Environmental Challenges 12 (2023) 100735

Contents lists available at ScienceDirect

Environmental Challenges

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

Comparative study of deterministic and probabilistic assessments of microbial risk associated with combined sewer overflows upstream of drinking water intakes

Raja Kammoun[a,*], Natasha McQuaid[a], Vincent Lessard[b], Michèle Prévost[a], Françoise Bichai[a], Sarah Dorner[a]

[a] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Quebec, Montreal H3C 3A7, Canada
[b] Conseil des bassins versants des Mille-Îles, 15 Chemin de la Grande-Côte bureau 105, Saint-Eustache, Quebec J7P 5L3, Canada

A R T I C L E I N F O

Keywords:
Deterministic risk assessment
Probabilistic risk assessment
Bayesian network
Combined sewer overflow
Urbanized watershed
R studio

A B S T R A C T

Combined sewer overflows (CSOs) are a source of microbial contamination of drinking water intakes located downstream from their discharge. To safeguard the quality of the source water, it is essential to evaluate the risk levels associated with these municipal structures. This study compares two risk assessment approaches to test their applicability for assessing the risk of CSOs to drinking water intakes in a highly urbanized watershed. The first approach was based on a deterministic equation that combines the characteristics of an overflow structure allowing the risk to be rated as very low, low, medium, high, or very high. The second probabilistic risk assessment approach yielded findings that are probabilistically distributed across the five levels of risk. This approach was developed by constructing a novel Bayesian network to probabilistically link the different factors defining the exposure of water intakes to the hazards of CSOs. The comparison between the results of these two approaches highlighted the importance of simultaneously considering many scenarios for assessing the risk of contamination of source waters. It was possible to use the Bayesian network rather than the deterministic equation, which only supports one scenario at a time. It was also shown that the deterministic approach often overestimated risk levels for CSO outfalls close to the water intake. This occurred because the assessment process emphasized the distance factor between the discharge point and the water intake, while neglecting other crucial characteristics of the overflow, such as duration and frequency. In particular, the deterministic approach tended to underestimate risk for CSOs associated with low overflow frequencies as it did not support scenarios of overflow duration, unlike the probabilistic approach. The validation and sensitivity analysis of the Bayesian model revealed that the population residing in the CSO's drainage basin, along with the frequency and duration of the overflows, exerted the greatest influence on the resulting risk levels. These factors outweighed other variables utilized in the risk assessment, including vulnerability of the drinking water intake, the type of overflow recorder, pipe diameter, and variables defining the exposure of the water intake to the discharge. In the context of implementing action plans, the Bayesian network is estimated as a cost-effective technique as it prioritized overflow structures needing special attention in a highly urbanized watershed, where the same CSOs were deterministically rated as having the same risk level. The results also demonstrated the effectiveness of the Bayesian model in addressing data gaps faced by water managers and stakeholders. The Bayesian model proved capable of assessing risks with uncertainties for CSOs, even with limited input data available. These findings can assist managers in identifying problematic structures by considering various scenarios, unlike the deterministic approach, which left almost half ($n = 42$) of the study site's overflow structures unassessed due to data limitations.

1. Introduction

Combined sewer overflows (CSOs) effluents contain a mixture of microbial pathogens (Passerat et al., 2011; Madoux-Humery et al., 2016; Jalliffier-Verne et al., 2017), physico-chemical contaminants (Birch et al., 2011; Gasperi et al., 2012), and emerging contaminants (Hajj-Mohamad et al., 2017; Petrie, 2021). Therefore, government entities are focusing on assessing potential risks from these anthropogenic activities to protect drinking water sources. A review of risk assessment approaches (Prévost et al., 2017) for these activities highlighted that

* Corresponding author.
E-mail address: raja.kammoun@polymtl.ca (R. Kammoun).

https://doi.org/10.1016/j.envc.2023.100735
Received 14 April 2023; Received in revised form 24 May 2023; Accepted 25 May 2023

laws, regulations, and recommendations of several countries use similar conceptual risk assessment methodology: a deterministic model. This model is commonly based on setting risk levels according to a risk matrix that combines the category of probability, or likelihood, against the category of consequence severity. Risk matrices are used in Canada, New Zealand, and Australia (Table S1). Instead of the risk matrix, the overall degree of risk can be assessed based on descriptive information on the pollutants that may be discharged by human activities, the location of the pollution source, water quality, land use, etc. Depending on the extent to which numerical measurements are available, these descriptions may be either general or precise. This is the case for Maine, New York, and California risk assessment methods (Table S1).

These methods present a quick and easy approach (Peace, 2017) to produce a comprehensive portrait of pollution hazards in all watersheds. However, improperly designed, or deployed risk matrices may raise uncertainty in outcomes (Peace, 2017; Cox, 2008). As explained by Grafton and Little (2017), the risk matrix deficiency is revealed when events with high consequences and low frequency are compared to those with low consequences and high frequency as having the same risk level. In addition, effective risk management decisions cannot be derived from the translation of risk ratings into protection plans (Cox, 2008; Vatanpour et al., 2015).

Researchers have developed improved deterministic methods to evaluate the level of microbial risk associated with CSOs using targeted water sampling and laboratory analysis (Madoux-Humery et al., 2015; Calderon et al., 2017; Al Aukidy and Verlicchi, 2017), or real-time measurement of a biochemical indicator of fecal pollution using ColiMinder (Burnet et al., 2021; Sylvestre et al., 2021) to identify water quality and analyze wastewater impact on receiving waterways. Therefore, statistical analysis (Madoux-Humery et al., 2015; Sylvestre et al., 2020), Quantitative Microbial Risk Assessment (QMR(A) (Sylvestre et al., 2021; McGinnis et al., 2022), and coupled hydrodynamic and water quality modeling (Locatelli et al., 2020; Taghipour et al., 2019) are also used as a risk assessment methods to better evaluate the effect of CSOs on surface water quality and public health. These deterministic modeling methods simulate scenarios where the input values are known. These strategies rely on specialized and large-scale data collected from advanced equipment or laboratory analysis, often unavailable, to calibrate models and reduce uncertainty regarding the interaction of different system components. Several studies have used artificial neural network (ANN), including uncertainty, to model and forecast fecal indicator bacteria in CSOs (Vijayashanthar et al., 2018), and to investigate different modalities related to CSOs, such as the flow rate (El Ghazouli et al., 2022), water level in the CSO structure (Zhang et al., 2018), and hydraulic performance of CSOs (Aziz et al., 2013; Mounce et al., 2014). ANN is appropriate for complex problems and adaptive learning (Sojobi and Zayed, 2022). However, it requires a large dataset for neural network learning (Mounce et al., 2014; Rosin et al., 2021).

To support source water protection by prioritizing high-risk combined sewer systems at the regional level, it is unrealistic to conduct wastewater modeling and analysis of each CSO due to the substantial resources and data required, especially in highly urbanized jurisdictions. Decision-makers must adopt a practical and accurate strategy for assessing risk (Dirckx et al., 2022) while realistically demanding limited resources. Recent studies (Kaikkonen et al., 2020; Phan et al., 2016; Yu and Zhang, 2021) suggest that Bayesian networks (BNs) are ideally a suitable tool for addressing the challenges associated with data scarcity in probabilistic risk assessment, combining different types of knowledge, and uncertainty in describing the interaction between a random set of variables that outline the interaction of system components. The BN is an acyclic graph composed of nodes linked by conditional dependencies represented as probability distributions (Shan et al., 2019; Stritih et al., 2020).

Few studies have developed a BN to study different features related to CSOs. Wijesiri et al. (2018) examined the potential health risks associated with poor water quality based on physicochemical attributes. Other studies used BN to assess the CSO pipe failure and their degradation (Hahn et al., 2002; Elmasry et al., 2017). To our knowledge, only Goulding et al. (2012) built a BN model that includes QMRA to evaluate public health risks exposed to microbial contamination in sewage overflows in Australia. Given that this BN uses analytical findings not required by North American regulations, such as raw sewage concentration, it cannot be extended to all drinking water source protection authorities. Thus, this study investigates exposure to the threat from irrigation water and recreational activities. There have been no studies to date that have used BN to predict the microbiological risk assessment for drinking water intakes from CSOs. Therefore, there is an urgent need for studies in this field to advise drinking water treatment plant managers and water stakeholders the necessary safeguards to protect drinking water intakes (DWIs) from microbial contamination.

The main objective of this study is to assess the microbial risk associated with CSOs events on drinking water sources using a limited dataset. To achieve this objective, a Bayesian model was developed, which considers various scenarios related to CSO features, such as frequency and duration of overflow, and their impact on the water quality of DWIs. Moreover, this study aims to compare two different risk assessment methods: deterministic and probabilistic. It seeks to illustrate the applicability and limitations of each method in assessing risks in a highly urbanized watershed, and to demonstrates the most effective strategy that could help water managers and stakeholders make informed decision about drinking water sources protection in Quebec (Canad(A).

