Strategic sensor placement for the identification of disinfection byproducts from chlorinated drinking water: Case study in the water distribution network of Coimbra, Portugal

Aristotelis Magklis[1] and Andreas Kamilaris[1]

[1] CYENS Center of Excellence, Nicosia, Cyprus.

Abstract. Providing clean drinking water to people worldwide is a challenging task due of the contaminants that tend to form in water distribution systems at any point in time, causing various health impacts to the consumers. An emerging contaminant with not well-known effects on human body is the family of disinfection byproducts (DBPs) occurring in chlorinated drinking water. Due to the complexity of their formation, it is very challenging to monitor and detect DBPs in a water network. By placing sensors throughout the distribution system, it becomes feasible to monitor certain environmental parameters that affect their formation and identify locations in the network which have high probabilities for the detection of DBPs. The problem is complicated due to the large number of possible DBPs and the constrained number of sensors available, thus a clever approach needs to be followed, especially in cases where prioritization of DBPs needs to be considered. In this work, we propose a methodology for detecting DBPs, considering a real-world case study of a calibrated water distribution network in Coimbra, Portugal. Different algorithms are investigated to satisfy different performance objectives, aiming to locate the best possible locations for sensor placement, towards maximizing possible detection and minimizing overall impact of DBPs.

Keywords: Disinfection Byproducts, Sensor Placement, Algorithms, Water Distribution Network, Chlorinated Drinking Water.

1 Introduction

Water quality is essential to civilian daily life. Detecting contaminants in water distribution networks before they reach into households is vital in order to reduce the number of people exposed to health risks due to polluted water. In order to detect these contaminants quickly, a common practice is to install sensors strategically inside a water distribution network (WDN), to monitor environmental parameters that can lead to the formation of these contaminants. However, due to very large WDNs and the complex nature of the problem, considering that numerous contaminants can be formed inside a WDN, the problem of detecting those contaminants using proper sensor placement is very difficult to solve.

The most common chemical disinfectant that is used for water treatment is chlorine. Disinfectant byproducts (DBPs) constitute an emerging category of contaminants, formed when chlorine is used for the disinfection of drinking water. When chlorine interacts with natural organic matter (NOM) [1], then it is likely for a variety of DBPs to be formed. More than 700 DBPs have been reported to date. DBPs are generally grouped into three main categories: (i) aliphatic, (ii) alicyclic and (iii) aromatic [2]. Some examples of NOM include humic substances and fulvic acids, that can impact the color, taste and odor of drinking water. Furthermore, many NOMs are results of human activities, such as polycyclic and heterocyclic aromatic hydrocarbons which are formed due to industrial processes, agricultural runoff and wastewater discharge [3]. Apart from various types of NOMs, other parameters can affect the creation of DBPs, such as temperature, turbidity, pH, topology, temperature as well as the increasing human population, which leads to higher needs for purified drinking water [4]. Being able to detect these DBPs early would allow to minimize the human population exposed to contaminated drinking water. The minimization of the time of detection can lead to fast and drastic measures by water operations before the contaminated water reaches the consumer. The long-term impact of these DBPs is severe and can create significant health risks, such as genotoxicity and cytotoxicity [5]. Genotoxicity can damage the DNA, which can result in mutation and increased risk of cancer. On the other hand, cytotoxicity can harm or kill cells, lead to tissue damage and organ failure. The precise effects of DBPs on human body are still under investigation.

Identifying DBPs formed in chlorinated drinking water using sensors is an issue that has not yet been well studied because of the complexity of the formation of DBPs, as well as the difficulty of obtaining an accurately calibrated model of a WDN. Moreover, as previously mentioned, having hundreds of DBP compounds makes it hard to prioritize certain DBPs over others without having historical data of monitored environmental parameters, to understand which DBPs may form at certain locations or conditions inside a water network. Another challenge is the fact that usually operators of WDNs have limited sensory equipment available, thus this equipment needs to be used as efficiently as possible. It is important to note the difference between the type of sensory equipment available: low-end sensors monitor only relevant environmental parameters, such as pH, temperature, conductivity and more. On the other hand, high-cost sensors can directly identify DBPs, such as electrochemical sensors [6], gas and liquid chromatography and mass spectrometry [7], and others. Being able to perform strategic sensor placement of both low-cost and high-cost sensors is a challenge in many WDN cases.

In this paper, a calibrated model of an existing WDN is utilized for addressing the problem of strategic placement of low-end sensors for the detection of DBPs formed in chlorinated drinking water. The contribution of this paper is the proposal of a novel methodology for addressing this problem, employing various techniques related to prioritization of DBPs, modelling of a WDN and use of algorithms for solving the optimization problem. The novelty of this paper lies in the proposed methodology to address a complex challenge which is still generally poorly investigated by the scientific community.

2 Related Work

Several studies have explored the impact of DBPs on human health as well as their formation, examining which environmental parameters are the most relevant. Most literature regarding chlorinated DBPs focuses on the categorization of the families of DBPs that can be identified in an WDN. In order to estimate the formation and fate of DBPs, predictive models have been developed and analyzed in scientific literature. These models can assist in the decisions that need to be taken by the drinking water industry [8]. The prediction of the concentration of DBPs in the water distribution networks is based on the effects of different water quality and operational parameters in controlling DBP formation, under different environmental conditions [9].

