Journal of Hydrology 637 (2024) 131333

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Categorization of precipitation for predicting combined sewer overflows. Application to the City of Montréal

Jonathan Jalbert [a,*], Claudie Ratté-Fortin [b], Jean-Baptiste Burnet [c], Émilie Papillon [d]

[a] Department of Mathematical and Industrial Engineering, Polytechnique Montréal, Montréal (QC), Canada
[b] Department of Decision Sciences, HEC Montréal, Montréal (QC), Canada
[c] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montréal (QC), Canada
[d] City of Montréal, Montréal (QC), Canada

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Shreedhar Maskey, Associate Editor.

Dataset link: https://github.com/jojal5/Publications

Keywords:
Combined sewer overflow
CSO
Precipitation
Rainfall
Decision tree

ABSTRACT

Combined sewer systems are widespread in America and Europe. They often face limitations in transport or treatment capacity, especially during heavy rain events or thaw periods, resulting in combined sewer overflows (CSOs). Predictive modeling for CSOs is essential in a risk management context, and some studies have presented methods to categorize precipitations based on their potential to generate overflows. However, the precipitation classification is usually based on a few characteristics, and its predictive power is limited. The objective of this study is to present a simple yet powerful method to categorize precipitation for predicting CSO occurrences. A prediction model, based on an optimized classification tree, is proposed to predict CSO occurrences as a function of publicly accessible precipitation data. We fit the model on 9 overflow outlets in Montréal city from 2013 to 2019 and use this model to predict CSOs in 2020. The results showed a very good predictive power of overflows, with a prediction rate of 89%, a sensitivity rate of 83%, and a specificity rate of 91%. The method is also more accurate than the 5-category classification currently used by the City of Montréal. The proposed method could be easily applied to another region where CSO data are available, providing a simple and rigorous method for predicting CSOs across urban drainage networks containing many overflow outlets.

1. Introduction

Wastewater and stormwater management are crucial services for water resources quality, infrastructure cost-efficiency, public health and sustainable environmental protection. In Europe, 70% of the total sewer network is comprised of combined sewers, a system that collects and transports both municipal wastewater and stormwater/snowmelt runoff. In North America, most combined systems are concentrated in the older cities (*e.g.* New York and Philadelphia) in the Northeastern and Great Lakes regions (Environmental Protection Agency, 1994). This sewer infrastructure can be limited in its transport or treatment capacity (especially during heavy rain events or snowmelt periods) leading to combined sewer overflows (CSOs) that result into the discharge of untreated wastewater into receiving waterbodies. Between 2013 and 2017, about 2.7% of all wastewater collected and discharged by municipal wastewater systems in Canada were untreated wastewater from combined sewer overflows (CSOs), the equivalent of an average 160 million cubic meters annually (Environment and Climate Change Canada, 2020).

CSOs threats to human health and environment are well documented (Gasperi et al., 2008; Al Aukidy and Verlicchi, 2017; Munro et al., 2019; Soriano and Rubió, 2019; Madoux-Humery et al., 2016), and several studies focused on quantifying water quality dynamics during CSOs events (Casila et al., 2020; Kim et al., 2020) to better understand CSOs impacts on water quality and ecosystem degradation. Among others, Madoux-Humery et al. (2016) analyzed the impact of CSOs on *Escherichia coli* (*E. coli*) concentrations at two drinking water intakes in the Greater Montréal area. Jalliffier-Verne et al. (2017) investigated the impact of climate change on *E. coli* concentrations on CSOs, as well as Iqbal et al. (2019) who evaluated the impact of potential future socio-economic scenarios, such as changes in population and livestock densities, urbanization and climate change. Results showed that *E. coli* concentrations were influenced by socio-economic development and climate change, and are expected to increase in future horizons.

Modeling of CSOs is essential to establish the risks of overflow and to quantify the impact of mitigation measures. Little guidance is available on CSO analysis methods for optimal design of combined sewer

* Corresponding author.
E-mail address: jonathan.jalbert@polymtl.ca (J. Jalbert).

https://doi.org/10.1016/j.jhydrol.2024.131333
Received 3 April 2023; Received in revised form 21 November 2023; Accepted 28 April 2024
Available online 20 May 2024

solutions (Environmental Protection Agency, 1999; Jean et al., 2018). To address these issues, some studies investigated the relationships between CSOs and rainfall characteristics to estimate overflow occurrence or discharged volumes from rainfall series (Sandoval et al., 2013; Yu et al., 2013; Fortier and Mailhot, 2015; Mailhot et al., 2015; Madoux-Humery et al., 2016). Some of them used hydraulic and/or hydrologic models to route precipitation into the drainage system (Thorndahl and Willems, 2008; Yu et al., 2013; Jean et al., 2018). For instance, Thorndahl and Willems (2008) described CSOs occurrence with depth and duration of rainfall events using a commercial urban drainage model. Jean et al. (2018) used PCSWMM software to evaluate how different types of rainfall data impacts CSO volume threshold estimations. Others have used probabilistic methods, such as Mailhot et al. (2015) who developed a simple method to estimate the average number of CSOs per year using rainfall estimates and records of basic spill measurements for 4285 combined sewer overflow outlets.