2. Material and methods

In this study, an assessment of microbial risks associated to CSOs upstream DWIs was performed using both deterministic and probabilistic approaches. Fig. 1 summarizes the steps followed in this study, and a more detailed description of these steps is provided in Sections 2.1 and 2.2.

2.1. Study site and CSOs

This risk assessment research focuses on four DWIs (DWI_1, DWI_2, DWI_3, and DWI_4) supplied from the same river in southern Quebec, Canada. This study location was chosen because it exemplifies the challenges of highly urbanized areas. There are a total of 89 CSO outfalls (from CSO_1 to CSO_89) upstream of these DWIs. These outfalls are in the immediate (500 m upstream DWI) and intermediate (10 km upstream DWI) protection zones, which are delineated according to the requirements of the Water Withdrawal and Protection Regulation (Quebec Government, 2022).

Some overflow structures in Quebec have real-time telemetry systems that track the frequency and duration of overflows. Other structures have a marker that shifts position due to overflows (Quebec Government, 2022; MELCC, 2021). In the latter case, overflow duration cannot be recorded, but the presence or absence of an overflow is noted during a weekly technical visit. Overflow data obtained manually or automatically was analyzed and the number of sewer structures, the frequency, and the duration of their overflows are summarized in the supplementary Table S2.

2.2. Deterministic risk assessment method

This research used the deterministic approach developed by McQuaid et al. (2019) to estimate the risk associated with all CSOs upstream of the four DWIs. This method was chosen because it presents an improvement over the regulatory approach in Quebec which is based on the risk matrix (Quebec Government, 2022). The risk level was calculated using metrics representing the wastewater flow, microbial concentration, the magnitude of the overflow, and the DWI's proximity to the outfall. The overflow index (OI) was computed using the formula below

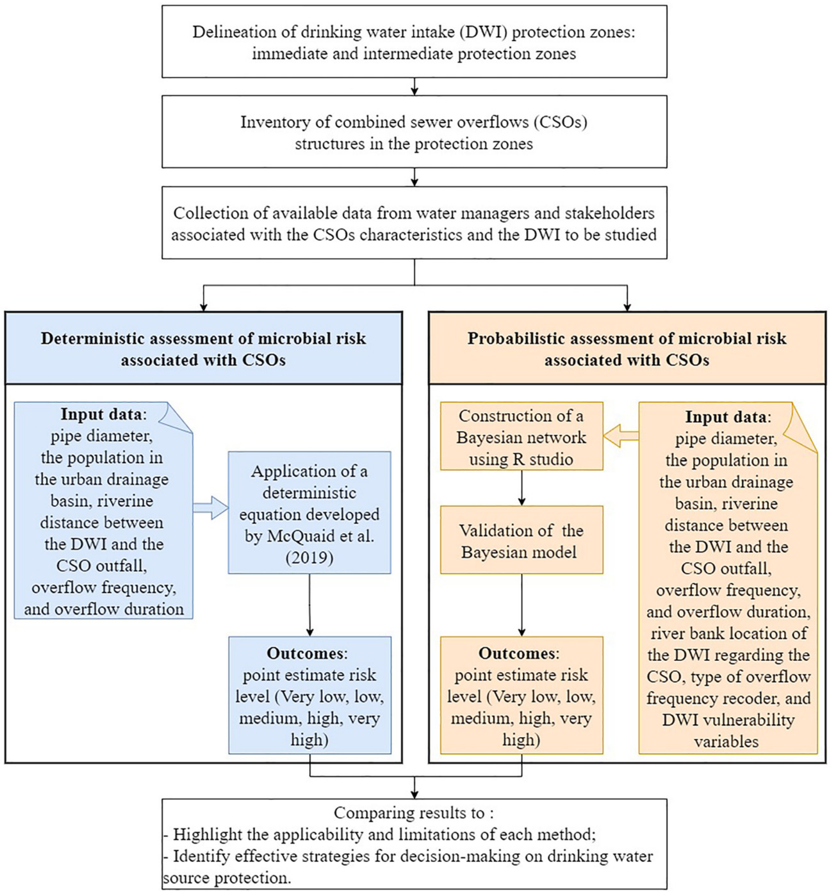

Fig. 1. Methodological flowchart for assessing microbial risks associated with CSOs.

(Eq. (1)). It generates a single numeric score, which is then classified into one of five risk categories, ranging from very low to very high.

$$OI = \left(D^2 .\, Max \sum_{i=1}^{n=5} X .\, Pop \right) / \ln(RD) \tag{1}$$

where D is the pipe diameter (m) at the outfall, Pop is the estimated population for the dissemination block boundary present in the urban drainage basin (UD(B), RD is the riverine distance (m) between the DWI and the overflow discharge point, $Max \sum_{i=1}^{n=5} X$ is the maximum annual cumulative duration (h) of overflows observed during five years, X is substituted with the overflow frequency if data on the duration is unavailable. The data used to apply this deterministic approach and their sources of collection are reported in the supplementary Table S3.

The overflow index was computed for all CSO structures with outfall in the intermediate protection zone of the DWI. A very high risk level was automatically assigned to any structure located in the immediate protection zone, regardless of its overflow characteristics.

2.3. Probabilistic risk assessment method

2.3.1. Bayesian network construction

The BN model presented in Fig. 2 was designed to assess the microbial risk of CSOs to DWIs by combining three sub-models – microbial hazard, exposure to the hazard, and global vulnerability of the DWI. The Bayesian network sub-model that assesses DWI vulnerability was developed by Kammoun et al. (2023.) using Netica (Norsys Software and Corp. Netica 2021) for agricultural risk assessment. This sub-model (Supplementary Fig. S1) was replicated using the R programming language (R studio, version 4.1.3) and interconnected with the CSO risk assessment sub-model, which includes key variables that describe CSO

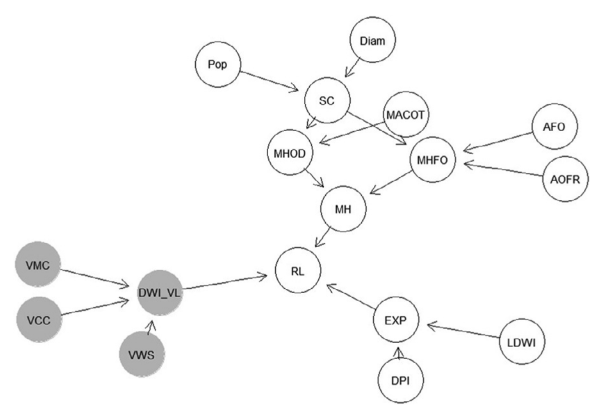

Fig. 2. Bayesian network for CSOs risk assessment. Abbreviation used in the model: Pop: population in urban drainage basin, diam: pipe diameter, SC: severity of consequences, AFO: annual frequency of overflows, AOFR: automatic overflow frequency recorder, MHFO: microbial hazard based on frequency of overflows, MACOT: maximum annual cumulative overflows time, MHDO: microbial hazard based on overflow duration, MH: CSO microbial hazard, DPI: distance prioritization index, LDWI: location of the DWI regarding the CSO, EXP: exposure of the DWI to the hazard of CSO, DWI_VL: DWI vulnerability level, VCC: vulnerability to chemical contamination, VMC: vulnerability to microbial contamination, VWS: vulnerability to water scarcity, RL: CSO risk level. The gray filled nodes are the key nodes of a BN sub-model proposed by Kammoun et al. (42).

structures, overflow events, DWI and overflow outfall location. Each variable constituting the BN is represented as one node. Child nodes get input from one or more parent nodes (Kaikkonen et al., 2020). The data used as input in this Bayesian Network, along with their corresponding sources of collection, are listed in supplementary Table S3.

Probabilistic risk analysis was conducted by filling in conditional probability tables (CPTs), which describe the probabilistic relationships between the parent nodes and their influence on the child nodes. The conditional probability distributions in CPTs were determined using an adapted mixed aggregation method ($F_{add\text{-}max}$) for BNs as proposed by Kammoun et al. (2023). This method, which is described in more details in the supplementary materials (Text S1), involves combining an additive ($f_{ad(D)}$ and a maximum (f_{max}) aggregation. To make one aggregation more important than the other, an alpha (α) weight between 0 and 1 is given to f_{add}. Otherwise, a weight of 1-α should be given for f_{max}. The additive aggregation requires assigning weights (W_n) (from 0 to 1) that show how much each parent node affects the child node.

The conditional probabilities distribution of each CPT and the optimal α and W_n combinations for each parent node were set by developing an algorithm that tests all possible iterations, incorporates interpreted preconditions deduced from the literature, and relies on expert knowledge. This algorithm is effective in identifying the optimal combination among thousands. It was developed by generally answering the following questions for each CPT:

- Which parent node contributes the most to the child node?
- How does the status of a child node change when the status of a parent node increases or decreases by one level?
- Are there any extreme probability values in the CPT that need to be excluded by evidence?