Literature regarding strategic sensor placement for the detection of DBPs formed in chlorinated drinking water is limited due to the complexity of the subject, as well as the difficulty to understand their formation. The optimal placement of water quality sensors has been extensively studied in [10], [11], focusing on indirect detection of disinfection byproducts by means of monitoring relevant environmental parameters, instead of actual DBP concentration in the network.

Nevertheless, many studies have been conducted utilizing different algorithms for sensor placement in WDNs for detecting other contaminants such as microorganisms, inorganic and organic chemicals, etc. [12]. For example, strategic placement of sensors for the detection of Escherichia coli and organophosphates was investigated in [13] and [14] respectively. Some literature focuses on the intelligent sensor placement for the detection of leakages based on the changes found in the flow of the water in the network [15]. Moreover, multiple experiments were conducted using specific field sensors for the monitoring of environmental parameters that help in the formation of DBPs, such as chlorine [16], pH [17], water temperature [18] and more. The most important parameter that is hard to detect is NOM, however advances have been made to steadily monitor and characterize it [19]. In some cases, the optimal sensor placement needs to take into consideration topological and connectivity features of the WDN [20]. The most common algorithms for addressing the problem of sensor placement include heuristic approaches, genetic algorithms [21], the NSGA-II algorithm [22], or mixed-integer programming techniques [23].

3 Methodology

The methodology for approaching the challenge of strategic sensor placement in WDNs for detection of DBPs formed in chlorinated drinking water is depicted in Figure 1.

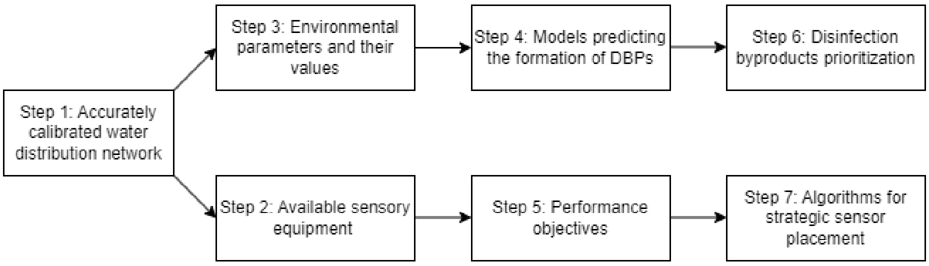

Fig. 1. Methodology

Foremost, the WDN under study needs to be modeled in a digital form. Then, the environmental parameters that affect the formation of DBPs need to be selected, and some or all of these parameters need to be monitored, to extract their values inside the WDN. Such data is valuable to assess the seasonal and spatial variation of the values of the relevant environmental parameters, which in turn affect the likelihood of forming DBPs inside the WDN. In the following sub-sections, the different steps of the general methodology proposed will be analyzed in more detail, such as prioritization of the families of disinfection byproducts depending on the available environmental parameters, algorithms that could be harnessed to solve the strategic placement problem and performance objectives tackled for the early detection of DBPs.

3.1 Step 1: Water distribution network modelling

Access to an accurately calibrated water distribution system is very important. Most research papers consider hypothetical networks and many assumptions [12], [22], and it is difficult to assess whether their proposed algorithms work in a real-world setting. Collecting available environmental data from water utility operators is vital, to understand the range of possible values, especially as the chlorine diffusion inside the network and its concentration at each node/junction inside the pipe system (see also Step 3). For example, having higher doses of chlorine in the water network leads to the formation of trihalomethanes in a faster pace in the water distribution system [24]. If the samples contain smaller doses of chlorine, it is more likely that haloacetic acids and haloacetonitriles are present in the WDN [25].

3.2 Step 2: Sensory equipment

Afterwards, it is important to consider the resources available for sensory equipment (number and type of sensors available: low-cost vs. high cost), examine the size and type of the network, the locations where sensory equipment can be placed inside the network, etc. There are many sensors that can assist in the monitoring of the environmental parameters, as well as analyzers that directly measure the DBPs found in the drinking water. However, depending on the nodes that must be covered, a decision needs to be made depending on the number of sensors that must be placed throughout the network and the cost of each sensor. Sensory equipment can be either low-end or high-end sensors: low-end sensors monitor only relevant environmental parameters (e.g. chlorine, pH, temperature, conductivity) while high-cost sensors can directly detect DBPs.

3.3 Step 3: Environmental parameters and their values

There are multiple environmental parameters that assist in the formation of DBPs. These parameters include: (i) bromide, (ii) dose of disinfectant, (iii) water temperature, (iv) potential of hydrogen (pH), (v) composition and concentrations of NOM, and (vi) chlorine. As previously mentioned, the most common treatment for disinfecting water is by using disinfectants like chlorine. Depending on the amount of chlorine concentrated in the network, different families of DBPs are formed. Governing organizations have set safety standards for the acceptable amount of disinfectant dosage, to balance the amount of disinfection needed with the potential of harmful DBPs being formed. Regarding pH, a neutral range is vital (6.5 to 8.5), to minimize the formation of DBPs. Higher ranges (9.0 – 10.0) reduce the reaction of the chlorine in the system, while lower ranges (5.0 – 6.0) help in the formation of haloacetic acids [26]. The water temperature varies in values significantly in WDNs. Generally, higher temperatures in the water distribution system can increase the reaction rate between the disinfectant and the natural organic matter, leading to an increase in the formation of disinfection byproducts [27]. Lastly, turbidity is a measure of level of particles, such as sediment, or organic by-products in the water [28]. The compliance for turbidity standards is below 1 Nephelometric Turbidity unit (NTU), in order for water utility operators to provide safe drinking water.