These studies helped understanding the relationship between rainfall events and CSO occurrence, and identifying rainfall variables to establish key thresholds that may lead to sewer failures in a specific area. For example, Yu et al. (2018) identified CSO occurrence thresholds using rainfall depth, maximum intensity and duration for 67 urban outfalls in Tokyo. Their choice of selecting the same three variables for all 67 outfalls was supported by previous studies (e.g. Mailhot et al., 2015; Schroeder et al., 2011). Discrimination between overflow and non-overflow events was done using maximum coincidence rates calculated for each coupled rainfall variables. Day and Seay (2020) used the same approach by classifying overflow and non-overflow events with coincidence rates calculated for each coupled rainfall variables.

Although these studies presented valuable methods to categorize precipitations according to their potential to generate overflows, quantitative analysis is not often well detailed. Few rainfall variables are tested to identify occurrence thresholds and most studies used rainfall intensity, duration and depth, or a combination of these variables (Mailhot et al., 2015; Schroeder et al., 2011). In addition, thresholds are fixed separately whereas depth, intensity and duration are dependant.

The goal of the present paper is to develop a general and simple method for predicting CSO binary occurrences as a function of precipitation characteristics. The proposed methodology uses classification trees and is illustrated with an application for the management of a public beach in the City of Montréal. Although the study is illustrated for the City of Montréal, the method could easily be applied to other smaller or larger cities where CSO and rainfall data are readily available. The remainder of the paper is as follows: Section 2 describes the publicly available CSO and precipitation data. Section 3 provides an overview of classification trees, and Section 4 presents the results for predicting CSOs as a function of precipitation characteristics using a classification tree. Section 5 discusses the proposed approach and compares it to the existing classification used by the city. Finally, Section 6 provides a conclusion. Note that the Julia code and the data for reproducing all the results are available on the following public repository: https://github.com/jojal5/Publications.

2. Data

2.1. Combined sewer overflows

In the city of Montréal, 60% of the total sewer network is combined sewers. Montréal sewer system has 170 overflow outlets located around the island (Fig. 1). On the South shore, overflows are discharging into the St. Lawrence River while on the North shore, they are discharging into the Des Prairies River. Both rivers are the major drinking water supplies of the city and they serve as recreational areas.

The City of Montréal has made the daily records of combined sewer overflows openly accessible, and the 2013–2020 data can be accessed publicly.[1] CSOs caused by planned construction works, emergencies, and snowmelt have been excluded from our analysis to focus solely on overflows triggered by precipitation. Therefore, only CSOs occurring between May and October were considered to emphasize rainfall-related events. The recorded overflows from the years 2013 to 2019 were retained as the training set, and the recorded overflows in the remaining year of 2020 correspond to the test set. There are a total of 1288 overflow observations between the months of May to October (184 days) for the 7 years of the training set (2013 to 2019 inclusive).

Table 1
Instances of daily overflow outlets spilling among the 9 considered between 2013 and 2019.

Number of individual CSOs	Occurrence
0	1101
1	105
2	31
3	14
4	11
5	10
6	8
7	4
8	3
9	1

This study focuses on the 9 overflow outlets, indicated by the darker red dots in Fig. 1, that influence the public Verdun beach upstream, marked by the blue dot, which includes swimming areas and other recreational activities. These outlets are identified by the following codes in the dataset publicly provided by the city: 4370-03D, 4370-05D, 4430-04D, 4430-01D, 4430-02D, 4420-01D, 4795-01D, 4400-02D, and 4400-01D. Table 1 compiles the number of days in the training set based on the number of overflowing outlets. There are 1101 days out of 1288 where no CSO occurred at any of the 9 outlets, and there is only one day where all 9 outlets overflowed.

The daily overflows from the 9 sites influencing the public beach area are aggregated to offer a comprehensive depiction of CSO dynamics at the beach scale. For a given day, an aggregated overflow is defined if an overflow was recorded at least one out of the 9 overflow outlets. This stringent definition ensures the inclusion of every CSO of the 9 outlets in the overall picture at the beach scale. Therefore, there are 1101 days where no aggregated overflow occurred, and 187 days where at least one occurred. These aggregated overflows are represented by a binary variable.

2.2. Precipitations

Precipitation data recorded every 15 minutes at the Pierre-Elliott-Trudeau International Airport in Montréal from 2013 to 2020 were retrieved from Environment and Climate Change Canada. Fig. 1 shows the location of the station. Precipitation data from May to October were extracted, and rain accumulations over 15 minutes, 30 minutes, 1 hours, 2 hours, 3 hours, 4 hours, 6 hours, 8 hours, 12 hours, and 24 hours were calculated using the appropriate duration sliding window based on previous records. For each duration, the maximum daily accumulation was retained. For example, on a given day, the retained accumulation during a 1 hour period is the largest among the 24 hourly amounts recorded during that day. The initial hourly accumulation of the day at 12:00 A.M. consists of the accumulations over the one-hour duration of the previous day between 11:00 P.M. and 12:00 A.M.

