



EPiC Series in Engineering

Volume 3, 2018, Pages 1841–1850

HIC 2018. 13th International
Conference on Hydroinformatics



Pre-conditioning approach to Bayesian Decision Network for water quality sensors positioning in urban drainage systems

Mariacrocetta Sambito¹, Cristiana Di Cristo², Gabriele Freni¹, Angelo Leopardi², Claudia Quintiliani²

¹ School of Engineering and Architecture, University of Enna “Kore”, Cittadella Universitaria, 94100 Enna, Italy

² Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, Cassino 03043, Italy

mariacrocetta.sambito@unikore.it

Abstract

In the last decades, the growth of mini- and micro-industry in urban areas has produced an increase in the frequency of xenobiotic polluting discharges in drainage systems. Such pollutants are usually characterized by low removal efficiencies in urban wastewater treatment plants and they may have an acute or cumulative impact on environment. In order to facilitate early detection and efficient containment of the illicit intrusions, the present work aims to develop a decision-support approach for positioning the water quality sensors. It is mainly based on the use of a decision-making support of the BDN type (Bayesian Decision Network), specifically looking soluble conservative pollutants, such as metals. In the application and result section the methodology is tested on two sewer systems, with increasing complexity: a literature scheme from the SWMM manual and a real combined sewer.

1 Introduction

In the water sector, the problem of polluting source identification was mainly investigated regarding pressurized distribution networks respect to sewers. A drinking water contamination event determines an immediate alarm for public health. On the other hand, an abnormal discharge into the sewer system has usually a lower impact on the general public even if it has its relevance on the environment. The quality of wastewater acts both on the proper functioning of the sewer system and of the wastewater treatment plant (WWTP) and on the receiving water body in case of combined sewer overflow (CSO)

activation. Since the adoption of the Water Framework Directive 2000/60/EC, Member States must apply local measures to address pollution affecting their surface waters; thus, decreasing the occurrence of overflows and improving discharged water quality are important parts of pollution reducing strategies.

Boenne et al. (2014) showed that the contribution of WWTP overflows can produce up to 22% of the measured nutrient load in a river due to just two events of a few hours. *Jiang et al. (2015)* asserted that the identification of polluting sources after river spill occurred is critical to improve decision-making on emergency response to sudden water pollution. For this reason, models for the characterization of the wastewater have been widely studied for assessing the pollution load overflowed and/or transferred to WWTP.

Montserrat et al. (2015) developed a methodology to evaluate the performance of combined sewer systems (CSS) using low-cost monitoring; in this case, the collection and analysis of real data is indispensable to assess, improve, and maintain CSSs in order to reduce the number and impact of overflows. *Irvine et al. (2011)* have implemented a project with municipalities in Western New York State to evaluate the low-cost options for illicit discharge trackdown. In *Vezzano et al. (2014)*, the potential for including water-quality in real time control (RTC) schemes was unveiled, providing a further option to urban water managers to improve the performance of their systems.

Furthermore, because CSOs contain untreated domestic and industrial waste, toxic materials, and debris, they impact the physicochemical, biological, hydraulic, and aesthetic status of receiving water bodies. For example, overflows can result in oxygen depletion, increased turbidity, and higher concentrations of micropollutants, heavy metals, and pathogenic and fecal organisms in surface waters (*Passerat et al., 2011*). The xenobiotic substances, unlike the organic ones, are only slightly affected by biological degradation processes. In particular, metals often show a remarkable tendency for bio-accumulation. As such substances are often unaffected by common waste water treatment plant technologies, an efficient contamination prevention strategy should involve the identification and elimination of illicit intrusions, especially in urban drainage systems. Therefore, the implementation of a chemical monitoring network is necessary to promptly detect the event of contamination. The installation and maintenance cost can be reduced optimizing the position of the sensors so obtaining, at the same time, a reliable and cheap monitoring infrastructure.