For example, a single CPT of a child node (x_1) that is linked to two parent nodes (x_2 and x_3) with five statuses for each node (k_i=5) must be filled up after testing 121 iterations, each of which has 125 probability values ($P(x_1=a_k|x_2, x_3)$, where a_k is the possible value of the node x_1). The 121 iterations depict all conceivable combinations of α, W_{X2} and $W_{X3,}$ where the sum of the weights equals 1. This approach must be applied to each CPT constituting the BN while adjusting the parameters that define the number of nodes and their status. A R code was developed for this purpose (supplementary Text S2). It illustrates all the criteria and hypotheses (Table 1) implemented in the algorithm to determine the most appropriate conditional probabilities distribution to fill the CPTs. The status of all nodes shown in Table 1 was mainly determined based on the levels used in the deterministic approach of McQuaid et al. (2019), as well as expert opinion.

After defining the probabilistic connection between nodes, the "bnlearn" package (Scutari, 2010) in R was used to learn the structure of the BN, and the "Rgraphviz" package (Hansen et al., 2022) was used to plot the network. The R code used for this purpose is given in the supplementary Text S3.

2.3.2. Bayesian network validation

Once the risk assessment model was constructed, the acceptability of the probabilities that this BN inferred was assessed to determine the model's reliability and credibility. This assessment was computed using the "querygrain" function of the "gRain" package (Hojsgaard, 2012), as demonstrated in the R code shown in Supplementary Text S4. By altering one BN input node at a time, outcomes enabled the examination of the direct perturbation of the risk level node (RL). The probability findings shown in Supplementary Fig. S2 indicated that changes to the population in UDB (Pop), the frequency (AFO) and the duration of the overflow (MACOT) nodes had the greatest impact on the "RL" node compared to the other nodes. This supports the existence of a substantial causal link between these three nodes and the relative risk of CSOs compared to the remaining input nodes. These results were further validated by examining the outcomes of the sensitivity analysis conducted for this Bayesian model. This sensitivity analysis refers to assessing the degree of influence exerted by the parent nodes on the outcomes of a target node (Grêt-Regamey and Straub, 2006; Harris et al., 2017; Zou and Yue, 2017), using the calculation of mutual information between variables, as described by Pearl (1988) . In this study, the target node was specified by the node representing the risk levels (RL). The results of this sensitivity analysis (Supplementary Table S4) identified the key factors that influence the risk levels associated with CSOs, ranked in descending order of influence: the population in the UDB (Pop), the frequency of overflow (AFO), the duration of overflow (MACOT), the DWI vulnerability (DWI_VL), the automatic overflow frequency recorder (AOFR), the pipe diameter (diam), the distance prioritization index (DPI), and the location of the DWI relative to the CSO (LDWI). These findings provide insights into the most critical factors that should be monitored and controlled to reduce the risk of source water contamination.

The validity of the BN was completed by running known scenarios, and the results were then evaluated using expert judgment. First, we examined the most extreme combinations of the scenario analysis, by thinking about the output in the best (S1) and worst case (S2) scenarios, and then we looked at the more generic version of the scenario analysis (S3 to S11). Supplementary Fig. S3 depicts the hierarchical grouping of scenarios based on nodes "Pop", "AFO", and "MACOT" that have the largest influence on the risk level. The scenario groups labeled "Population max", "Frequency max" and "Duration max" are defined by population (Pop), the frequency (AFO), and the duration overflow (MACOT) nodes at their maximum statuses, respectively. Similarly, the groups linked with the "min" indices are related with the nodes having statuses at a low range. As an example, scenario S5 is defined by a very low population (0 to 50 person), very high overflow frequency (≥14), and extremely long overflow duration (≥262). Supplementary Table S5 provides details of all modeled scenarios.

The outcomes of the RLs related to the 11 scenarios (from S1 to S11) are shown in Supplementary Fig. S4. The probability of a very high risk level was highest (99%) for the worst-case scenario S2. In comparison, the probability of a very low risk level (97%) was highest for the best-case scenario S1. These estimations were expected given that the very high risk for S2 was associated with the combination of the critical nodes (Pop, AFO, and MACOT) in their worst range, and the overflow was predicted to take place at a short distance from the DWI on the same riverbank. In comparison, S1 implied all nodes that were coupled at their best range during BN construction.

Very high risk probabilities of 24%, 8%, and 22%, respectively, were observed in scenarios S6, S7, and S8. For these three scenarios, the large estimated population (≥ 5 000 people) was the leading factor explaining these probabilities. Scenario S6 was characterized by a low frequency of overflow (AFO = 1 to 2) and a short period of overflow (MACOT = 1 to 23 h). However, it was assumed that the discharge would be extremely close to the DWI, which supported the 24% very high risk rating. The outfall's placement on the opposite riverbank and at the end of the intermediate protection zone boundary justified a 16% reduction in the very high risk rating for S7 compared to S6. Also, S8 had a 14% higher probability that the risk would be classified as "very high" compared to S7 due to its more frequent and longer period of overflows.

Due to the uncertainty around the frequency and duration of overflows in S9, which were not captured by an automated recorder as they were in S3, the medium level risk predicted by S9 was 26% higher than that predicted by S3. Additionally, S4 had a higher probability of the medium risk level than scenarios S5 and S10 because of the higher overflows' frequency and proximity to the DWI. The results of S3 and S10 showed that varying the overflow duration from a minimum to a maximum state while maintaining all other variables in the same state led to a 54% increase in the medium risk level. It was also shown by scenarios S3 and S11 that changing the overflow frequency from the lowest to the highest state shifted the probability of the medium risk level by 80%.

A quantitative technique could not be used to validate and improve the precision of the Bayesian model that was designed in this study. However, in qualitative terms, we may infer that the findings connected with the various scenarios outlined below were plausible and consistent with the experts' expectations. Accordingly, this Bayesian model can be

Table 1
Criteria and hypothesis used to fill conditional probability tables constituting the BN.

Child node	Parent nodes: Attribute	Parent nodes: Status*	Criteria and hypothesis for determining the link between parent and child nodes	CPT filling parameters selected
Severity of consequences (S(C)	Population (Pop)[1] Diameter (diam)[2]	< 50 50 to 1000 1000 to 2500 2500 to 5000 ≥ 5000 < 0.9 0.9–1 1–1.2 1.2–1.5 ≥ 1.5	- The number of individuals that are housed in the UDB has a substantial correlation with the load of microbiological pollutants that are released from the overflow structure. Consequently, the severity of the consequences tends to be proportional to the population that occupies the UDB. According to Madoux-Humery et al. (2013) research, the concentration of *E. coli* in raw sewage water varies depending on the kind of land use, the UDB area, and the population density. The average *E. coli* concentration per person per combined sewer overflow event was estimated by Jalliffier-Verne et al. (2017) to be 5.5 CFU/100 mL/p in a southern Quebec urban region that is fairly similar to our study site. Olds et al. (2018) also demonstrated a causal relationship between the level of sewage contamination and the degree of urbanization by using genetic markers for human-associated indicator bacteria such human Bacteroides (H(B) and human Lachnospiraceae (Lachno2). - It was considered that the diameter of the overflow pipe does not completely represent the flow rate of the overflow discharge into the surface water because, in reality, pipe design is done to ensure that the maximum flow occurs within 93% of the total pipe diameter (MELCC and MAMROT, 2014). In instances, the rate of wastewater outflow fluctuates with precipitation intensity. Furthermore, the overflow pipe's diameter reflects the scale of the urban area drained upstream. - The weight assigned to the parent node "Pop" should be greater than that assigned to the node "diam". - The uncertainties on the conditional probability distributions are mainly related to flow rate of the overflow discharge.	α= 0.9 W_{Pop}= 0.9 W_{Diam}= 0.1
Microbial hazard based on overflow duration (MHO(D)	Maximum annual cumulative overflow time (MACOT)[3] Severity of consequences (S(C)	[0, 1), [1, 23), [23, 96), [96, 262), [262, +∞) Very low Low Medium High Very high	- Studies on how *E. coli* concentrations change during an overflow event have shown inconsistent findings. According to the research of Madoux-Humery et al. (2013) and Passerat et al. (2011), the *E. coli* concentration is higher at the beginning of overflow events due to leaching from the sewer systems. However, the more extended the overflow event lasts, the peak E. coli contamination decreases. Other research has shown that depending on the peak flow, *E. coli* peaks may happen later (Taghipour et al., 2019) or be randomly spread during excessively long overflow events or events caused by snowmelt mixed with precipitation (McCarthy et al., 2012). Indeed, theses variations in the outcomes are influenced by a multitude of factors, most notably the duration and intensity of rainfall, as well as the duration of the dry period preceding the overflow episodes (Madoux-Humery et al., 2015). Consequently, the uncertainties on the conditional probability distributions are mainly related to these factors.	α= 1 W_{MACOT}= 0.4 W_{CS}= 0.6
Microbial hazard based on frequency overflows (MHFO)	Annual frequency of overflows (AFO)[4] Automatic overflow frequency recorder (AOFR)[5] Severity of consequences (S(C)	[0, 1), [1, 2), [2, 5), [5, 14), [14, +∞) Yes No Very low Low Medium High Very high	- The frequency of overflows that are manually recorded may not accurately represent the microbiological hazard, because the real number of overflows could be higher when an automatic recorder is used; if there is more than one overflow in a week, they are all recorded as one event, and the duration of the overflow is not recorded. Consequently, there are more uncertainty and variability in the probability distribution of the CPT linked to the MHFO child node while using manual recorder. - A high number of overflows is not particularly associated with the highest levels of *E. coli*. However, high *E. coli* concentrations are often correlated with the extensive sewershed population that discharges upstream of the intakes (Jalliffier-Verne et al., 2016).	If the overflow structure is equipped with an automated overflow frequency recorder: α = 0.7 W_{AFO}= 0.3 W_{CS}= 0.7 Otherwise, the conditional probability distributions are adjusted based on expert judgment.
CSO microbial hazard (MH)	Microbial hazard based on overflow duration (MHO(D) Microbial hazard based on frequency overflows (MHFO)	Very low Low Medium High Very high	- Microbial hazard can be estimated with more certainty based on the overflow duration rather than the frequency. CSO frequencies do not represent contaminant loads properly (Taghipour et al., 2019). - The uncertainties on the conditional probability distributions are mainly related to the CSO microbiological load.	α= 0.5 W_{MHOD}= 0.6 W_{MHFO}= 0.4