To establish the optimal classification of precipitation accumulation causing combined sewer overflows, a classification tree is proposed. The daily occurrences of overflows correspond to the classification variable: 1 if an overflow occurred and 0 otherwise. The explanatory

[1] https://donnees.Montreal.ca/ville-de-montreal/debordement

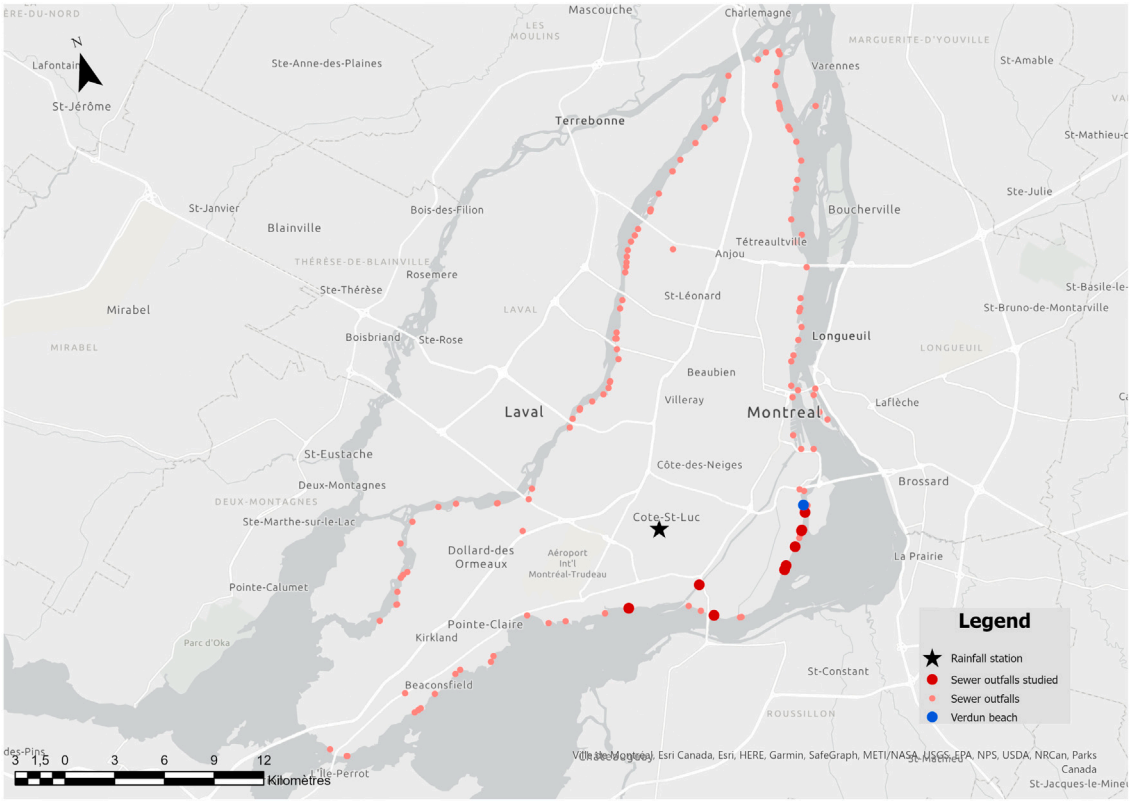

Fig. 1. Location of the 170 overflow outlets (red dots), rainfall station (black star), and the public beach (blue dot) on the Montréal Island. The darker red dots indicate the outlets studied in this paper. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variables (also referred to as *features*) correspond to the maximum daily rain accumulations over the various periods considered.

In order to obtain a simple categorization of precipitation, the maximum depth of the tree was set at 2, producing a partition composed of at most 4 categories. Some leafs could be merged afterwards if the cross-entropy does not increase too much.

3. Classification trees

In the present paper, we propose using a classification tree to categorize precipitation based on its potential to generate CSOs. The binary variable represents the occurrence or absence of a daily aggregated overflow, and the feature space consists of the daily maximal precipitation accumulation during various durations. This section provides a brief overview of classification trees. For more in-depth information, interested readers can refer to dedicated references on classification trees, such as Hastie et al. (2009).

A classification tree is a supervised learning algorithm, notably useful for predicting the class of an observation based on its features. It involves partitioning the feature space into a set of rectangles. For each rectangle, a class value is assigned to the data inside the rectangle using the majority rule. In graph theory, these rectangles in the feature space are referred to as tree leaves. Classification trees are simple yet powerful models that perform well even with numerous correlated explanatory variables. The ease of interpreting binary recursive trees is a major advantage, supporting their widespread use.