3.4 Step 4: Models predicting formation of DBPs

Probabilistic models use certain criteria to predict, for example, chlorine concentrations in the network, or DBP concentrations depending on the parameters available for input. For the development of these models, machine learning (ML) methods are used, such as back propagation artificial neural networks (BPANNs) [29], Random Forests (RFs) [30], support vector machines (SVMs) [31] and many more [32]. Prediction models can also be utilized for identifying concentrations of DBPs in the network. Depending on the DBP family under focus, different parameters that assist in the formation of relevant DBP compounds are used as inputs, such as pH, temperature, dissolved organic carbon and more.

3.5 Step 5: Performance objectives

Performance metrics need to be selected wisely, and they directly relate to the decisions about sensor placement. Based on literature, there are two approaches to handle the strategic sensor placement: either with a single performance or multi-performance objective approach. The aim can be: a) to maximize the coverage of the network based on the selected number of sensors, b) to minimize the time of detection of the contaminants; and c) to reduce the infected water consumed by the consumers. This leads to a multi-performance objective approach, with the main focus being the minimization of time of detection, in order to effectively react fact to any potential dangers that might occur in the network. Most of the metrics mentioned benefit from one another, since maximizing the coverage allows the minimization of time of detection at the same time.

3.6 Step 6: Disinfection byproducts prioritization

To effectively place sensors for disinfection byproducts, it is of outmost importance to highlight which families of DBPs are targeted by the sensors. For example, if the WDN has higher dosages of chlorine, then trihalomethanes are the DBPs targeted for monitoring. Another factor to consider when prioritizing DBPs, is the longevity of each contaminant in the network. For example, trihalomethanes can survive longer in the pipes of the network, while, on the other hand, haloacetonitriles have very low longevity [33]. Considering the longevity of the DBPs, compounds that have higher lifetime should be prioritized. Lastly, data for certain families of DBPs is available, depending on the region where the water distribution is located, and the legal restrictions that have been applied for maintaining an average value of chlorine concentration in the network. Moreover, each family of disinfection byproducts has different long-lasting effects on the human health, making it another factor to consider when prioritizing compounds for sensor placement [2]. Lastly, assuming it is possible to have data for each node at specific times throughout the day for the network, we categorize the possible families of DBPs that can be formed in the network (trihalomethanes, haloacetic acids, haloacetonitriles) and which precursors are dominant (e.g. total organic carbon, dissolved organic carbon, higher temperature of water). Kalita et al. [2] attempted a prioritization of DBPs based on a range of relevant factors.

3.7 Step 7: Algorithms for strategic sensor placement

After the identification of the performance metrics that need to be tackled, the selection of the algorithm which will solve the strategic placement problem needs to be decided. As previously mentioned, the most common algorithms for addressing the problem of sensor placement include heuristic approaches, genetic algorithms [21], the NSGA-II algorithm [22], or mixed-integer programming techniques [23]. There can be two approaches when examining the issue of strategic sensor placement. The empirical approach and the optimization-based approach. The empirical approach does not consider the hydraulic data of a network, instead it focuses on certain criteria that are vital for the placement of the sensors. For example, nodes with high numbers of consumers nearby, points in the network that can alter the quality of water instantly (source nodes) and accessible nodes where installing the equipment is feasible. On the other hand, the optimization-based approach, focuses on the implemented performance objectives, such as time of detection, population exposed to contamination, maximum coverage, thus aiming to detect the most optimal locations for sensor deployment. This approach requires a calibrated model, to examine the network thoroughly using the hydraulic data provided and run different simulations and what-if scenarios to discover the best possible algorithm. Moreover, for smaller networks mixed-integer programming techniques are more effective due to their higher precision while, for larger networks, heuristic approaches can be used due to their faster execution, with small penalties in accuracy.

4 Implementation

We have implemented the methodology described in Section 3 using an accurately calibrated model of a WDN, at the civil parish Ameal e Arzila in Coimbra, Portugal (see Figure 1).