Freni & Sambito (2017) proposed a probabilistic approach to the positioning of water quality sensors in urban drainage networks, showing the progressive increase in identification probability obtained through the Bayesian approach. They carried out 2 tests, with and without pre-conditioning; the pre-conditioning approach was based on the study of *Banik et al. (2017a)* and showed an improvement of results with more efficiency in term of computational efforts. However, the proposed pre-conditioning approach was mainly influenced by network topology. Following the past literature, the present work aims to improve it, inserting new information beyond network topology, i.e. a grey information from commercial/industrial activities inventory.

The methodology is applied to two different networks with increasing size and complexity: the literature network Example 8 of the SWMM application manual (*Gironàs et al., 2009*) and the real test-case represented by the sewer system Massa Lubrense (Italy). The study is mainly aimed to the solution of a sensors location problem, assuming the positioning of fixed-type sensors. Contrarily to what was done in previous works, the proposed approach was also evaluated considering the inclusion in the network of groups of sensors. Initially the analysis was carried out making the hypothesis that each node of the network has the same probability of being the polluting source, successively the hypothesis that some nodes can be more frequently polluted than others is introduced. A further objective is to show the adaptability of Bayesian approach in sensors location.

2 Materials and Methods

By using the Bayesian approach, new information, coming from the analysis, is incorporated in the approach allowing the operator to gain insight on the system once new contamination events are detected and identified. In this way, the approach is suitable for solving problems in which data are initially piecemeal and the operator plans to improve the monitoring strategy.

For the solution of this problem, two main components are required: a calibrated model for hydraulic and water quality simulations in sewer systems and a Bayesian solver for likelihoods estimation and probability update.

In this case, the EPA SWMM model was used to perform the hydraulic and water quality simulations and a decision-making support of the Bayesian Decision Network (BDN) type was implemented for the positioning the water quality sensors.

A Bayesian network (BN) is a graphical structure that allows us to represent an uncertain domain. The nodes represent a set of random variables: $X = X_1, \dots, X_n$, from the domain. A set of directed arcs connects pairs of nodes, $X_i \rightarrow X_j$, represent the direct dependencies between variables. There are at least three distinct forms of uncertainty which an intelligent system operating in anything like our world shall need to cope with: ignorance, the limits of our knowledge lead us to be uncertain about many things, physical randomness or indeterminism and vagueness.

Bayesianism is the philosophy that asserts that in order to understand human opinion as it ought to be, constrained by ignorance and uncertainty, the probability calculus is the single most important tool for representing appropriate strengths of belief. In fact, the probability calculus allows to represent the interdependencies which other systems require, but also allows to represent any dependencies. Therefore, the BN is a very robust and particularly useful method for assessing risk and uncertainty, providing a complete framework for analyzing all cause and effect relationships (*Korb & Nicholson, 2010*). Bayesian approaches starts from the formulation of prior knowledge in terms of probability of events to be representative of the truth (in the presented case, the probability that a sensor or a group of sensors to correctly identify the source of contamination). After that the system is solicited and investigated in order to obtain a series of events able to confirm or deny prior assumptions (in the presented case, a set of simulated or real events of contamination in which sensors were in place). The number of events considered sufficient to update prior knowledge (population of the update) is a parameter in the Bayesian approach. After this number is reached, the posterior probability is calculated incorporating new information in prior knowledge by means of the application of the Bayes' theorem. The number of updates for which additional information does not significantly implement previous knowledge is also a parameter of the approach and it depends on the complexity of the analyzed problem and case study.