The partition is achieved through recursive binary splits. The initial split divides the feature space into two rectangles, aiming to create the purest possible rectangles, where each containing observations belonging to a single class. While achieving perfect purity is impractical, the feature and split point are chosen to minimize an impurity measure. The cross-entropy, or deviance, is a commonly used impurity measure for a rectangle. Let $\hat{p}_{ij}$ represent the proportion of class $j \in 0, 1$ observations in rectangle $i \in \mathcal{V}$, where $\mathcal{V}$ denotes the set of rectangles. Define $H(i)$ as the cross-entropy of node i:

$$H(i) = \sum_{j=0}^{1} \hat{p}_{ij} \log \hat{p}_{ij}.$$

If the leaf is pure, the cross-entropy is null since $p_{ij} \in \{0, 1\}$. Otherwise, the cross-entropy increases with the heterogeneity of the rectangle.

After the first split, each of the two rectangles is further divided into two more rectangles using the same method. This process continues until a stopping criterion is met. Typically, the number of binary splits indicating the tree depth is fixed as the stopping criterion. Once the set depth is reached, the merging of some rectangles is considered, provided it does not significantly increase the impurity measure. This merging criterion is also determined by the analyst and this procedure is referred to as *pruning the tree*. There is a trade-off between capturing the data characteristics and avoiding overfitting, a consideration when determining the number of rectangles.

4. Predicting CSOs with a classification tree

4.1. Partioning precipitation of the training set

A classification tree has been fitted to model the aggregated overflows (the binary variable) as a function of the precipitation features (the daily maximal precipitation accumulations during various durations) in the training set. The depth was fixed at 2, and cross-entropy was used as the impurity measure. Fig. 2 illustrates the fitted tree. Among all possible durations and split points, the 1 hour accumulation with a threshold of 2.95 mm proves to be the most effective for partitioning precipitation based on its potential to trigger overflow, as it corresponds to the first split. The subsequent second split further refines the classification of precipitation according to its potential to

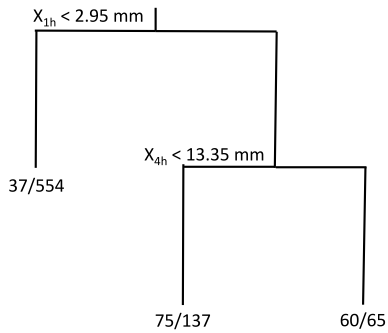

Fig. 2. Classification tree for precipitation for predicting CSO. X_d stands for the precipitation accumulation during the time period d and the ratio at each leaf indicates the number of CSOs over the number of rainfall events.

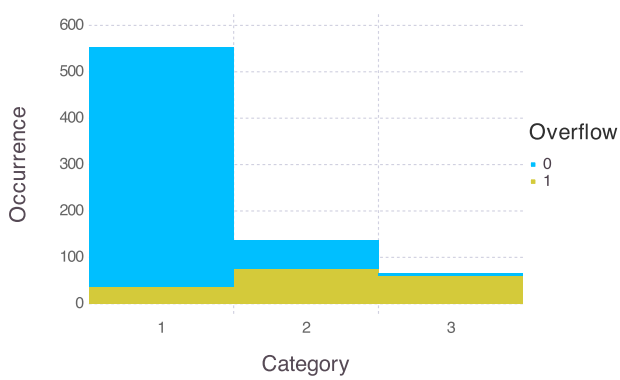

Fig. 3. CSO occurrences as a function of precipitation categories.

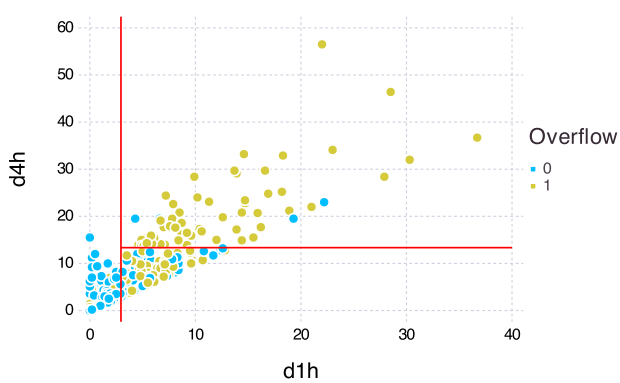

Fig. 4. Overflow occurrences as a function of the rectangles in the precipitation feature space obtained with the classification tree.

cause overflow. The two leaves on the left branch at the second level were merged during the pruning procedure.