(*continued on next page*)

Table 1 (*continued*)

Child node	Parent nodes: Attribute	Parent nodes: Status[*]	Criteria and hypothesis for determining the link between parent and child nodes	CPT filling parameters selected
Exposure of the DWI to the hazard of CSO (EXP)	Location of the DWI regarding the CSO (LDWI)[6] Distance prioritization index (DPI)[7]	ORB[8] MOR[9] SRB[10] [0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), [0.8, 1]	- Peak *E. coli* concentrations are typically higher in DWI raw water located longitudinally downstream of the overflow outfalls, — in other words on the same riverbank, compared to the intake placed opposed the outfalls (Taghipour et al., 2019). - Different CSO loading scenarios and simulated river hydrodynamic conditions demonstrated that maximum concentrations reached at drinking water treatment facilities within a few hours for near-intake overflow structures (Taghipour et al., 2019). - Position of the DWI relative to a CSO outfall locations may reflect a significant impact of higher flowrates on *E. coli* concentrations at the intake (Jalliffier-Verne et al., 2017). - The DPI is generated using the maximum and minimum distances between the drinking water intake and CSO outfalls within the same riverbank. The closer the index score is to one, the higher the microbiological hazard to water quality at the intake. - Uncertainties in the conditional probability distributions are mainly related to microbial contaminant fate and transport from CSO outfalls to the DWI.	$\alpha = 0.2$ $W_{LDWI} = 0.6$ $W_{DPI} = 0.4$
CSO risk level (RL)	CSO microbial hazard (MH) Exposure of the DWI to the hazard of CSO (EXP) DWI vulnerability level (DWI_VL)	Very low Low Medium High Very high Low Medium High	CPT filling parameter were selected based on the study conducted by Kammoun et al. (2023).	$\alpha = 1$ $W_{MH} = 0.8$ $W_{EXP} = 0.1$ $W_{DWI\_VL} = 0.1$

[*] The status of each node are listed in ascending order from the best to the worst range.

[1] The population “Pop” estimated in the UDB supplied by the combined overflow structure.

[2] Diameter (in meters) of the wastewater discharge pipe to the surface water.

[3] Maximum annual cumulative duration (in hours) of overflow events occurred during a 5-year period.

[4] Maximum annual cumulative number of overflow events occurred during a 5-year period.

[5] This node specifies whether or not the overflow structure is occupied by an automated recorder of the frequency and duration of overflows.

[6] The riverbank where the DWI is situated in regard to the overflow structure’s outfall.

[7] The distance prioritization index was determined based on the riparian distance (m) between the CSO outfall and the DWI.

[8] Opposite riverbank.

[9] Middle of the river.

[10] Same riverbank.

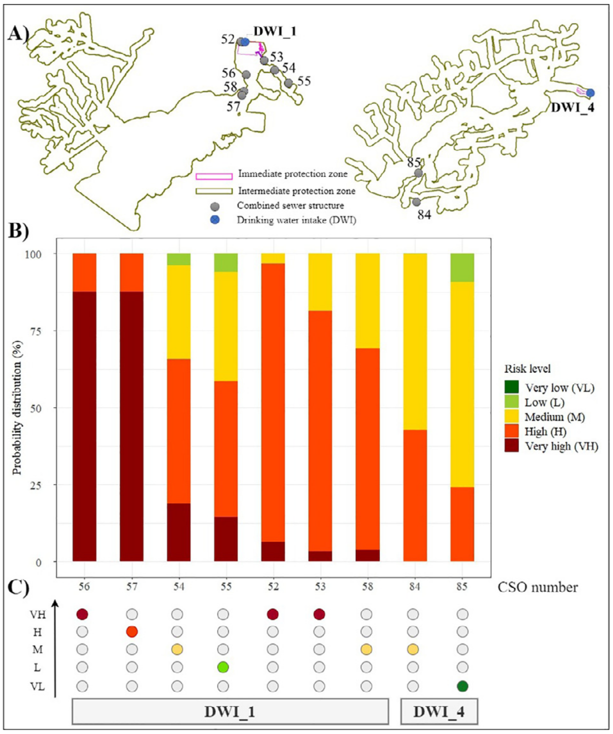

Fig. 3. Results of the risk assessment of the CSOs inventoried upstream of DWI_1 and DWI_4. (A) Maps illustrate overflow structure's locations upstream of DWIs. (B) Bar plots feature the probabilities (in%) across different risk levels determined based on the probabilistic approach. (C) The bottom plot indicates the risk levels of CSOs using the deterministic approach.

used to accurately assess the risk of the overflow structures at the study site.

3. Results

3.1. Overflows frequency and duration

Both deterministic and probabilistic approaches were applied to assess the CSO risk at the study site. Fig. 3 illustrates the results for upstream (DWI_1) and downstream (DWI_4) river intakes. Figs. 4 and 5 illustrate middle river section intakes (DWI_2 and DWI_3, respectively).

Using the probabilistic approach, CSO_2 posed the most significant risk for the study site, with a 99% probability of being at the very high risk level. This CSO outfall is located within the DWI_3′s immediate protection zone and has the highest number of recorded overflow occurrences ($n = 343$), therefore, the high risk it presents is justified. CSO_11, situated upstream of DWI_2 and DWI_3, had the longest cumulative annual overflow duration (1186 h). This also explains the very high risk probabilities of 86% and 83% regarding DWI_2 and DWI_3, respectively. The exposure to this CSO and the vulnerability of each DWI led to a 3% variation in their very high risk probabilities. The deterministic method estimated that CSO_2 and CSO_11 posed a very high risk to these two DWIs.

Using a probabilistic method, the probabilities of very high risk posed by CSO_9, CSO_36, and CSO_50, upstream DWI_2 and DWI_3, ranged from 23% to 62%. Nonetheless, the risk level was underestimated using the deterministic approach; it was stated that CSO_9 and CSO_36 posed a medium risk level and CSO_50 a high risk level. A lack of data on the overflow durations of CSO_36 and CSO_50 caused this underestimation. The deterministic assessment, in this instance, was carried out based on their overflow frequencies, regardless of their durations; CSO_36 had a medium overflow frequency, and CSO_50 a low one. The risk is also underestimated, although the frequency and duration of CSO_9 overflow events are known.

Risk levels at 13 other overflow structures (CSO_7, CSO_10, CSO_22, CSO_34, CSO_39, CSO_43, CSO_45, CSO_47, CSO_48, CSO_49, CSO_52, CSO_54, CSO_55) upstream from all DWIs were assessed using the deterministic method without overflow duration data, as they were not equipped with automatic overflow recorders. The deterministic method implies a low- to medium-frequency overflow and poses, in general, a low to medium risk to the DWI.

3.2. Hazard exposure

Over the study site, six outfalls of CSOs (CSO_1, CSO_2, CSO_3, CSO_44, CSO_52, and CSO_53) are located in the DWIs' immediate protection zones (Figs. 3–5). Due to their discharges occurring no more than 500 m upstream of the intakes, all these structures were estimated to be associated with a very high risk using the deterministic method. The Bayesian model revealed, however, that only CSO_2 upstream of DWI_3 was 99% likely to present a very high risk due to its characteristics: all the nodes AFO, MACOT, DPI, and LDWI were at their worst status. In contrast, for CSO_52, CSO_53, and CSO_3, the predominant risk level was deemed high, with probabilities of 90%, 78%, and 70%, respectively. The lower population density in their UDBs, as compared to CSO_2, seemed to be mostly responsible for these findings. Furthermore, CSO_3 and CSO_53 are positioned on the opposite riverbank as the DWI, making it uncertain that all microbiological contaminants could reach the intake. The risk associated with CSO_44 was rated as medium (60%) to low (29%) because of its low frequency of overflows (1 to 2), a small population (50 to 1000), and intake position (MOR). According to Taghipour (2019), pollutants disperse predominantly along the riverbank where they are discharged, which supports BN risk probabilities beyond the deterministic risk level. As the population parameter is unavailable, CSO_1′s risk is calculated as very high (38%), high (48%), and medium (14%) (Fig. 2). This uncertainty was related to the population variable that reflects the microbial contaminant load.