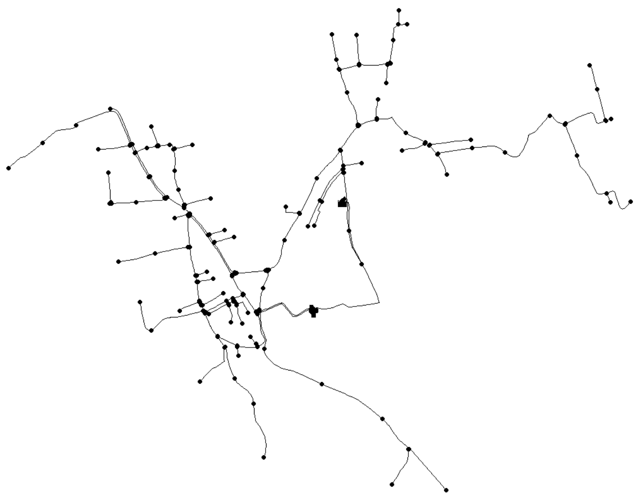

Fig. 2. Water distribution network of civil parish Ameal e Arzila in Coimbra, Portugal

The WDN provided by the utility operators is of moderate size, includes 227 nodes, a tank and a reservoir and is categorized as a dead-end network. Portugal is characterized by higher temperatures during the summer season, and relatively low temperatures during the winter season. Higher doses of chlorine are required to be added to the network when the temperatures are rising, so the formation of trihalomethanes in the network is favored. The environmental parameters that are monitored in the WDN under study are: the dose of disinfectant, water temperature, pH and turbidity. In our case study, the pH is in the neutral range between 6.5 to 8.5. Lastly, the turbidity value is steadily below 1.0, meaning that the water is purified, thus the levels of turbidity do not assist in the formation of harmful contaminants in the drinking water [28]. These range of values have been provided to us by the local operator.

Using the software tool EPANET [34], it was possible to examine the chlorine concentration over time at each node. One of the first tests for simulating DBPs in the network was by applying the 2R bulk-reaction model using the extension of EPANET, EPANET-MSX. The environmental parameters considered are: chlorine dosage added to the network through the tank and the reservoir, as well as water temperature. Without having ground-truth data, it was impossible to include more environmental parameters and characterize their values. The DBP families of trihalomethanes [35] and haloacetic acids [36] were considered in this simulation, due to the lack of available data for DBPs' historical occurrence from the water utility operators. The EPANET-MSX model, which simulates the concentration of DBPs in the network, has two reactants, one fast and one slow after the initial dosing of chlorine [37]. This simulation creates an output file which contains the ID of each node, the timestep at each step of the simulation (e.g. report of chlorine concentration and DBP concentration every 15 minutes).

The two algorithms utilized in this case study are: mixed-integer programming (MIP), and the greedy randomized adaptive search procedures (GRASP), For the MIP algorithm, a Python model has been developed based on two Python frameworks: (i) Chama [38], which specializes in the placement of the sensors and (ii) WNTR [39], a framework specialized for simulating and analyzing resilience of water distribution networks. Moreover, for the GRASP algorithm, the Threat Ensemble Vulnerability Assessment and Sensor Placement Optimization Tool (TEVA-SPOT) [40] was utilized, that allows the simulation of many different scenarios for any contaminant the user wants to create, strategic sensor placement, planning utility response to sensor detection and the design of the basis threat.

The performance metrics of maximum coverage and time of detection were selected, to analyze the trade-offs between these performance metrics, due to their correlation. Having a larger number of sensors in the network automatically decreases the time of detection, due to the coverage. The minimization of mass consumption of polluted water is also another performance metric considered. The prioritization of the DBPs was based on the region of the WDN and the data available. The region has a history of formation of trihalomethanes in most networks in Portugal, this is why trihalomethanes are better regulated than other families of DBPs. Then, the effects of each compound on the consumers as well as its longevity in the network were considered. Thus, the selected family we focused on was trihalomethanes, with a focus on the following compounds: chloroform (CF), Dibromochloromethane (DBCM) and their brominated counterparts. Our prioritization strategy is aligned with the prioritization exercise performed in [2].

Another important aspect of strategic placement is the type of sensors deployed, as previously mentioned. A range of sensors was considered during experimentation, assuming various scenarios of available resources and budget for sensory equipment by the water utility operator.

The equations selected for the DBP predictive models developed are listed in Table 1. The reasoning behind the selection of these equations is their high accuracy and the wealth of environmental parameters involved. The selection of these equations is based on an extensive literature review we performed. The equations selected are based on the family of DBPs and not on individual compounds (e.g. chloroform), since the simulations of the strategic placement focus on the selection of the major DBP families.

Table 1. Equations for the detection of disinfection byproducts.

DBP	Example equation	R^2	Reference
THMs	THM = $10^{-0.038} \times (Cl2)^{0.654} \times (pH)^{1.322} \times (time)^{0.174} \times (SUVA)^{0.712}$	0.88	Uyak et al. [41]
HAA9	HAA9 (μg / L) = −345 + 1.695(Temperature) + 93.1(pH) − 226(UVA254) + 4.95(Cl2) + 5.66(NO−2 −N) + 16.6(DOC) + 0.325(NH+4 −N) − 0.0693(Temperature)^2 − 6.41(pH)^2 + 190821(UVA254)^2 − 1.73(NO−2 −N)^2 −3.77(DOC)^2 − 0.01663(NH+4 −N)^2	0.811	Okoji et al. [42]
HANs	T-HANs = $10^{-1.065} (Br)^{0.346}(DOC)^{0.369} \times (Cl2/DOC)^{0.520} (t)^{0.238} (Temp)^{0.373}$ × (R^2 = 0.943, p<0.0005, n = 36)	0.943	Hong et al. [43]

Through the simulation, the names of the nodes/junctions that have been selected for sensor placement are outputted to an excel file, in order to compare the results of the two algorithms.