This tree implies a 3-class categorization of precipitation after pruning. Category 1 is defined when less than 2.95 mm falls in a 1 hour period. For precipitation in this category, overflow occurs very rarely: 37 times out of 554. Category 2 is defined when the accumulation over a 1 hour period is greater than 2.95 mm, and the accumulation over a 4 hour period is less than 13.35 mm. For precipitation belonging to Category 2, overflow occurs often: 75 times out of 137. Category 3 is defined when the accumulation is greater than 2.95 mm over the 1 hour duration and greater than 13.35 mm over the 4 hour duration. Precipitation in Category 3 triggers overflow most of the time: 60 out of 65. Fig. 3 illustrates the purity of the categorization, while Fig. 4 shows the resulting rectangles in the precipitation feature space.

Table 2
Confusion matrix of the CSO prediction model for 2020 calibrated with the 2013–2019 CSO dataset. Condition 1 stands for aggregated overflow, and condition 0 stands for no overflow.

		True condition	
		1	0
Prediction	1	20	7
	0	4	70

4.2. Precipitation category and number of individual overflows

Although CSOs have been aggregated to identify the precipitation categories triggering them for the 9 outlets influencing the public beach, it is interesting to visualize the number of CSOs for each individual outlet when there is an aggregated overflow. The number of individual CSOs can be seen as a measure of the aggregated CSO severity. An overflow could be deemed *severe* when many outlets overflow. Fig. 5 shows the number of overflowing outlets when there is an aggregated overflow as a function of precipitation category. For Category 1 precipitation, a total of 37 aggregated overflows out of 554 rainy days were recorded. When an aggregated overflow occurs, the number of outlets overflowing is limited to one or two, as shown in Fig. 5(a). Therefore, not only is an overflow uncommon for Category 1 rainfall, but when they occur, they are restricted to a limited number of sites.

For Category 2 precipitation, a total of 75 aggregated overflows out of 137 rainy days were recorded. When an aggregated overflow occurs, it is more severe than for Category 1 precipitation, as the number of overflowing outlets tends to be larger (Fig. 5(b)). Out of the 65 days with Category 3 precipitation in the training set, 60 of them generated an aggregated overflow, and when it occurs, the overflow tends to be the most severe, as shown in Fig. 5(c).

4.3. CSO prediction on the test set

To assess the predictive capability of the model, it was trained on the data from 2013 to 2019 and then used to predict the aggregated CSOs on the test set from 2020. For prediction purposes, we assume that overflow is triggered by Category 2 and Category 3 precipitation (majority rule), while Category 1 does not lead to overflow.

Between May 1st and October 31st, 2020, 74 Category 1 precipitations were recorded, along with 17 Category 2 and 10 Category 3. Table 2 shows the confusion matrix for CSO predictions. The prediction accuracy is 89%, known as the coincidence rate of predicted and observed CSOs and non-CSOs, with a sensitivity of 83% (proportion of correctly predicted CSOs) and a specificity of 91% (proportion of correctly predicted non-CSOs). Using only two explanatory variables for three categories—precipitation accumulation over 1 hour and 4 hour durations, the predictive power of CSOs is very good.

Predicting CSOs for Category 2 precipitation increased the sensibility to 84% but the specificity and the accuracy decreased to 80% and 82%, respectively. It was decided to maximize specificity to increase confidence in predicted absence of overflows.

5. Discussion

5.1. Model fit and predictive power

The aim of the present paper was to identify a small set of simple rules for predicting CSOs using easily accessible data. This is why