In the intermediate protection zones, CSO_44, CSO_47, CSO_48, and CSO_49 are near the DWI (DPI = 0.8 to 1) and on the same riverbank. Using the BN, it was assumed that these CSOs had a medium risk level (54 to 61%). Compared to these CSOs, CSO_32 had a high risk (88%), even though it is located on the other shore and distant from the DWI.

3.3. Lack of data

The lack of data on any of the deterministic equation's parameters makes it impossible to apply it to estimate the overflow risk level. Almost half ($n = 42$) of the study site's overflow structures were not deterministically assessed because of lack of data. Using the Bayesian network, it was feasible to evaluate these CSOs by assigning equal probability to all CSOs with missing data, therefore evaluating risk with uncertainty, as shown in Fig. 6.

4. Discussion

4.1. Prioritization of overflow structures

In this study, the assessment of microbial contamination risk levels of DWIs located downstream from CSOs discharge is conducted using deterministic and probabilistic approaches. The findings in Section 3 indicate that overflow structures with both high frequency and long duration of overflow events are often linked to a high microbial risk for DWIs, particularly when the discharge outlet is located on the same riverbank and near the intake (Sections 3.1 and 3.2). The results also showed that the deterministic approach could underestimate the level of risk due to a lack of data regarding the duration of overflow events, as was observed for CSO_36 and CSO_50. Even with information on the frequency and duration of overflows events, as in the case of CSO_9, the risk level

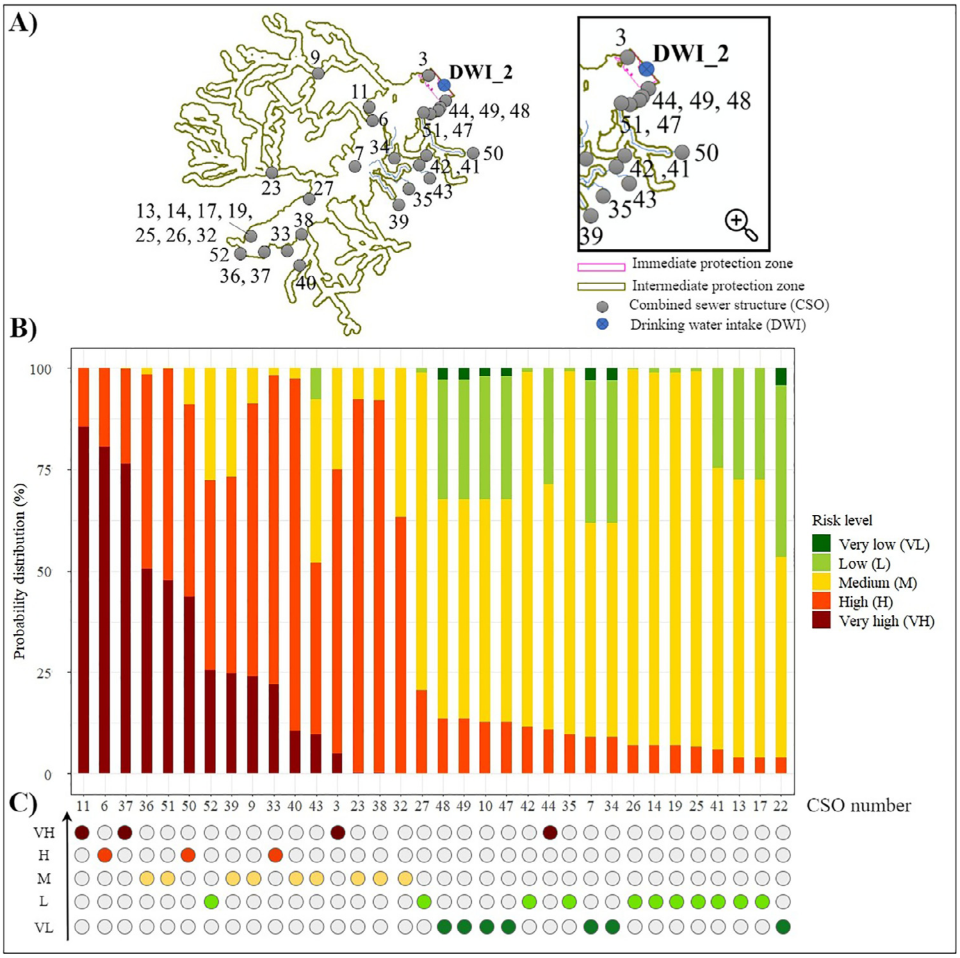

Fig. 4. Results of the risk assessment of the CSOs inventoried upstream of DWI_2. (A) Maps illustrate overflow structure's locations upstream of DWI_2. (B) Bar plots feature the probabilities (in%) across different risk levels determined based on the probabilistic approach. (C) The bottom plot indicates the risk levels of CSOs using the deterministic approach.

can still be underestimated (Section 3.1). This is because the deterministic method is not capable of handling many scenarios simultaneously; for example, a single overflow might result in several waterborne epidemics, mainly when the CSO structure serves a large population. In fact, the population estimated in the UDB of CSO_9, CSO_36, and CSO_50 ranged from high (2500 to 5000) to very high (≥ 5000). However, as explained in Table 1, the probabilistic method considers that the overflow may last for a long time, and it may alter the quality of water sources. As a result, for all these CSOs, the probabilistic method produces levels of risk were higher than that given by the deterministic method. These findings demonstrate the Bayesian model's capability to account for the overflows' inherent uncertainty as well as the uncertainties in the recorded data, particularly when employing a manual technique, giving it an advantage over the deterministic assessment.

These findings are supported by the research of Taghipour et al. (2019) demonstrating with hydrodynamic and water quality modeling that overflow frequency is not always more indicative of risk as compared to microbial contaminant load. Additionally, the World Health Organization (WHO) (World Health Organization (WHO), 2016) has shown that 13 events of microbial contamination intrusion into surface waters resulted in more than 600,000 waterborne disease outbreaks over 14 years. The probabilistic risk assessment approach, which includes uncertainty and accounts for several scenarios, yields more reliable findings than the deterministic one.

Based on the results of Section 3.2, it can be deduced that the Bayesian model prioritizes the CSO that is overflowing more and for a longer period, and it is not simply the overflow outfall's position that is of greatest significance and the only parameter of interest. The very high frequency and duration of overflows in the CSO_32 explained the high level of risk compared to other structures such as CSO_44, CSO_47, CSO_48, and CSO_49, which exhibited very low to low UDB populations.

The deterministic approach produced a point estimate risk level that may be either very low, low, medium, high, or very high using a simple equation, whereas the probabilistic approach generated distributions of probable risk levels across various risk ranges (Section 3). The point risk levels of the deterministic approach are difficult to compare as they are less informative and may give similar ratings to quantitatively distinct risks. Using DWI_3 as an example, it was assessed that 10 CSOs pose a medium risk level with 100% evidence (Fig. 4). Nevertheless, using the Bayesian model, it was feasible to identify the CSOs that require greater attention. These 10 CSOs may be sorted from most to least attention needed: CSO_36, CSO_51, CSO_39, CSO_9, CSO_40, CSO_43, CSO_43, CSO_38, CSO_3, CSO_23, and CSO_32. Results demonstrated that the other deterministic risk levels (very low, low, high, and very high) show the same pattern for all DWIs. Thus, the outputs of the deterministic approach make it challenging for managers to identify CSOs that need extensive attention when many CSO are evaluated as having the same risk level. This may be handled by using the sorted results obtained from the probabilistic model.

When implementing a source protection action plan, it is essential to identify and especially prioritize the pollution sources that cause the greatest risk to raw water. The probabilistic method allows for this, as opposed to the deterministic approach, by considering uncertainties in the risk assessment process.