The problem of sensor location for identifying the illicit intrusion for the considered case-study, has been already investigated in *Banik et al. (2017b)* but using genetic algorithms. In particular, different single and multi-objective optimization procedures to optimally locate sensors in the sewer have been compared. The procedures adopt different objective functions combined with different optimization algorithms. The multi-objective ones are solved with the NSGA-II algorithm, while for the single objective ones a Greedy Algorithm is adopted. The results indicate that good performances can be obtained using Greedy Algorithms. In all cases the obtained sensor displacement is more efficient, compared with the existing one established without any optimization. In 2015, *Banik et al. (2015)*, introduced a pre-screening approach to reduce the number of nodes that should be considered as possible source, solving a contamination source identification problem in urban drainage systems.

Kim et al. (2013) aimed our study at developing a decision-support model for identifying the location of the pathogenic intrusion as a means of facilitating rapid detection and efficient containment, using Artificial Neural Networks (ANNs). The results showed that ANNs identified the location of the injection sites with 57% accuracy; ANNs correctly recognized eight out of fourteen expressions with

relying on data from a single detection sensor. Increasing the available sensors within the basin significantly improved the accuracy of the simulation results (from 57% to 100%).

In *Freni & Sambito (2017)* without pre-conditioning, initially, all nodes had equal probability to be the candidate nodes for sensor placement and all nodes had equal probability to be an illicit contamination source. As a consequence, the contamination events are randomly simulated in order to evaluate the probability of each sensor to identify the source of contamination. Set a number of updates, the BDN tool is applied recursively to update sensor locations based on the results obtained from the previous run of the decision tool.

Applying a pre-conditioning initialization based on network topology, some nodes had higher initial probability to be selected for sensor positioning. As a result, nodes characterized by higher passing flows had more probability to be selected as candidates for sensor positioning. The study demonstrated that such a preconditioning is able to reduce computational time to converge towards an optimal sensor configuration even if approach efficiency strictly depends on the network and on the characteristics of the contamination (duration and magnitude of contamination load).

The analysis was carried out on a literature example and on a real case study using SWMM model as simulation tool for analysing hypothetical contamination events. Each contamination event is simulated by a random mass of contaminant (ranging from 10g to 500g) constantly injected in a node for a random time (ranging from 15 min to 3 hours).

In the present sensors location exercise, two of the previous limiting hypotheses were removed:

- Contamination probability is no more equal in all network nodes, but some of them have higher probability to be object of an illicit intrusion;
- The pre-conditioning approach is based not only on network topology but considering also data regarding commercial and industrial activities in the area.

The considered contaminant is conservative xenobiotic, such as, heavy metals. The analysis was initially run in wet weather conditions in which deliberate discharge of pollutants more likely occurs.

The literature example, i.e. the network Example 8 presented in the EPA SWMM reference manual, is a combined sewer network serving an area of 0.12 km² and consists of 31 nodes, 29 pipes and a pump. According to the problem previously exposed, the chosen network has two outputs: the wastewater treatment plant and the overflow. As it is reasonable to assume that the nodes downstream of commercial/industrial activities may have a greater probability of being subject to illegal spill of contaminants into the sewers, the cumulative probability distribution function of contamination, reported in Figure 1, was not considered uniform. Three nodes (J4, J10 and J118) were hypothetically considered to host industrial activities having double probability to present illicit contamination with respect to the others. This hypothesis does not modify the general applicability of the method in cases in which industrial nodes may change in number, contamination probability or location.