Regarding the MIP algorithm, for each node, up to one sensor is assigned, in respect to the number of sensors available. Since this algorithm has a binary nature, each option is either chosen (1) or not (0). Based on the defined constraints, this algorithm aims to provide an optimal or near optimal solution regarding the strategic sensor placement on a provided WDN. To satisfy the constraints, the following formula was implemented:

$$\sum_{a \in A} \alpha_a \sum_{i \in L_a} d_{ai} x_{ai} \tag{1}$$

$$\sum_{i \in L_a} x_{ai} = 1 \; \forall \, a \in A \tag{2}$$

$$x_{ai} \le s_i \; \forall \, a \in A, i \in L_a \tag{3}$$

$$\sum_{i \in L} c_i s_i \le p \tag{4}$$

$$s_i \in \{0, 1\} \forall \, i \in L \tag{5}$$

$$0 \le x_{ai} \le 1 \; \forall \, a \in A, i \in L_a \tag{6}$$

The formulation is based on the p-median facility location problem and the goal is to minimize objective (1) subject to constraints (2 to 6), where A is the set of all scenarios, L is the set of all candidate locations, L_a is the set of all scenarios capable of detecting scenario α, α_a is the probability of the scenario α happening, d_{ai} is the impact assessment, x_{ai} is an indicator variable that will be 1 if sensor i is installed, s_i is a binary variable that will either be 1 if sensor i is installed or 0 otherwise, c_i is the cost of the sensor i and lastly, p is the sensor budget, meaning the number of sensors available for deployment.

Regarding the GRASP algorithm, the greedy randomized adaptive search procedure is used to solve optimization problems. It combines elements of greedy algorithms and randomization and through iterations it aims to identify the best possible solutions. A candidate list is constructed with all the best possible solutions (in our case, the selection of nodes depending on the performance objective defined) and the ranking of the list is based on the parameters defined (time, mass consumed, etc.). The selection of the solution however is randomized, to avoid selecting only local optima solutions (e.g. multiple nodes near the injection location). This process is repeated with respect to the provided budget (number of sensors) and each iteration starts with a new randomized solution.

5 Evaluation

5.1 Maximum coverage

For the performance objective of maximum coverage, both the MIP and the GRASP algorithm were used, to identify the optimal locations for the potential detection of DBPs. The Pareto front in Figure 2 showcases the results of the maximum coverage objective, including the minimization of time of detection at the same time. After a certain percentage of the network has been covered (e.g. 30% to 35%) by sensors, the effectiveness of further increasing the number of sensors is very low. The same procedure was conducted using the GRASP algorithm to determine any major changes in the locations of the sensors between the two methods. Only eight nodes selected were different in the two approaches, without a big distance between the locations selected, as the next node in the network was selected instead. Thus, the two algorithms had similar behavior and results.

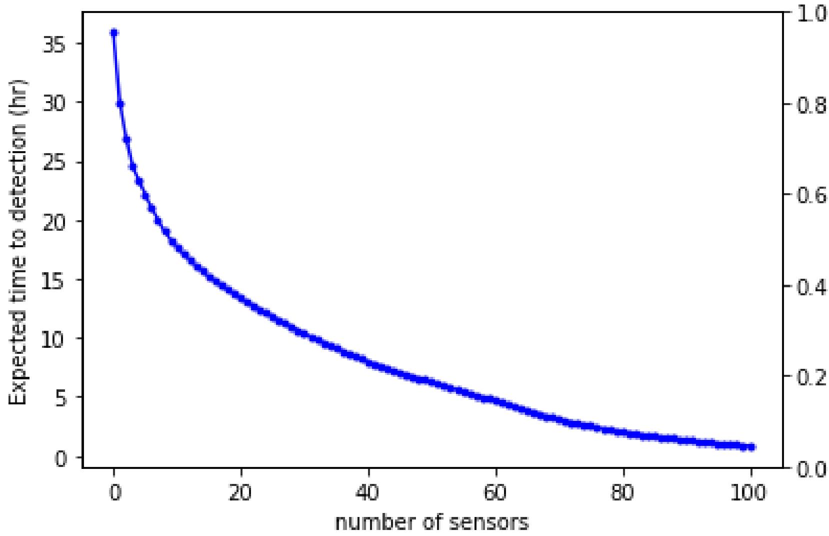

Fig 3. Pareto front for maximum coverage of the network and minimization of time of detection.

5.2 Minimization of time of detection and mass consumption

A more realistic number of 10 sensors, assuming lower-cost sensors and considering the size of the network, was used for examining the minimization of time of detection and reduction of mass consumption. For the prioritization of the disinfection byproducts family, the diameter size of the pipes was used as a parameter for the sensor placement, to consider the spatial variation for the formation of haloacetic acids and haloacetonitriles [44]. The results are depicted in Figure 4, with the left image showing the sensor placement prioritizing nodes with higher concentrations of chlorine, while the right image includes the diameter size of the pipes as a parameter for the placement.