4.2. Risk mitigation

The first step in safeguarding a water source is recognizing the most important risks to the quality of that water. As described in Section 4.1,

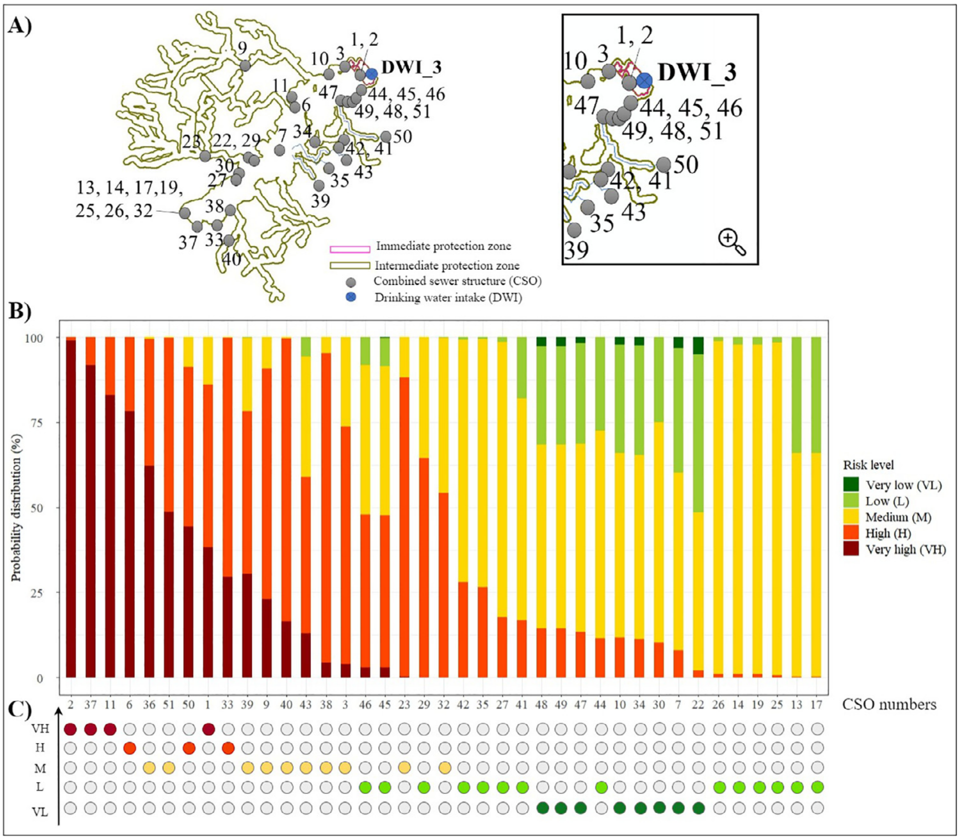

Fig. 5. Results of the risk assessment of the CSOs inventoried upstream of DWI_3. (A) Maps illustrate overflow structure's locations upstream of DWI_3. (B) Bar plots feature the probabilities (in%) across different risk levels determined based on the probabilistic approach. (C) The bottom plot indicates the risk levels of CSOs using the deterministic approach.

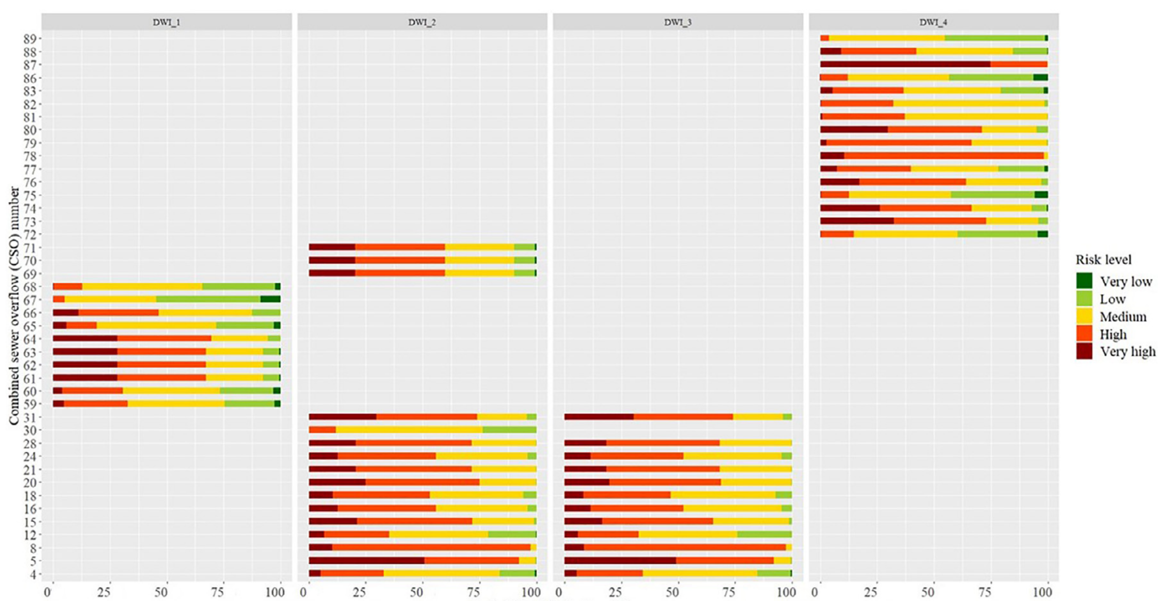

Fig. 6. Probabilistic risk assessment results for CSOs inventoried upstream of drinking water intakes (DWI 1, DWI 2, DWI 3, and DWI 4) that could not be analyzed using the deterministic approach owing to lack of data.

it was challenging to prioritize overflow structures in an urbanized watershed using the deterministic approach. However, the concept of a deterministic approach could be used at a larger scale (e.g., provincial) for a preliminary examination to evaluate the key sources of risk from a range of human activities. For higher-level risk evaluation on a smaller scale (watershe(D), a probabilistic approach based on BN is useful. This strategy could be beneficial to support water safety plans.

Water managers and stakeholders face several obstacles, including data gaps, and the complexity of interdependent components that define water resource systems (Phan et al., 2016). This can make it challenging to implement effective action plans to protect DWIs, especially using deterministic risk outcomes that do not support the lack of data as demonstrated in the results of Section 3.3. In this study, the Bayesian model addressed these issues by defining the relationship between variables that account for various types of uncertainties, including those related to data, knowledge gaps, and technical components that are related to the nature of the equipment (manual or automati(C) used to record overflow characteristics. Thus, it can be stated that a single model may be used to estimate the risk associated with overflow structures in all watersheds, despite the limited amount of data available. These findings might help managers detect problematic structures by covering several eventualities. This information may be used to make a judgment on whether to install automated recorders to fill in the gaps in data about the duration and frequency of overflows. Subsequently, the updated data might be incorporated into the BN to reevaluate the risk posed by these structures with less uncertainty. In this study, it is important to acknowledge the limitations of the constructed Bayesian model, which relies on simplifying hypotheses used to fill conditional probability tables (Table 1). These hypotheses can lead to approximate outcomes or oversimplifications of the complex reality of the assessed system. This simplification of complexity affects various elements of the study, such as dispersion and diffusion patterns, as well as the growth, reproduction, and mortality of *E. coli* during the process of spreading and diffusion to the DWI. To incorporate these elements in more detail, it would be necessary to conduct hydrodynamic and water quality modeling using a three-dimensional (3(D) model to include the physical, biochemical and biological processes.

Regulations aim at reducing the frequency of CSOs (Jalliffier-Verne et al., 2016). This goal could be achieved by various strategies such as green infrastructures, Water Sensitive Urban Design (WSU(D), improved operation practice and collection system improvements (Botturi et al., 2020; Fry and Maxwell, 2017; Fu et al., 2019; Jean et al., 2021; Ryu et al., 2015; Wong, 2006) including the separation of the combined sewers (Abbas et al., 2019). However, not all these risk mitigation options are compatible with the characteristics of the studied watershed and available budget. A feasibility evaluation is needed to select the optimal (or a combination of) options. The BN developed in this study could be extended to a Bayesian decision network (BDN) by adding decision nodes representing various risk mitigation options and utility nodes indicating the benefits of outcomes and the cost of action (Kaikkonen et al., 2020; Phan et al., 2019). This BDN could compare the costs and benefits of several options and settle on the most cost-effective strategy for mitigating the risk (Notaro et al., 2014; Penman et al., 2020).

4.3. Applicability of the Bayesian model on a wider scale

The BN model proposed in this study was developed with a particular emphasis on the hazard that CSOs represent to DWIs. The factors characterizing an overflow and its impact on water quality (such as population in the UDB, frequency and duration of overflows, pipe diameter, exposure, etc.) are similar across different countries and regions in Canada. The readily available data used to construct this Bayesian model makes it suitable for extrapolation to other major, densely urbanized watersheds, as well as watersheds with different levels of urbanization. It is possible to customize the nodes considering new data describing the characteristics of the CSOs, the DWI to be studied, and the local regulations. To further strengthen and validate this study's conclusions, future research endeavors may explore expanding this study to multiple locations. This Bayesian model could also be optimized in diverse global settings and contexts to protect other uses of surface water from CSOs, like recreational and agricultural uses, as well as to protect aquatic species and the ecological health of the receiving water.

It is possible to improve the developed Bayesian model by considering the simultaneous occurrence of overflows, because the overflow events accumulation on the same riverbank increases the contaminant concentration downstream of the DWI (Jalliffier-Verne et al., 2016). Additionally, this study's Bayesian model is deemed static because it uses the statistics of the data gathered over a period of five years without considering their evolution over time. This static BN might be made dynamic by adding additional time-dependent variables and features (Marcot and Penman, 2019) as the probability distribution may change during its lifespan (Khosravi-Farmad and Ghaemi-Bafghi, 2020). In other words, climate or global change can be integrated into this model using a dynamic BN. The main challenge of this application is integrating multiple simulations into the model, given that temporal development is coupled with a change in many variables, generating various scenarios.