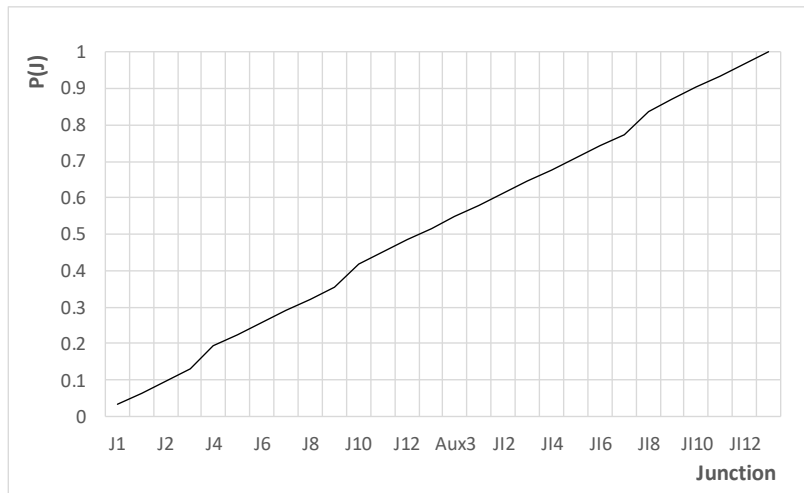


Figure 1: Non-uniform cumulative distribution function of contamination probability for Example 8 network.

The real case study is the sewer of Massa Lubrense (Figure 2), a town located near Naples, Italy. It is a combined sewer system, covering an area of 19.71 km², divided into 12 sub-catchments, serving a population of 14,087 (year 2011). The network consists of 1,909 circular conduits connecting 1,902 junctions, 14 pumps, 14 storage units, and 1 treatment plant. The daily average values of the dry weather flows in the 1866 input nodes are estimated considering the population connected to each one. The input file for the SWMM model has been calibrated using discharge measurements. The system has 12 monitoring stations, which displacement was decided on the basis of practical considerations.

Similar to the literature example, the analysis was carried out making the hypothesis that the area in red in Figure 2 has double probability to be subjected to contamination than any other node in the network. The area was selected for demonstrating the response of the proposed methodology.

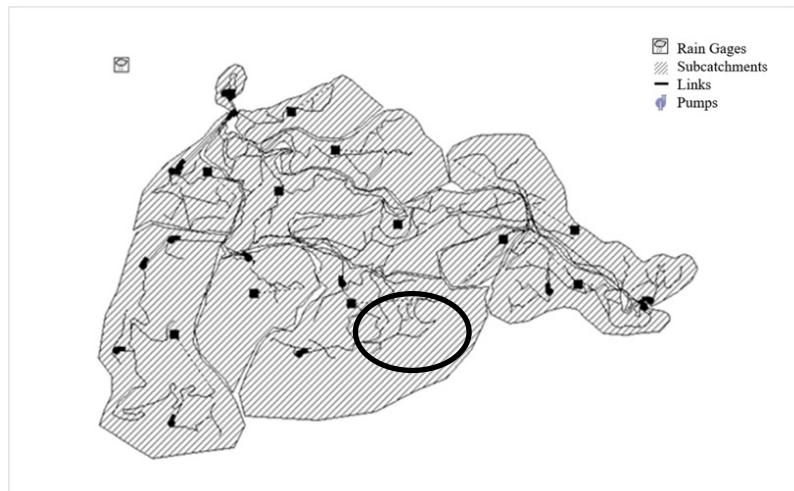


Figure 2: The Massa Lubrense network with the indication (circle) of the area having double probability of being the source of contamination.

3 Results and Discussion

Initially, the Bayesian analysis was carried out on the Example 8 network considering the possible implementation of 1, 2 or 3 sensors. Initially any node in the network has the same probability to be the location of a sensor. In a second step, prior knowledge was based on information derived from the location of industrial activities and the most probable path of contaminants propagation. Such information was obtained through a statistical analysis of 50 random simulations. The analysis was carried out by means of Monte Carlo Simulations in which a single source of contamination was located in one of the nodes of the network according to the contamination probability presented in the previous paragraph (Figure1).

BDN parameters such as the number of Bayesian updates and the population of each update were modified to understand their impact on the analysis (Table 1), especially with respect to the minimum number of simulations useful to highlight the most efficient configuration of sensors. In this way, it is possible to understand how the value of BDN parameters can affect the selection of most relevant nodes for sensor placement and the number of model simulations needed to obtain a stable configuration.