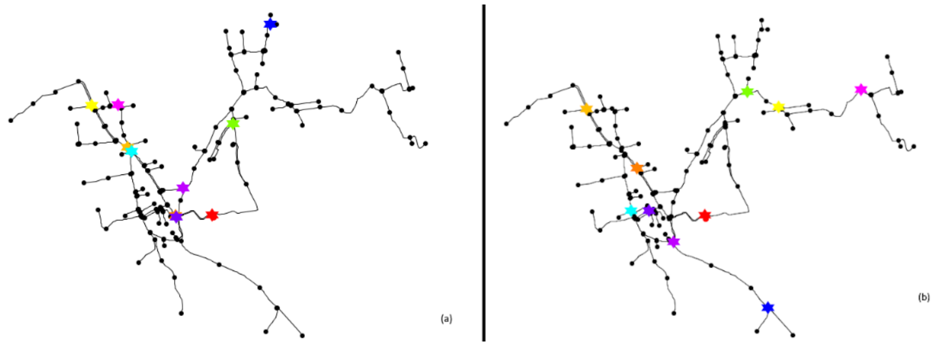

Fig 4. (a) Sensor placement of 10 sensors targeting trihalomethanes, (b) Sensor placement of 10 sensors focusing on haloacetonitriles and haloacetic acids. Injection of chlorine begins from tank and reservoir.

For the selection of nodes where spatial variation is prioritized, locations closer to the dead-end nodes are preferred. These nodes have a lower concentration of chlorine, which as previously mentioned, benefits in the formation of haloacetic acids and haloacetonitriles.

5.3 Randomized injections in the network

To further examine optimal sensor placement, different simulations were conducted based on the possibility of injecting chlorine in different locations inside the WDN. For example, injections of chlorine were performed at the south side of the network and at the east side of the network at the same time. Figure 5 shows the differences in placement depending on the performance objective, left image focusing on the time of detection and the right image focusing on the minimization of mass consumption.

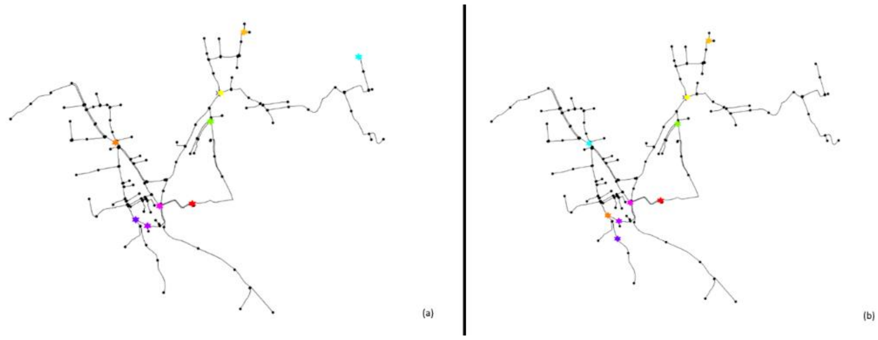

Fig 5. (a) Minimization of time of detection, (b) Minimization of mass consumption. Both scenarios are based on 10 sensors in the network.

Another scenario included a set number of sensors (10) and a completely randomized injection in the network, in order to compare the results of the two algorithms regarding the optimal solutions they provide for minimization of time of detection. In Figure 6, the left image showcases the selection of 10 sensors utilizing the MIP algorithm and the right image shows the results based on the GRASP heuristic approach.

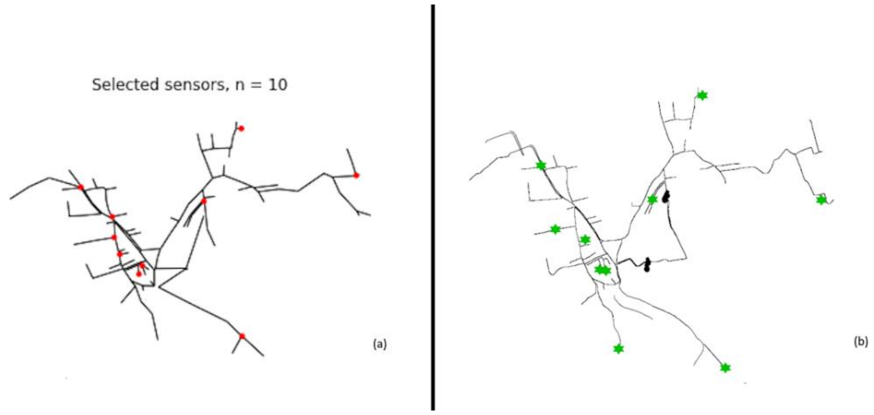

Fig 6. (a) Minimization of time of detection (MIP algorithm), (b) Minimization of time of detection (GRASP algorithm). Both algorithms consider the availability of 10 sensors.

6 Discussion

The case study at Coimbra and the methodology followed helped to understand the number of sensors that should be placed inside the network, the best candidate locations, as well as the trade-offs between budget/cost (i.e. purchase of sensory equipment) and efficiency, based on the performance metrics used. Based on the results showcased, the optimal number of sensors for the performance metric of maximum coverage would be 80 sensors, to cover 35% of the network and 60 sensors to minimize effectively the time of detection of the contaminants, as shown in Figure 3. Regarding these two objectives, the related work in [22] shows similar results following this optimization-based approach for sensor placement.

Due to the fact that the network provided is of moderate size, running the simulations alone is enough to provide high-confident results about chlorine concentrations in different nodes of the network, i.e. the chlorine diffusion algorithm works well. Additionally, having data related to the population of the network was very important in order to tackle the performance metric of minimizing the mass consumption effectively.