5. Conclusions

Results of this paper highlight how the outcomes of the deterministic and probabilistic approaches differed:

- When all variables defining an intake's exposure to an overflow event are set to their lowest or highest values, the deterministic method yields steady and equivalent outcomes to the probabilistic approach. However, when a structure is linked to many features at various degrees, the deterministic findings generally underestimate risk compared to the probabilistic ones. Exceptionally, the deterministic approach overestimates the risk level of the CSO located in the immediate protection zone because it does not consider the microbial contaminant load presented by the population in the UDB. This inaccurate classification of risks may lead to unreliable decision making.
- Single-valued outputs from a deterministic risk assessment approach are less relevant than those depicted in a probability distribution using a probabilistic approach. A valuable aspect of the probabilistic model is its capacity to integrate probabilities to consider the uncertainty related to the intrinsic characteristics of overflows as well as external influences such as DWI vulnerability and exposure to the threat.
- In contrast to the deterministic approach, the Bayesian model enables ranking and prioritizing CSOs to support risk management at the watershed level and provides considerable assistance in establishing more effective action plans for drinking water safety.
- The assessment of risk associated with CSOs requires a substantial amount of data, which is not readily available or necessitates in-depth and sophisticated analyses based on advanced equipment that may not easily accessible to managers. Therefore, it is essential to rely on knowledge and experience, including theory, to conduct risk assessment. In this research project, this approach was facilitated by utilizing BN to handle various types of data, scientific knowledge expert knowledge, and regulatory requirements in defining variables and the probabilistic relationships among them. The BN model shows the important ability to overcome limitations that arise from a lack of data when using the deterministic method. In the field of source water protection, it is crucial to identify risk sources even with uncertainty, to plan subsequent advanced and sophisticated analyses for risk mitigation.
- The Bayesian model can make a significant contribution to effective learning process in the fields of source water protection. Sensitivity analysis help identify variables that have the greatest influence on risk levels and require further investigation. This enables targeted and systematic learning, leading to improved scientific knowledge in this field.

Funding

This work was supported by NSERC PURE CREATE and the partner municipalities.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Sarah Dorner reports financial support was provided by Partner municipalities. Raja Kammoun reports financial support was provided by NSERC PURE CREATE.

Data availability

I have shared my code in the supplementary materials.

Acknowledgments

We thank the watershed organization team responsible for the study area for contributing to the data collection and vulnerability assessments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.envc.2023.100735.

References

Abbas, A., Carnacina, I., Ruddock, F., Alkhaddar, R., Rothwell, G., Andoh, R., 2019. An innovative method for installing a separate sewer system in narrow streets. J. Water Manag. Model..

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 -608.

Aziz, M., Imteaz, M., Choudhury, T., Phillips, D., 2013. Applicability of artificial neural network in hydraulic experiments using a new sewer overflow screening device. Australasian Journal of Water Resources 17 (1), 77–86.

Birch, H., Mikkelsen, P.S., Jensen, J.K., Lutzhoft, H.C., 2011. Micropollutants in stormwater runoff and combined sewer overflow in the Copenhagen area, Denmark. Water Sci. Technol. 64 (2), 485–493.

Botturi, A., Ozbayram, E.G., Tondera, K., Gilbert, N.I., Rouault, P., Caradot, N., et al., 2020. Combined sewer overflows: a critical review on best practice and innovative solutions to mitigate impacts on environment and human health. Crit. Rev. Environ. Sci. Technol. 51 (15), 1585–1618.

Burnet, J.B., Habash, M., Hachad, M., Khanafer, Z., Prévost, M., Servais, P., et al., 2021. Automated targeted sampling of waterborne pathogens and microbial source tracking markers using near-real time monitoring of microbiological water quality. Water. 13 (15).

Calderon, O., Porter-Morgan, H., Jacob, J., Elkins, W, 2017. Bacterial diversity impacts as a result of combined sewer overflow in a polluted waterway. Glob. J. Environ. Sci. Manag. 3 (4), 437–446.

Cox Jr., L.A., 2008. What's wrong with risk matrices? Risk Anal. 28 (2), 497–512.

Dirckx, G., Vinck, E., Kroll, S., 2022. Stochastic determination of combined sewer overflow loads for decision-making purposes and operational follow-up. Water 14 (10).

El Ghazouli, K., El Khatabi, J., Soulhi, A., Shahrour, I., 2022. Model predictive control based on artificial intelligence and EPA-SWMM model to reduce CSOs impacts in sewer systems. Water Sci. Technol. 85 (1), 398–408.

Elmasry, M., Hawari, A., Zayed, T., 2017. Defect based deterioration model for sewer pipelines using Bayesian belief networks. Can. J. Civ. Eng. 44 (9), 675–690.

Fry, T.J., Maxwell, R.M., 2017. Evaluation of distributed BMPs in an urban watershed—High resolution modeling for stormwater management. Hydrol. Process. 31 (15), 2700–2712.

Fu, X., Goddard, H., Wang, X., Hopton, M.E., 2019. Development of a scenario-based stormwater management planning support system for reducing combined sewer overflows (CSOs). J. Environ. Manag. 236, 571–580.

Gasperi, J., Zgheib, S., Cladiere, M., Rocher, V., Moilleron, R., Chebbo, G., 2012. Priority pollutants in urban stormwater: part 2 - case of combined sewers. Water Res. 46 (20), 6693–6703.

Goulding, R., Jayasuriya, N., Horan, E., 2012. A Bayesian network model to assess the public health risk associated with wet weather sewer overflows discharging into waterways. Water Res. 46 (16), 4933–4940.

Grêt-Regamey, A., Straub, D., 2006. Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS. Nat. Hazards Earth Syst. Sci. 6 (6), 911–926.

Grafton, R.Q., Little, L.R., 2017. Risks, resilience, and natural resource management: lessons from selected findings†. Nat. Resour. Model. 30 (1), 91–111.

Hahn, M.A., Palmer, R.N., Merrill, M.S., Lukas, A.B., 2002. Expert system for prioritizing the inspection of sewers: knowledge base formulation and evaluation. J. Water Resour. Plan. Manag. 128 (2), 121–129.

Hajj-Mohamad, M., Darwano, H., Duy, S.V., Sauve, S., Prevost, M., Arp, H.P.H., et al., 2017. The distribution dynamics and desorption behaviour of mobile pharmaceuticals and caffeine to combined sewer sediments. Water Res. 108, 57–67.

Hansen K.D., Gentry J., Long L., Gentleman R., Falcon S., Hahne F., et al., 2022. Package 'Rgraphviz': provides plotting capabilities for R graph objects. Version 2.41.1.

Harris, M.J., Stinson, J., Landis, W.G., 2017. A Bayesian approach to integrated ecological and human health risk assessment for the South River, Virginia Mercury-Contaminated Site. Risk Anal. 37 (7), 1341–1357.

Hojsgaard, S., 2012. Graphical independence networks with the gRain Package for R. J. Stat. Softw. 46 (10), 1–26.

Jalliffier-Verne, I., Heniche, M., Madoux-Humery, A.-.S., Galarneau, M., Servais, P., Prevost, M., et al., 2016. Cumulative effects of fecal contamination from combined sewer overflows: management for source water protection. J. Environ. Manag. 174, 62–70.

Jalliffier-Verne, I., Leconte, R., Huaringa-Alvarez, U., Heniche, M., Madoux-Humery, A.-.S., Autixier, Laurène, et al., 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., Morin, C., Duchesne, S., Pelletier, G., Pleau, M., 2021. Optimization of real–time control with green and gray infrastructure design for a cost-effective mitigation of combined sewer overflows. Water Resour. Res. 57 (12).

Kaikkonen, L., Parviainen, T., Rahikainen, M., Uusitalo, L., Lehikoinen, A., 2020. Bayesian networks in environmental risk assessment: a review. Integr. Environ. Assess. Manag. 17 (1), 62–78.

Kammoun R., McQuaid N., Bichai F., Goitom E.A., Lessard V., Prévost M., et al., 2023. Risk assessment of drinking water intake contamination from agricultural activities using Bayesian Network (Article submetted). PLOS Water Unpublished results:45.

Khosravi-Farmad, M., Ghaemi-Bafghi, A., 2020. Bayesian decision network-based security risk management framework. J. Netw. Syst. Manag. 28 (4), 1794–1819.

Locatelli, L., Russo, B., Acero-Oliete, A., Sánchez-Catalán, J.C., Martínez-Gomariz, E., Martínez, M, 2020. Modeling of <i>E. coli</i> distribution for hazard assessment of bathing waters affected by combined sewer overflows. Nat. Hazards Earth Syst. Sci. 20 (5), 1219–1232.

Madoux-Humery, A.S., Dorner, S., Sauve, S., Aboulfadl, K., Galarneau, M., Servais, P., et al., 2013. Temporal variability of combined sewer overflow contaminants: evaluation of wastewater micropollutants as tracers of fecal contamination. Water Res. 47 (13), 4370–4382.

