an accumulated depth of less than 0.5 mm. For each identified event, the simulation period was defined to include 48 hours of antecedent conditions, consistent with the maximum forecast horizon provided by Environment Canada's High Resolution Deterministic Prediction System (HRDPS) (Environment and Climate Change Canada, 2025). The post-event simulation duration was adjusted based on event magnitude to ensure that the complete lake drawdown was captured. This preprocessing yielded 1158 rainfall events ranging from 0.50 mm to 131.25 mm in depth, providing a wide representation of the site's hydrological regime. Simulation results revealed some system weaknesses; for example, water levels reached 16.1 m during high-intensity rainfall events, which is 0.7 m higher than the operational upper limit of 15.4 m.

### 2.2.2 Control algorithm design

The procedure is based on generating optimal control patterns through learning from MPC simulations to operate the system in predictive mode. The MPC uses 48-hour rainfall forecasts from the HRDPS to optimize the gate setting at each 5-minute time interval, minimizing deviations from the 15.1 m operational level and penalizing excessive gate movements and outflow limit violations. The optimization problem is solved using convex optimization in python (CVXPY) with the operator splitting quadratic program (OSQP) solver over a 2-hour control horizon (24-time steps), considering forecasted rainfall, system states, and operational constraints. Ten representative rainfall events that occurred during the 2022-2024 period were simulated using PySWMM, generating time series of optimal gate positions, obtained from MPC. From these simulations, a supervised learning dataset was constructed with 17 features based on short-term rainfall forecasts (6-12 hours in advance), recent rainfall accumulation (1-24 hours ago), and current conditions (rainfall, lake level, outflow, inflow, and lake volume).

Next, a C4.5 DTs regression algorithm (Quinlan, 1993) was trained on the simulated dataset to extract interpretable IF-THEN control rules. This training was performed using event-based splitting (80% training, 20% testing), configured with a maximum depth of 8 levels (i.e., no branch of the decision tree could exceed eight successive splits), a minimum of 100 samples per split, and sample weighting that favored forecast-based decisions. The resulting decision tree was converted into executable IF-THEN rules that approximate MPC performance using only rainfall information, allowing real-time gate control without online optimization. In addition, the rules obtained were analyzed to remove redundancies.

## 3 RESULTS AND DISCUSSION

Applying the proposed methodology resulted in the decision tree shown in Figure 2, structured as IF-THEN rules. The importance analysis of the features extracted from the system (based on the relative contribution of each input variable to the model's predictions) revealed that the 6-hour rainfall forecast is the main decision variable, confirming the reliability of the methodology in extracting the predictive control logic of the behavior considered as optimal.