As example, Figure 3 shows the graphical representation of prior knowledge and the subsequent updates for the placement of 1 sensor for the case with 50 updates and 10 simulations. From the results obtained, the BDN configuration that allows to select more efficiently (in terms of Bayesian probability) sensors location refers to 50 Bayesian updates, each populated by 10 simulations. Increasing the Bayesian updates to 100 the improvement is limited. The cases with 50 and 100 Bayesian updates agree in the selection of the following optimal sensor locations:

- 1 sensor: nodeJI10,
- 2 sensors: nodes J11 and J11
- 3 sensors: nodes J11, J15 and JI10.

The cases with only 10 Bayesian updates agree on the same nodes only if population is 500 or higher.

Example 8				
n° sensors	N° of Bayes updates	Pop. of updates	Highest sensor posterior probability	Prior probability of each sensors combination
1	10	100	0.10	0.034
2	10	500	0.0068	0.0025
3	10	5000	0.00062	0.00027
1	10	10	0.1315	0.034
2	10	10	0.0076	0.0025
3	10	10	0.00068	0.00027
1	50	10	0.16	0.034
2	50	10	0.0085	0.0025
3	50	10	0.00075	0.00027
1	100	10	0.178	0.034
2	100	10	0.0085	0.0025
3	100	10	0.00077	0.00027

Table 1: Results in terms of bayesian probability in the various BDN configurations (Example 8) without using prior knowledge.

Starting from this result, the probability that the sensors positioned in those junctions intercept the contamination is equal to 47% with one sensor, 78% with two sensors and 84% inserting three sensors.

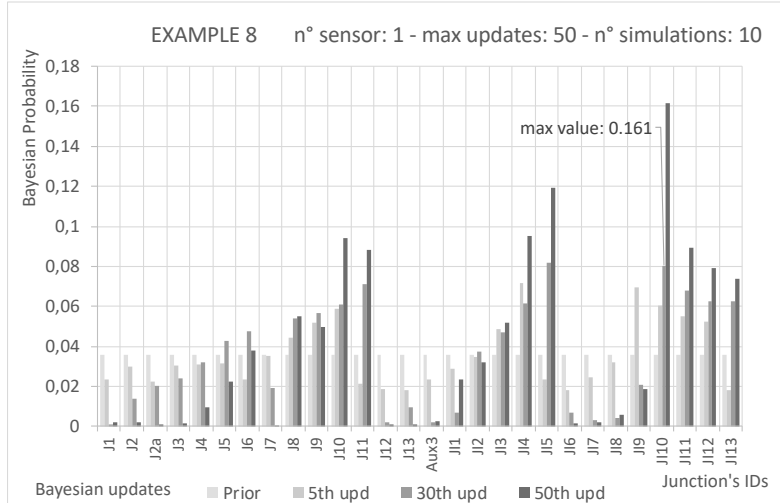


Figure 3: Graphical representation of the Bayesian probability for Example 8 network starting without an informative prior distribution.

The use of a non-uniform informative “a priori” sensor location distribution allows for saving simulation time. Figure 4 shows the graphical representation of prior knowledge and the subsequent updates for the placement of 1 sensor for the case with 50 updates and 10 simulations for the same example. This second analysis shows the same results of the previous in term of optimal sensor location so demonstrating that the use of different prior distributions does not affect sensor maximum efficiency in polluting source location. The use of informative prior distributions helps reducing the number of simulations needed to reach a stable optimal configuration (after 15 updates only, Bayesian probabilities are equal to the 50th update of the first analysis).