6.1 Limitations

In order to achieve better results in such a complex subject, more data needs to be gathered from the water utility operators at a daily basis. The environmental parameters that relate to the formation of DBPs need to be monitored consistently, so methods such as machine learning can also be implemented to have a functioning model to predict more precisely the family that can be formed at each node of the network.

Furthermore, the installation of these sensors in the nodes in the system is sometimes a challenging task. The equipment needs to be located somewhere where it can be easily maintained and be safe from any dangers and to also have access to power and electricity. These restrictions make it almost impossible to place sensors at every node in the system. Most of the limitations relate to the fact that data is not available regarding the environmental parameters that can assist in the development of predictive models as well as create better simulations and what-if scenarios.

6.2 Future work

As future work, constant monitoring of the environmental parameters that assist in the formation of DBPs will be performed inside the pilot site, to extract useful data that will help to develop better prediction models. These data and models will then allow algorithms and simulations, such as the ones used in this paper, to achieve more realistic and optimized sensor placement. Regarding the prioritization of the DBPs in the network for the placement of the sensors, more research needs to be conducted regarding the longevity, the health risks associated at each DBP family or even at compound level. Having such data easily accessible can assist in the field of strategic sensor placement with diverse conditions in different water networks. Prioritization of DBPs needs to be embedded inside the constraint formulas when considering integer programming approaches, such as the ones we employ. The same goes for differentiating between low-end and high-end sensor devices, which is also an important aspect of future work.

Moreover, we plan to deploy both low- and high-end sensor devices inside the actual WDN of Coimbra, Portugal, using the approach and algorithms developed in this paper as guidance for strategic placement. This is part of the H20FORALL project [45], and the pilot will run for 12 months.

Finally, we intend to apply our methodology in other pilot sites as well, in the near future.

7 Conclusion

In this work, we proposed a methodology for strategic placement of sensors for disinfection byproducts formed from chlorinated drinking water, using a real-world water distribution network in Coimbra, Portugal as a case study. The results of our simulations varied depending on the family of disinfection byproducts prioritized, as well as on the number of sensory equipment available for placement. Different algorithms were utilized to compare different scenarios based on different performance objectives. Our results showed minimal changes to the placement problem when trying different algorithms or performance metrics. We have highlighted some of the challenges involved, such as the lack of data publicly available from actual water distribution networks, which could assist in the development of predictive models, as well as the shortage of accurately calibrated models. By having public data available, for both the water distribution networks and the impact of the disinfection byproducts on human health, it would be feasible to simulate more realistic scenarios in existing, water distribution networks, allowing higher accuracy in the placement based on the ground-truth data provided. This would facilitate better policy- and decision-making in the future, tackling successfully this emerging problem of potential contamination.

Acknowledgment

Funded by the European Union, under the Grant Agreement GA101081953 attributed to the project H2OforAll Innovative Integrated Tools and Technologies to Protect and Treat Drinking Water from DBPs. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

1. Plewa, M.J. et al.: Chemical and Biological Characterization of Newly Discovered Iodoacid Drinking Water Disinfection Byproducts. 38, 18, (2004).
2. Kalita, I. et al.: Assessing the Health Impact of Disinfection Byproducts in Drinking Water. 4, (2024).
3. Tak, S., Vellanki, B.P.: Natural organic matter as precursor to disinfection byproducts and its removal using conventional and advanced processes: state of the art review. Journal of Water and Health. 16, 5, 681–703 (2018).
4. Moreno-Andrés, J. et al.: Operational and environmental factors affecting disinfection byproducts formation in ballast water treatment systems. Chemosphere. 232, 496–505 (2019).
5. Wagner, E.D., Plewa, M.J.: CHO cell cytotoxicity and genotoxicity analyses of disinfection by-products: An updated review. J Environ Sci (China). 58, (2017).
6. Zhang, W. et al.: Recent Advances on Electrochemical Sensors for the Detection of Organic Disinfection Byproducts in Water. ACS Sensors. 4, 5, 1138–1150 (2019).
7. Yang, M. et al.: Current methods for analyzing drinking water disinfection byproducts. Current Opinion in Environmental Science & Health. 7, 98–107 (2019).
8. Sadiq, R., Rodriguez, M.J.: Disinfection by-products (DBPs) in drinking water and predictive models for their occurrence: a review. Sci Total Environ. 321, 1, (2004).
9. Chowdhury, S. et al.: Models for predicting disinfection byproduct (DBP) formation in drinking waters: a chronological review. 407, 14, (2009).