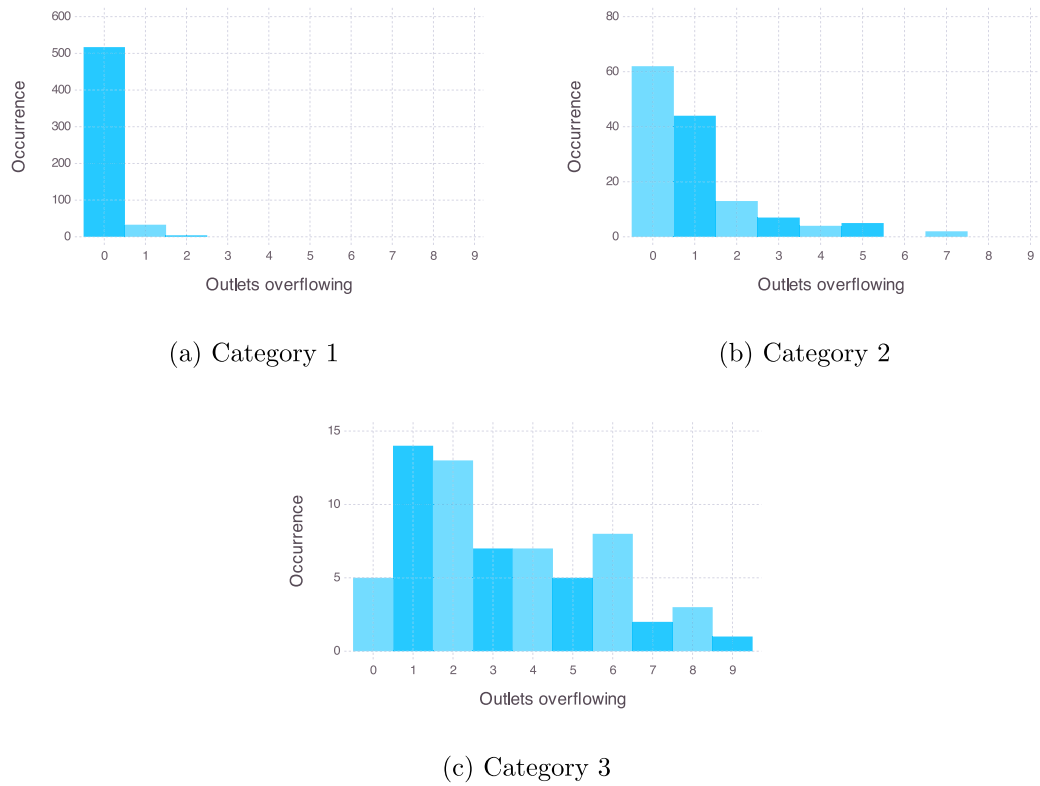

Fig. 5. Empirical distribution of the number of outlets that overflow among the 170 based on rainfall categorization when there is an aggregated overflow.

only rainfall data recorded at Pierre-Elliott-Trudeau Airport was used. Despite the fact that these data do not capture the spatial heterogeneity of precipitation on the Montréal island, the model fit and predictive power are still very good, as shown respectively in Fig. 2 and in Table 2. Precipitation recorded by the meteorological station network utilized by the City of Montréal could potentially enhance the fit and prediction power of the model. These data could be treated as additional explanatory variables, and the same tree-fitting procedure could find the best variable and split point among all additional possibilities. However, these data are not publicly available, and, in general, they do not undergo rigorous quality checks to meet the World Meteorological Organization standards, as the data from the Pierre-Elliott-Trudeau Airport. Therefore, we propose a simple yet powerful model using data that are readily available to scientists who need to predict CSOs.

The three categories of precipitation, defined by the precipitation accumulation over 1- and 4 hour durations, made it possible to achieve a good fit and predictive power for overflows. The performance may be further improved by considering a tree with more depth. However, the precipitation categories would consequently increase, and we have shown that the proposed use of three categories constitutes an excellent trade-off between simplicity, fit, and predictive power.

Precipitation accumulations over several periods of time constitute a powerful set of explanatory variables for predicting CSOs that are easy to obtain and interpret. Nevertheless, other explanatory variables could be considered, particularly the duration of the period of dry weather between two rainy events. This variable could be integrated into a more complex model to attempt to explain the heterogeneity of overflows for the same rainfall intensity. In the proposed 3-category model, this variable is not easily usable. It could be incorporated into another type of model, such as logistic regression, but this is beyond the scope of the present paper.

5.2. Binary splits

As shown in Fig. 4, the rainfall categories are defined by rectangles. In cases where the two explanatory variables are strongly dependent, it would be possible to split the feature space using another rule, such as triangle partitioning. This could be achieved using multiway splits, but the resulting model would be generally less intuitive and simple to interpret (Hastie et al., 2009). Also, multiway splits can be accomplished by a series of binary splits, maintaining a simple interpretation. Additionally, because it is impossible for precipitation accumulated over 3 hours to exceed accumulation over 4 hours, the rectangles shown in Fig. 4 effectively act as triangles since no points can lie below the unit slope line. We opted for the rectangle interpretation because it directly stems from binary splits and is more versatile for explanatory variables with lower levels of dependence.

5.3. Comparison with the existing categorization

The city engineers are currently using a rainfall categorization based on the characteristics of rain events for various purposes, including CSO prediction (Mailhot et al., 2019). A rain event is defined as a continuous streak of precipitation, which can be extracted from the first non-zero record to the last one in the 15-minute time step records. The 5-category classification utilizes the volume S of precipitation (in mm) that has accumulated during the rain event, as well as the maximum intensity I in mm/h, calculated over a two-hour period during the event:

Category 1 : $S < 10$ mm.

Category 2 : $10 \text{ mm} \leq S < 20 \text{ mm}$ and $I < 5$ mm/h.

Category 3 : $10 \text{ mm} \leq S < 20 \text{ mm}$ and $I \geq 5$ mm/h.

Category 4 : $S \geq 20$ mm and $I < 5$ mm/h.

Category 5 : $S \geq 20$ mm and $I \geq 5$ mm/h.

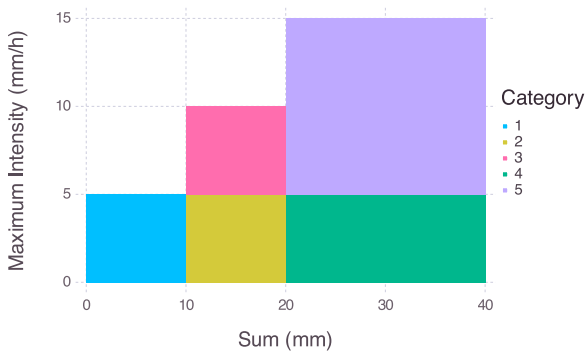

Fig. 6. Precipitation categories used by the City of Montréal for CSO management.