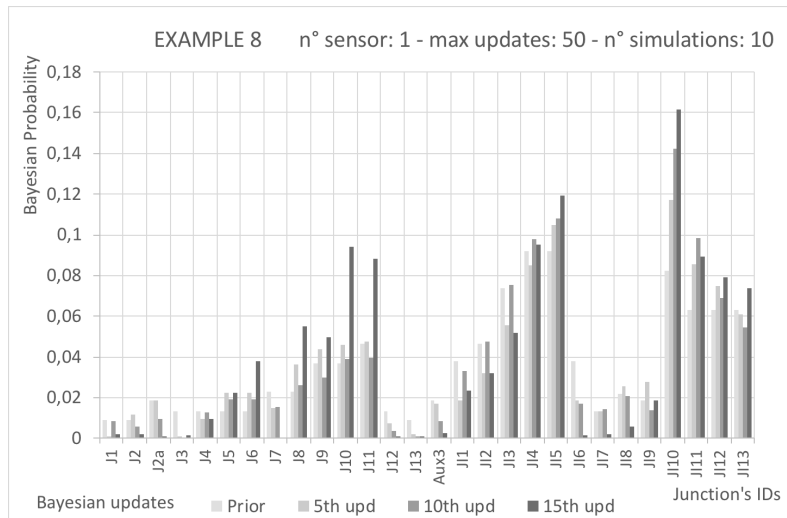


Figure 4: Graphical representation of the Bayesian probability for Example 8 network starting from an informative prior distribution.

Massa Lubrense network				
n° sensors	N° of Bayes updates	Pop. of updates	Highest sensor posterior probability	Prior probability of each sensors combination
6	10	100	0.04	5.26E-04
12	10	500	1.85E-5	1.53E-17
18	10	5000	3.12E-11	2.21E-31
6	10	10	0.003	5.26E-04
12	10	10	2.45E-8	1.53E-17
18	10	10	6.44E-16	2.21E-31
6	50	50	0.056	5.26E-04
12	50	50	4.55E-5	1.53E-17
18	50	50	5.12E-11	2.21E-31
6	100	100	0.126	5.26E-04
12	100	100	6.65E-4	1.53E-17
18	100	100	2.15E-9	2.21E-31

Table 2: Results in terms of Bayesian probability in the various BDN configurations (Massa Lubrense).

The BDN approach was successively applied to the real network of Massa Lubrense (Italy) considering three sensors configurations involving 6, 12 (the real number of implemented monitoring stations) and 18 sensors.

The analysis was carried out both analyzing the impact of BDN approach parameters and the ability of the best possible sensor locations. In order to save computational time, the analysis was only carried out starting from an informative priori distribution based on 100 random contamination simulations.

Table 2 reports the results in terms of Bayesian probability in the various considered BDN configurations. In this case, the best sensor configuration and the efficiency of the sensor network in identifying the polluting source depend on the number of simulations and on the number of procedure updates. The complexity of the analyzed network takes, as a consequence, that the best strategy is to increase the number of Bayesian updates reducing the population of each update: the use of 50 updates with 50 simulations each provides better results compared with the use of 10 updates with 1000 simulations each. This can be explained by the complexity of the system and the number of possible sensors combinations that should be considered.

Positively, when 50 or 100 updates are considered, all the analysis converge on the same set of sensor locations so confirming the robustness of the proposed approach. Analyzing the three configurations (6, 12 and 18 sensors), the probability that the sensors intercept the contamination is equal to 58% with one sensor, 81% with two sensors and 92% inserting three sensors.

4 Conclusion

The contamination of surface waters represents one of the most important aspects of the urban management, both for the environmental and sanitary implications, and for social-economic issues that may rise. The goal of the present work is to develop a decision-support approach to identify the location

of the water quality sensors in order to identify illicit intrusions. This is done, specifically looking at soluble conservative pollutants, such as heavy metals.

The analysis was based on a Bayesian approach in order to introduce data assimilation and identification probability in the procedure. The system was simulated by means of SWMM under different random contamination scenarios. The progressive updates were carried out considering contamination events and evaluating the probability of identifying the contamination source with the sensors located in a specific node. An appropriate pre-conditioning approach is relevant in terms of early identification of the most suitable sensor locations. The proposed approach showed the progressive increase in identification probability obtained through the use of Bayesian update.