10. Łangowski, R. et al.: An optimised placement of the hard quality sensors for a robust monitoring of the chlorine concentration in drinking water distribution systems. Journal of Process Control. 68, 52–63 (2018).
11. Waeytens, J. et al.: Adjoint-based numerical method using standard engineering software for the optimal placement of chlorine sensors in drinking water networks. Environmental Modelling and Software. 92, 229–238 (2017).
12. Rathi, S., Gupta, R.: Sensor Placement Methods for Contamination Detection in Water Distribution Networks: A Review. 89, (2014).
13. Weickgenannt, M. et al.: Risk-Based Sensor Placement for Contaminant Detection in Water Distribution Systems. Journal of Water Resources Planning and Management. 136, 6, 629–636 (2010).
14. Ohar, Z. et al.: Optimal sensor placement for detecting organophosphate intrusions into water distribution systems. Water Research. 73, 193–203 (2015).
15. Sophocleous, S. et al.: A Two-stage Calibration for Detection of Leakage Hotspots in a Real Water Distribution Network ☆. 186, (2017).
16. Wilson, R. et al.: Continuous Chlorine Detection in Drinking Water and a Review of New Detection Methods. Johnson Matthey Technology Review. 63, 2, 103–118 (2019).
17. Maineult, A. et al.: Detection of advected concentration and pH fronts from self-potential measurements. Journal of Geophysical Research. 110, (2005).
18. Qin, Y. et al.: Integrated water quality monitoring system with pH, free chlorine, and temperature sensors. Sensors and Actuators B-chemical. 255, 781–790 (2018).
19. Chen, W., Yu, H.-Q.: Advances in the characterization and monitoring of natural organic matter using spectroscopic approaches. Water Research. 190, 116759 (2021).
20. Perelman, L., Ostfeld, A.: Short communication: Topological clustering for water distribution systems analysis. Environmental Modelling and Software. 26, 7, 969–972 (2011).
21. Casillas, M.V. et al.: Optimal sensor placement for leak location in water distribution networks using genetic algorithms. 13, 11, (2013).
22. Diao, K. et al.: Sensor placement in water distribution networks using centrality-guided multi-objective optimisation. (2023)
23. Berry, J.W. et al.: Sensor Placement in Municipal Water Networks with Temporal Integer Programming Models. 132, 4, (2006).
24. Valdivia-Garcia, M. et al.: Climatic, Geographic and Operational Determinants of Trihalomethanes (THMs) in Drinking Water Systems. 6, 1, (2016).
25. Rodriguez, M.J. et al.: Formation and fate of haloacetic acids (HAAs) within the water treatment plant. 41, 18, (2007).
26. Hung, Y.C. et al.: pH effect on the formation of THM and HAA disinfection byproducts and potential control strategies for food processing. Journal of Integrative Agriculture. 16, 12, 2914–2923 (2017).
27. Xue, R. et al.: Evaluation of thirteen haloacetic acids and ten trihalomethanes formation by peracetic acid and chlorine drinking water disinfection. Chemosphere. 189, 349–356 (2017).
28. Liu, Z. et al.: Enhanced coagulation of low-turbidity micro-polluted surface water: Properties and optimization. Journal of Environmental Management. 233, 739–747 (2019).
29. Pérez-Beltrán, C.H. et al.: Artificial intelligence and water quality: From drinking water to wastewater. (2024).
30. Shin, H. et al.: Development of water quality prediction model for water treatment plant using artificial intelligence algorithms. Environmental Engineering Research. (2023).

31. Wang, D. et al.: Model predictive control for chlorine dosing of drinking water treatment based on support vector machine model. Desalination and Water Treatment. 173, 133–141 (2020).
32. Li, L. et al.: Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. Chemical Engineering Journal. 405, 126673 (2021).
33. Pavelic, P. et al.: Fate of disinfection by-products in groundwater during aquifer storage and recovery with reclaimed water. 77, 4, (2005).
34. Farina, G. et al.: Using EPANET for modelling water distribution systems with users along the pipes. Civil Engineering and Environmental Systems. 31, 1, 36–50 (2014).
35. Sathasivan, A. et al.: Trihalomethane species model for drinking water supply systems. Water Research. 184, 116189 (2020).
36. Premarathna, S.M. et al.: Model for halo-acetic acids formation in bulk water of water supply systems. Science of The Total Environment. (2022).
37. Fisher, I.H.: Integrated EPANET-MSX process models of chlorine and its by-products in drinking water distribution systems. Water Environment Research. 95, 12, e10949 (2023).
38. Overview — Chama 0.2.0 documentation, https://chama.readthedocs.io/en/latest/overview.html, last accessed 2024/08/29.
39. Haxton, T.: Case Study Applications of Water Network Tool for Resilience (WNTR). (2022).
40. USEPA/TEVA-SPOT-Toolkit, https://github.com/USEPA/TEVA-SPOT-Toolkit, last accessed 2024/08/29.
41. Uyak, V. et al.: Multiple linear regression modeling of disinfection by-products formation in Istanbul drinking water reservoirs. Science of The Total Environment. 378, 3, 269–280 (2007).
42. Okoji, A.I. et al.: Performance evaluation of artificial intelligence with particle swarm optimization (PSO) to predict treatment water plant DBPs (haloacetic acids). Chemosphere. 140238 (2023).
43. Hong, H. et al.: Using regression models to evaluate the formation of trihalomethanes and haloacetonitriles via chlorination of source water with low SUVA values in the Yangtze River Delta region, China. Environmental Geochemistry and Health. 38, 6, 1303–1312 (2016).
44. Soliyev, A.R. et al.: Spatio-temporal variability of halogenated disinfection by-products in a large-scale two-source water distribution system with enhanced chlorination. Journal of Hazardous Materials. 423, 127113 (2022).
45. H2oforall, https://h2oforall.eu, last accessed 2024/09/10.