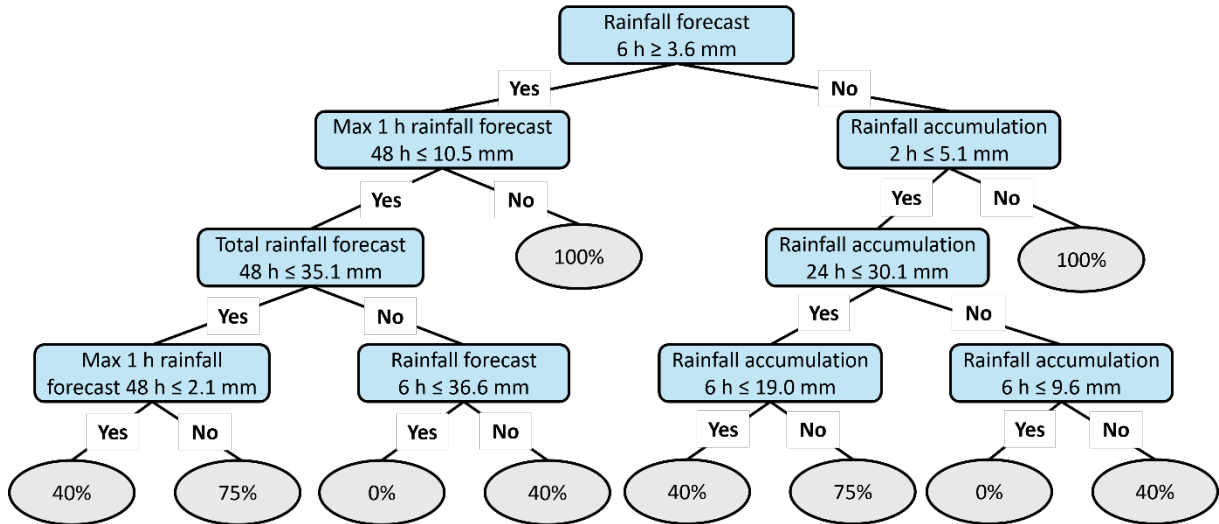


Figure 2. Decision tree. ‘Rainfall forecast X h’ refers to the predicted rainfall over the next X hours, ‘Rainfall accumulation X h’ refers to the total rainfall observed over the previous X hours, and ‘Max 1 h rainfall forecast X hours’ refers to the maximum predicted rainfall for 1 hour over the next 48 hours. Percentages in gray nodes represent the gate opening

Figure 3 shows a comparison of the performance of the three control strategies for a rainfall event of 45.25 mm. Results showed that the MPC and RBC derived from the ML behaved similarly, maintaining maximum water levels at 15.3 m, below the upper limit of 15.4 m and lower than in the baseline scenario. Under baseline conditions (constant gate level), the water level reached 15.5 m, exceeding the maximum limit and posing flood risks to surrounding areas. In contrast, active control strategies successfully prevented overflow and optimized storage utilization. A notable advantage of the ML-derived RBC was observed in its gate operating pattern and reduced computational time, as it opened the gate when necessary and closed it quickly when water levels approached toward the operational level of 15.1 m, achieving this 8 times faster than the MPC. This behavior reduced unnecessary gate movements while still achieving control objectives. In addition, RBC and MPC strategies facilitated rapid recovery of operational levels after rainfall, which restored available storage capacity for subsequent storms.

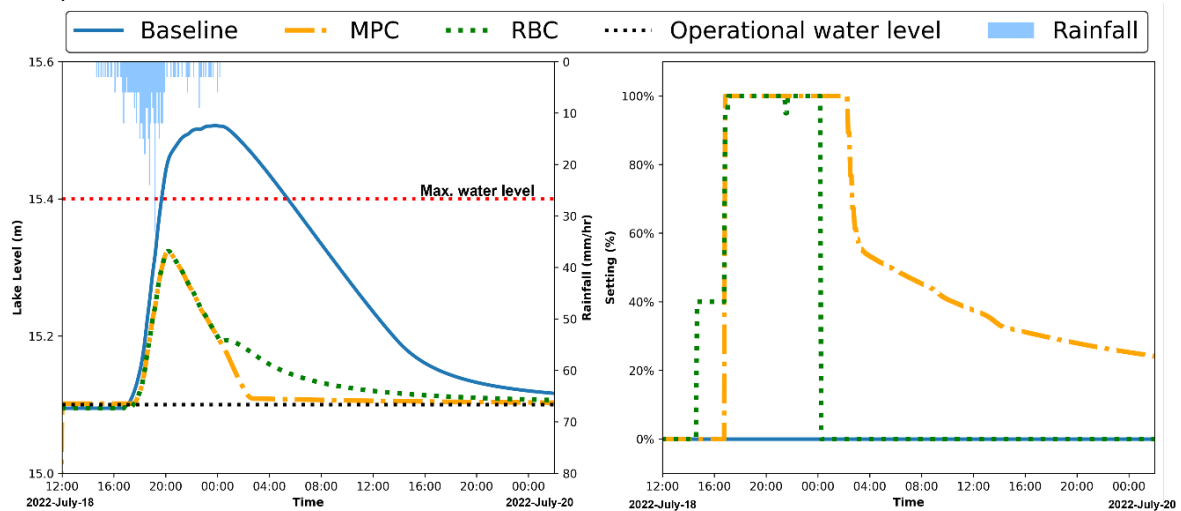


Figure 2. Lake behavior for passive system, MPC, and RBC.

These results validated the ability of the proposed framework to extract effective and transparent control rules from an optimal lake behavior pattern. The RBC derived from ML achieved optimal performance as well as offering computational simplicity and operational interpretability, addressing the main challenges for the implementation of advanced controls in UDS. Although the analysis focused on a single retention basin, the framework is also suitable for more complex drainage networks with multiple controlled facilities. In such systems, MPC can be simulated offline to generate optimal trajectories across basins and event types, and

decision trees can translate these patterns into interpretable rules that approximate system-level behavior. Extending this approach to systems where MPC is limited by real-time constraints would require incorporating additional state variables and more training scenarios to capture various operating conditions.