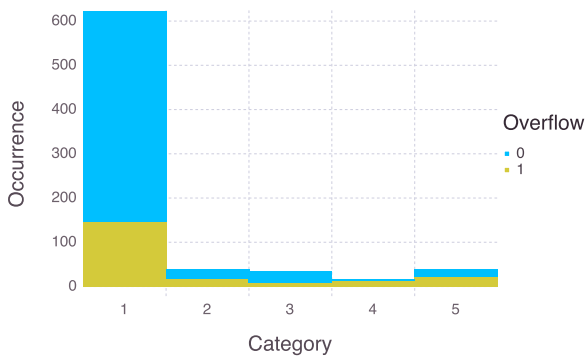

Fig. 7. Aggregated CSO occurrences (2013–2019) as a function of the rainfall event categories currently defined by the city.

Table 3
Confusion matrix for the CSO predictions on the test set using the city classification.

		True condition	
		1	0
Prediction	1	5	1
	0	24	65

The precipitation classification is illustrated in Fig. 6. According to Mailhot et al. (2019), events of categories 1 to 3 infrequently generate CSOs, whereas events of categories 4 and 5 are the most likely to trigger CSOs.

Fig. 7 summarizes CSO occurrences as a function of the rainfall event category, as currently defined by the city. However, this categorization does not discriminate well, as CSOs and non-CSOs are mixed in the same rainfall category. If we assume that rainfall events of categories 4 and 5 generated overflows, the confusion matrix illustrated in Table 3 is obtained on the test set. The correct rate of predictions is 74%, with a sensitivity of 17% and a specificity of 98%. Both the correct rate and the sensitivity of the predictions are lower compared to the proposed 3-category classification. This is mainly due to the fact that the definition of the categories has not been optimized for the prediction of overflows.

Additionally, the classification by rainfall events is a little more difficult to achieve because the streak has to be identified with several arbitrary choices, such as when to separate or merge rainfall events. For instance, if two rainy events are an hour apart, do they need to be merged? What if they are three hours apart? This problem does not arise with precipitation accumulations, as proposed in Section 2.2.

5.4. Dynamic sewer management

The City of Montréal implements real-time flow control in the sewer system during rainfalls to minimize overflows. The network configuration can be dynamically adjusted for each rainfall event by, for instance, opening or closing some gates. This dynamic nature might make it more challenging to establish a clear relationship between rainfall data and the occurrence of overflows. Therefore, we did not anticipate achieving perfect predictive power of CSOs as a function of precipitation. Nevertheless, the achieved predictive power without considering this dynamic management is satisfactory. Taking into account real-time control would necessitate hydraulic simulations, which is beyond the scope of this paper.

6. Conclusion

This paper proposed a simple yet powerful method to categorize precipitation for the purpose of predicting CSOs occurrence. We applied the method on 9 overflow outlets in the City of Montréal using data from 2013 to 2020, and used simple rules for predicting CSOs using easily accessible data. The resulting model, using only a 3-category precipitation classification, performed very well on the test set with a prediction rate of 87%, a sensibility rate of 76% and a specificity rate of 98% fro predicting CSOs. It also exceeded the performance of the model used by the city of Montréal using the 5-category classification (predictions rate of 74%, sensibility of 17% and specificity of 98%).

The framework developed in this paper can be easily adapted to another subset of overflow outlets or to another set of meteorological station observations. The resulting precipitation classification could be different, as the watershed characteristics of the outlet subsets could also vary. Additionally, the method can be easily applied to another region where CSO data are available.

Future work could focus on adding other features to the model. For example, a previous dry period, water flow in sewers before precipitation, snowmelt, etc. Future work should also focus on receiving water impact. Nevertheless, we believe that the proposed simple procedure can be a useful tool for managing CSO occurrences based on precipitation categories. The proposed model can help, for example, the beach manager to anticipate the decision to open or close the beach according to the rain category, even before performing water quality tests.

CRediT authorship contribution statement

Jonathan Jalbert: Conceptualization, Funding acquisition, Methodology, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **Claudie Ratté-Fortin:** Formal analysis, Methodology, Writing – original draft. **Jean-Baptiste Burnet:** Supervision, Writing – original draft. **Émilie Papillon:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jonathan Jalbert reports financial support was provided by IVADO. Jonathan Jalbert reports financial support was provided by Natural Sciences and Engineering Research Council of Canada.

Data availability

The code and data are provided on this public repository: https://github.com/jojal5/Publications.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number RGPIN-2018-04481] and IVADO, Canada [funding reference number PRF-2019-3295824760].

References

Al Aukidy, M., Verlicchi, P., 2017. Contributions of combined sewer overflows and treated effluents to the bacterial load released into a coastal area. Sci. Total Environ. 607, 483–496.