Madoux-Humery, A.S., Dorner, S.M., Sauve, S., Aboulfadl, K., Galarneau, M., Servais, P., et al., 2015. Temporal analysis of E. coli, TSS and wastewater micropollutant loads from combined sewer overflows: implications for management. Environ. Sci. Process. Impacts 17 (5), 965–974.

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

Marcot, B.G., Penman, T.D., 2019. Advances in Bayesian network modelling: integration of modelling technologies. Environ. Model. Softw. 111, 386–393.

McCarthy, D.T., Hathaway, J.M., Hunt, W.F., Deletic, A., 2012. Intra-event variability of Escherichia coli and total suspended solids in urban stormwater runoff. Water Res. 46 (20), 6661–6670.

McGinnis, S.M., Burch, T., Murphy, H.M., 2022. Assessing the risk of acute gastrointestinal illness (AGI) acquired through recreational exposure to combined sewer overflow-impacted waters in Philadelphia: a quantitative microbial risk assessment. Microb. Risk Anal. 20.

McQuaid N., Madoux-Humery A.-.S., Touttée J.-.M., Dorner S., Prévost M., 2019. Technical Sheet No. 2: Potential risk assessment of wastewater overflows (DEU) (Fiche Technique N°2: évaluation du potentiel de risque associé aux débordements d'eaux usées (DEU)). Montreal, Quebec: Polytechnique Montreal.

MELCC and MAMROT, 2014. Stormwater management guide: Development Strategies, Design Principles and Best Management Practices for Urban Drainage Systems (Guide de gestion des eaux pluviales: Stratégies d'aménagement, principes de conception et pratiques de gestion optimales pour les réseaux de drainage en milieu urbain). Quebec, Canada.

MELCC, 2021. Monitoring of the exploitation of the municipal wastewater treatment plants (OMAEU): Wastewater Treatment Plant and Overflow Works (Suivi d'exploitation des ouvrages municipaux d'assainissement des eaux usées (OMAEU): Station d'épuration et ouvrages de surverse. Quebec, Canada.

Mounce, S.R., Shepherd, W., Sailor, G., Shucksmith, J., Saul, A.J., 2014. Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. Water Sci. Technol. 69 (6), 1326–1333.

Norsys Software and Corp. Netica, Version 6.09: norsys; 2021 Available from: https://www.norsys.com/netica.html.

Notaro, V., Fontanazza, C.M., Loggia, G.L., Freni, G., 2014. In: Identification of the Best Flood Retrofitting Scenario in an Urban Watershed by Means of a Bayesian Decision Network, 139. WIT Press, p. 12.

Olds, H.T., Corsi, S.R., Dila, D.K., Halmo, K.M., Bootsma, M.J., McLellan, S.L., 2018. High levels of sewage contamination released from urban areas after storm events: a quantitative survey with sewage specific bacterial indicators. PLoS Med. 15 (7), e1002614.

Passerat, J., Ouattara, N.K., Mouchel, J.M., Rocher, V., Servais, P., 2011. Impact of an intense combined sewer overflow event on the microbiological water quality of the Seine River. Water Res. 45 (2), 893–903.

Peace, C., 2017. The risk matrix: uncertain results? Policy Pract. Health Saf. 15 (2), 131–144.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, United States, p. 552 340 Pine Street, Sixth Floor.

Penman, T.D., Cirulis, B., Marcot, B.G., 2020. Bayesian decision network modeling for environmental risk management: a wildfire case study. J. Environ. Manag. 270, 110735.

Petrie, B., 2021. A review of combined sewer overflows as a source of wastewater-derived emerging contaminants in the environment and their management. Environ. Sci. Pollut. Res. Int..

Phan, T.D., Smart, J.C.R., Capon, S.J., Hadwen, W.L., Sahin, O., 2016. Applications of Bayesian belief networks in water resource management: a systematic review. Environ. Model. Softw. 85, 98–111.

Phan, T.D., Smart, J.C.R., Stewart-Koster, B., Sahin, O., Hadwen, W.L., Dinh, L.T., et al., 2019. Applications of Bayesian networks as decision support tools for water resource

management under climate change and socio-economic stressors: a critical appraisal. Water. 11 (12).
Prévost M., Madoux-Humery A.S., Dorner S., 2017. Protection measures for surface water withdrawals for human consumption: protection areas and source vulnerability - literature review (Mesures de protection des prélèvements d'eau de surface effectués à des fins de consommation humaine : aires de protection et vulnérabilité des sources-Revue bibliographique.), p. 101, Polytechnique Montreal, Quebec.
Quebec Government, 2022a. Water Withdrawal and Protection Regulation- Q-2, r. 35.2. Quebec Official Publisher, Quebec, Canada.
Quebec Government, 2022b. In: Regulation Respecting Municipal Wastewater Treatment Works - chapter Q-2, r, 34. Quebec Official Publisher, Quebec, p. 1.
Rosin, T.R., Romano, M., Keedwell, E., Kapelan, Z., 2021. A Committee Evolutionary Neural Network for the Prediction of Combined Sewer Overflows. Water Resources Management 35 (4), 1273–1289.
Ryu, J., Baek, H., Lee, G., Kim, T.H., Oh, J., 2015. Optimal planning of decentralised storage tanks to reduce combined sewer overflow spills using particle swarm optimisation. Urban Water J. 14 (2), 202–211.
Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. J. Stat. Softw. 35 (3), 1–22.
Shan, K., Shang, M., Zhou, B., Li, L., Wang, X., Yang, H., et al., 2019. Application of Bayesian network including Microcystis morphospecies for microcystin risk assessment in three cyanobacterial bloom-plagued lakes, China. Harmful Algae 83, 14–24.
Sojobi, A.O., Zayed, T., 2022. Impact of sewer overflow on public health: a comprehensive scientometric analysis and systematic review. Environ. Res. 203, 111609.
Stritih, A., Rabe, S.E., Robaina, O., Gret-Regamey, A., Celio, E., 2020. An online platform for spatial and iterative modelling with Bayesian Networks. Environ. Model. Softw. 127.
Sylvestre, E., Burnet, J.B., Smeets, P., Medema, G., Prevost, M., Dorner, S., 2020. Can routine monitoring of E. coli fully account for peak event concentrations at drinking water intakes in agricultural and urban rivers? Water Res. 170, 115369.
Sylvestre, E., Prevost, M., Burnet, J.B., Smeets, P., Medema, G., Hachad, M., et al., 2021. Using surrogate data to assess risks associated with microbial peak events in source water at drinking water treatment plants. Water Res. 200, 117296.
Taghipour, M., Shakibaeinia, A., Sylvestre, É., Tolouei, S., Dorner, S., 2019a. Microbial risk associated with CSOs upstream of drinking water sources in a transboundary river using hydrodynamic and water quality modeling. Sci. Total Environ. 683, 547–558.
Taghipour, M., Tolouei, S., Autixier, L., Prevost, M., Shakibaeinia, A., Dorner, S., 2019b. Normalized dynamic behavior of combined sewer overflow discharges for source water characterization and management. J. Environ. Manag. 249, 109386.
Taghipour, M., Shakibaeinia, A., Sylvestre, E., Tolouei, S., Dorner, S., 2019c. Microbial risk associated with CSOs upstream of drinking water sources in a transboundary river using hydrodynamic and water quality modeling. Sci. Total Environ. 683, 547–558.
Taghipour M., 2019. Characterization of CSO microbial contamination and their risks to drinking water sources. Montral, Canada: Civil, Geological and Mining Engineering Department, Polytechnique Montreal.
Vatanpour, S., Hrudey, S.E., Dinu, I., 2015. Can public health risk assessment using risk matrices be misleading? Int. J. Environ. Res. Public Health 12 (8), 9575–9588.
Vijayashanthar, V., Qiao, J., Zhu, Z., Entwistle, P., Yu, G., 2018. Modeling fecal indicator bacteria in urban waterways using Artificial Neural Networks. J. Environ. Eng. 144 (6).
Wijesiri, B., Deilami, K., McGree, J., Goonetilleke, A., 2018. Use of surrogate indicators for the evaluation of potential health risks due to poor urban water quality: a Bayesian Network approach. Environ. Pollut. 233, 655–661.
Wong, T.H.F, 2006. An overview of water sensitive urban design practices in Australia. Water Pract. Technol. 1 (1).
World Health Organization (WHO), 2016. Protecting surface water for health. Identifying, assessing and managing drinking-water quality risks in surface-water catchments. Geneva. Report No.: ISBN 978 92 4 151055 4.
Yu, R., Zhang, C., 2021. Early warning of water quality degradation: a copula-based Bayesian network model for highly efficient water quality risk assessment. J. Environ. Manag. 292, 112749.
Zhang, D., Lindholm, G., Ratnaweera, H., 2018. Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. J. Hydrol. 556, 409–418.
Zou, X., Yue, W.L., 2017. A Bayesian network approach to causation analysis of road accidents using netica. J. Adv. Transp. 2017, 1–18.