## 4 CONCLUSIONS AND FUTURE WORK

This study presents a framework for the design of RBC strategies by extraction of interpretable decision rules from MPC optimization using machine learning. The methodology generates IF-THEN rules that achieve performance comparable to MPC as well as maintaining operational interpretability. Its application to a stormwater retention basin demonstrates that the obtained rules maintain water levels within operational constraints, matching MPC performance without computational burden. The findings constitute an initial validation of the framework using simplified conditions. Future research will focus on improving the formulation of the MPC objective function to better capture operational priorities, developing surrogate models to accelerate MPC simulations and enable broader scenario testing, exploring alternative AI-based control strategies (e.g., Pei et al., 2024) as additional sources of optimal behavior for rule extraction, conducting performance comparisons among all derived strategies, and assessing robustness to rainfall forecast uncertainty.

## BIBLIOGRAPHIE

- Bilodeau, K., Pelletier, G., and Duchesne, S. (2018). *Real-time control of stormwater detention basins as an adaptation measure in mid-size cities*. *Urban Water Journal*, 15(9), 858–867.
- Bouazza, W. (2024). *Machine Learning-Based Hyper-Heuristics: A Clear Insight*. In Proceedings of the 2024 7th International Conference on Computational Intelligence and Intelligent Systems (pp. 29-37).
- Bowes, B. D., Wang, C., Ercan, M. B., Culver, T. B., Beling, P. A., and Goodall, J. L. (2022). *Reinforcement learning-based real-time control of coastal urban stormwater systems to mitigate flooding and improve water quality*. *Environmental Science: Water Research & Technology*, 8(10), 2065–2086.
- CHI. (2025). PSCSWMM Software (Version 7.7) <https://www.pcswmm.com/>
- Environment and Climate Change Canada (ECCC). (2025). *High Resolution Deterministic Prediction System (HRDPS) - Continental (Version 7.0)*. Government of Canada; Meteorological Service of Canada (MSC). <https://open.canada.ca/data/en/dataset/5b401fa0-6c29-57f0-b3d5-749f301d829d>
- García, L., Barreiro-Gomez, J., Escobar, E., Téllez, D., Quijano, N., & Ocampo-Martinez, C. (2015). *Modeling and real-time control of urban drainage systems: A review*. *Advances in Water Resources*, 85, 120–132.
- Jean, M.-È., Morin, C., Ossa Ossa, J. E., Duchesne, S., Pelletier, G., and Pleau, M. (2024). *Optimal distribution of green and grey infrastructures coupled with real time control of the sewer for combined sewer overflows control as an adaptation measure to climate change*. *Urban Water Journal*, 21(4), 419–435.
- Kroll, S., Weemaes, M., Van Impe, J., and Willems, P. (2018). *A methodology for the design of RTC strategies for combined sewer networks*. *Water*, 10(11), 1675.
- McDonnell, B., Ratliff, K., Tryby, M., Wu, J., and Mullapudi, A. (2020). *PySWMM: The Python Interface to Stormwater Management Model (SWMM)*. *Journal of Open Source Software*, 5(52), 02292
- Meneses, E. J., Gaussens, M., Jakobsen, C., Mikkelsen, P. S., Grum, M., and Vezzaro, L. (2018). *Coordinating Rule-Based and System-Wide Model Predictive Control Strategies to Reduce Storage Expansion of Combined Urban Drainage Systems: The Case Study of Lundtofte, Denmark*. *Water*, 10(1), 76.
- Pei, S., Hoang, L., Fu, G., and Butler, D. (2024). *Real-time control of urban drainage systems using neuro-evolution*. *Journal of Environmental Management*, 360, 121133.
- Quinlan, J. R. (1993). *C4. 5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Shishegar, S., Duchesne, S., and Pelletier, G. (2019). *An integrated optimization and rule-based approach for predictive real time control of urban stormwater management systems*. *Journal of Hydrology*, 577, 124000.
- van der Werf, J. A., Kapelan, Z., and Langeveld, J. (2022). *Towards the long term implementation of real time control of combined sewer systems: A review of performance and influencing factors*. *Wat. Sci. & Tech.*, 85(4), 1295–1320.