Looking broadly to all the performed analyses:

- The value of the Bayesian probability tends to increase as the number of Bayesian updates increases; however, the number of simulations in each update should be based on the complexity of the problem to be solved as complex problems needs a larger population for each update;
- As the number of updates increases, with the same number of sensors, the max value of the Bayesian probability increases however the marginal increments are progressively lower;
- The proposed algorithm is able to locate sensors able to identify a large number of contamination events: up to 84% in Example 8 case with three sensors and up to 92% in Massa Lubrense with 18 sensors.

Even if the number of sensors is high to obtain a technically relevant probability of contamination source identification, the BDN approach allows to act a progressive implementation of the sensor network depending on water manager budget limitations. The use of prior knowledge allows to converge towards an optimal solution in less time and with smaller computational resources. Even if in a small network, this fact may be negligible, the impact on larger networks is relevant enabling to obtain the same optimal configuration halving the needed number of simulations.

Possible future research should focus on non-conservative contaminants (such as organic compounds and some priority pollutants), on the implementation of Lagrangian sensor deployment and on the introduction of distributed information about polluting sources in the definition of prior distributions.

References

Banik B. K., Di Cristo C., Leopardi A. (2015). A pre-screening procedure for pollution source identification in sewer systems. 13th Computer Control for Water Industry Conference, CCWI 2015, Leicester, UK. *Procedia Engineering*, 119(1), pp. 360-369.

Banik B. K., Di Cristo C., Leopardi A., and de Marinis G. (2017a). Illicit intrusion characterization in sewer systems. *Urban Water J.*, 14(4), pp. 416-426.

Banik B. K., Alfonso L., Di Cristo C., Leopardi A., Mynett A. (2017b). Evaluation of different formulations to optimally locate sensors in sewer systems. *Journal of Water Resources Planning and Management*, 143(7).

Boenne W., Desmet N., Van Looy S. and Seuntjens P. (2014). Use of online water quality monitoring for assessing the effects of WWTP overflows in rivers. *Environ. Sci.: Processes Impacts*, 16, 1510.

Freni G. & Sambito M. (2017). Probabilistic approach to the positioning of water quality sensors in urban drainage networks. *International Conference of Urban Drainage*, Prague.

Gironás J., Roesner L.A., Davis J., Rossman L.A, and Supply W. (2009). Storm water management model applications manual. Cincinnati, OH: national Risk Management Research Laboratory, Office of Research and Development, *US Environmental Protection Agency*.

Irvine K., Rossib M. C., Vermettea S., Bakerta J. and Kleinfelder K. (2011, December). Illicit discharge detection and elimination: Low cost options for source identification and trackdown in stormwater systems. *Urban Water Journal*, 8(6), pp. 379–395.

Jiang J., Shi B., Huang A., Wang N., Yuan Y. (2015). Inverse uncertainty characteristics of pollution source identification for river chemical spill incidents by stochastic analysis. *Journal of Hydroinformatics*.

Kim M., Choi C. Y., Gerba C. P. (2013). Development and evaluation of a decision supporting model for identifying the source location of microbial intrusions in real gravity sewer systems. *Water Research*, 47, pp. 4630 – 4638.

Korb K. B. & Nicholson A. E. (2010). Bayesian Artificial Intelligence. Second Edition. *CRC Press*.

Montserrat, L. Bosch, Kiser M.A., Poch M., Corominas L. (2015). Using data from monitoring combined sewer overflows to assess, improve, and maintain combined sewer systems. *Science of the Total Environment*, 505, pp. 1053–1061.

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

Vezzaro L., Christensen M.L., Thirsing C., Grum M., Mikkelsen P.S. (2014). Water quality-based real time control of integrated urban drainage systems: a preliminary study from Copenhagen, Denmark. 12th International Conference on Computing and Control for the Water Industry, CCWI2013. *Procedia Engineering*, 70, pp. 1707 – 1716.