Casila, J.C., Azhikodan, G., Yokoyama, K., 2020. Quantifying water quality and flow in multi-branched urban estuaries for a rainfall event with mass balance method. Water Sci. Eng..

Day, C.A., Seay, G., 2020. Impacts of storm characteristics on generating sanitary sewer overflow (SSO) events for an urban sewershed. Pap. Appl. Geogr. 6, 460–470.

Environment and Climate Change Canada, 2020. Canadian environmental sustainability indicators: Municipal wastewater treatment. https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/municipal-wastewater-treatment.html. (Accessed 13 March 2021).

Environmental Protection Agency, 1994. Combined sewer overflow (CSO) control policy. Fed. Regist. 59, 18687–18698.

Environmental Protection Agency, 1999. Combined Sewer Overflow Guidance for Monitoring and Modeling. Technical Report EPA/832-B-99-002, Office of Wastewater Management, Environmental Protection Agency, URL: https://www3.epa.gov/npdes/pubs/sewer.pdf.

Fortier, C., Mailhot, A., 2015. Climate change impact on combined sewer overflows. J. Water Resour. Plan. Manag. 141, 04014073.

Gasperi, J., Garnaud, S., Rocher, V., Moilleron, R., 2008. Priority pollutants in wastewater and combined sewer overflow. Sci. Total Environ. 407, 263–272.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: Data mining, inference and prediction, second ed. Springer.

Iqbal, M.S., Islam, M.M., Hofstra, N., 2019. The impact of socio-economic development and climate change on *E. coli* loads and concentrations in Kabul River, Pakistan. Sci. Total Environ. 650, 1935–1943.

Jalliffier-Verne, I., Leconte, R., Huaringa-Alvarez, U., Heniche, M., Madoux-Humery, A.-S., Autixier, L., Galarneau, M., Servais, P., Prévost, M., Dorner, S., 2017. Modelling the impacts of global change on concentrations of escherichia coli in an urban river. Adv. Water Resour. 108, 450–460.

Jean, M.-È., Duchesne, S., Pelletier, G., Pleau, M., 2018. Selection of rainfall information as input data for the design of combined sewer overflow solutions. J. Hydrol. 565, 559–569.

Kim, M., Ligaray, M., Kwon, Y.S., Kim, S., Baek, S., Pyo, J., Baek, G., Shin, J., Kim, J., Lee, C., et al., 2020. Designing a marine outfall to reduce microbial risk on a recreational beach: Field experiment and modeling. J. Hazard. Mater. 124587.

Madoux-Humery, A.-S., Dorner, S., Sauvé, S., Aboulfadl, K., Galarneau, M., Servais, P., Prévost, M., 2016. The effects of combined sewer overflow events on riverine sources of drinking water. Water Res. 92, 218–227.

Mailhot, A., Talbot, G., Bolduc, S., 2019. Évolution des régimes de précipitations en climat futur pour la région de Montréal. Research report AM03-2018, Institut National de la Recherche Scientifique, 112 p..

Mailhot, A., Talbot, G., Lavallée, B., 2015. Relationships between rainfall and combined sewer overflow (CSO) occurrences. J. Hydrol. 523, 602–609.

Munro, K., Martins, C.P., Loewenthal, M., Comber, S., Cowan, D.A., Pereira, L., Barron, L.P., 2019. Evaluation of combined sewer overflow impacts on short-term pharmaceutical and illicit drug occurrence in a heavily urbanised tidal river catchment (London, UK). Sci. Total Environ. 657, 1099–1111.

Sandoval, S., Torres, A., Pawlowsky-Reusing, E., Riechel, M., Caradot, N., 2013. The evaluation of rainfall influence on combined sewer overflows characteristics: The Berlin case study. Water Sci. Technol. 68, 2683–2690.

Schroeder, K., Riechel, M., Matzinger, A., Rouault, P., Sonnenberg, H., Pawlowsky-Reusing, E., Gnirss, R., 2011. Evaluation of effectiveness of combined sewer overflow control measures by operational data. Water Sci. Technol. 63, 325–330.

Soriano, L., Rubió, J., 2019. Impacts of combined sewer overflows on surface water bodies. The case study of the Ebro River in Zaragoza city. J. Clean. Prod. 226, 1–5.

Thorndahl, S., Willems, P., 2008. Probabilistic modelling of overflow, surcharge and flooding in urban drainage using the first-order reliability method and parameterization of local rain series. Water Res. 42, 455–466.

Yu, Y., Kojima, K., An, K., Furumai, H., 2013. Cluster analysis for characterization of rainfalls and CSO behaviours in an urban drainage area of Tokyo. Water Sci. Technol. 68, 544–551.

Yu, Y., Zhang, S., An, A.K., Furumai, H., 2018. Simple method for calculating hydraulic behavior of combined sewer overflow from rainfall event data. J. Water Resour. Plan. Manag. 144, 04018061.