



**HAL**  
open science

## **An Incremental Sensor Placement Optimization in a Large Real-World Water System**

Nicolas Cheifetz, Anne-Claire Sandraz, Cédric Féliers, Denis Gilbert, Olivier Piller, Anne Lang

► **To cite this version:**

Nicolas Cheifetz, Anne-Claire Sandraz, Cédric Féliers, Denis Gilbert, Olivier Piller, et al.. An Incremental Sensor Placement Optimization in a Large Real-World Water System. *Procedia Engineering*, 2015, 119, pp.947-952. 10.1016/j.proeng.2015.08.977 . hal-01199563

**HAL Id: hal-01199563**

**<https://hal.science/hal-01199563>**

Submitted on 28 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



13<sup>th</sup> Computer Control for Water Industry Conference, CCWI 2015  
**An Incremental Sensor Placement Optimization  
in a Large Real-World Water System**

Nicolas Cheifetz<sup>a,\*</sup>, Anne-Claire Sandraz<sup>a</sup>, Cédric Féliers<sup>a</sup>,  
Denis Gilbert<sup>b</sup>, Olivier Piller<sup>b</sup>, Anne Lang<sup>c</sup>

<sup>a</sup>Veolia Eau d'Ile de France, Le Vermont, 28, Boulevard de Pesaro, Nanterre F-92751, France

<sup>b</sup>Irstea, Water Infrastructure Asset Management Team, Bordeaux regional centre, Cestas F-33612, France

<sup>c</sup>Syndicat des Eaux d'Ile de France (SEDIF), 120 Boulevard Saint-Germain, Paris F-75006, France

---

**Abstract**

Supplying modern water systems in smart cities requires the ability to monitor water quality in the production plants and the distribution network. Protection against accidental or intentional events is usually based on an Early Warning Detection System (EWDS) to minimize the impact of any contamination [1]. A major issue is positioning online sensors along the water distribution network to ensure the best protection at minimum cost. Several contributions for tackling such problem have been proposed during the last decade, including a scientific challenge comparing 14 methodologies [2] and two reviews of about 150 articles [3,4]. Currently there is no consensus about algorithms or design objectives to use, and such a problem is NP-hard [5,6]. In this paper, a greedy approach is proposed to near-optimally solve the sensor placement problem dealing with large-scale water systems. This methodology is designed to integrate both the specificity of the studied network (expert/prior knowledge) and flexibility related to the uncertainty of the nature, time, and duration of contamination injections. The method is illustrated to minimize the expected fraction of the exposed population on the largest network in France (about 100,000 nodes, 600,000 connections, above 8,000 km of pipes). Costly approaches in terms of computation time, such as MIP (Mixed-Integer Programming) and exhaustive search cannot scale to large networks. A sensitivity analysis is presented using the greedy approach on a sub-network which leads to choose the number of sufficient contamination simulations and the concentration threshold used to detect contaminations. The extensive experiments allow us to highlight the effectiveness and the rapidity of the proposed approach.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Scientific Committee of CCWI 2015

**Keywords:** Early Warning Detection System (EWDS), Sensor Placement Optimization Problem (SPOP), Greedy algorithm, Water Distribution Network (WDN), Water Quality, Syndicat des Eaux d'Ile de France (SEDIF)

---

**1. Introduction**

The monitoring of Drinking Water Distribution Systems (DWDS) deals with the automatic surveillance of production plants and the distribution network in order to minimize the impact of any accidental or malicious contamination. In fact, these first infrastructures are usually monitored by sensors that were already deployed in pre-existing installa-

---

\* Corresponding author. Tel. : +33 1 55 67 66 82.

E-mail address: [nicolas.cheifetz@veolia.com](mailto:nicolas.cheifetz@veolia.com)

tions of water utilities. On the other hand, the detection of any injection into the network remains a very challenging problem due to the size and complexity of large scale WDN. Indeed, the placement of an “optimal” sensor network providing the best possible sensing quality is a combinatorial optimization problem that has been proven to be NP-hard [5,6]. In other words, it is unlikely that the exact solution will be found efficiently by any current algorithm.

Such sensing problem received significant interest over the last decade including a benchmark challenge (the Battle of the Water Sensor Networks, BWSN) comparing 14 methodologies [2]. Note that a greedy-like algorithm with lazy evaluations obtained the highest score (number of non-dominated solutions) at the BWSN challenge. A large variety of approaches have been proposed to locate sensors including single/multi objective optimizations, exact/suboptimal/expert methods (e.g. resp. mixed-integer programs, heuristics, human judgments) with performance guarantees or not, perfect sensors (accurate and reliable) or not, robust optimization (worst case contaminations) or minimization of the expected impact, etc. The interested reader can refer to reviews comprised of about 150 articles [3,4]. In this article, we are interested in the placement of quality sensors to ensure the best protection at minimum cost. Nevertheless, there is no clear consensus about algorithms or design objectives to use. At least four questions have to be addressed to design a sensor deployment [7] : What objective function(s) should be optimized ? How many sensors are required ? Should the worst or the average impact of (malicious or accidental) contaminations be minimized ? How scalable is the placement optimizer ? In this work, we present an efficient and flexible methodology to place about 100 sensors minimizing the average impact on population.

Within the framework of the project SMaRT-Online<sup>WDN</sup>[8], Veolia Eau d’Ile de France participated to the design and evaluated the performance of a methodology to deploy sensors onto the drinking distribution network of Syndicat des Eaux d’Ile de France (SEDIF). The next section describes the proposed methodology based on a greedy algorithm to perform the placement of quality sensors. The third section presents the results gained on the SEDIF network which is the largest WDN in France and one of the largest in Europe. A sensitivity analysis provides indications to choose the number of sufficient contamination simulations and the concentration threshold used to detect contaminations. Finally, some conclusions are drawn in section 4.

## 2. Placement of quality sensors

A large number of performance objectives have been proposed to quantify the effectiveness of the placement solutions [2,4]. Some criteria aim to facilitate the procedure of decontamination such as the minimization of the contaminated network water volume or the minimization of the contaminated network pipe surface, which are computed before any sensor detection [9]. The minimization of the public health impacts is a global objective usually well accepted by the scientific community [3]. In this work, the minimization of the population exposed to contaminations is illustrated in subsection 3.2. Note that the proposed methodology gives the opportunity to optimize different criteria that can be computed based on the occurrence of contamination incidents on the distribution network, assuming certain structural property in order to preserve near-optimal performance.

### 2.1. Characterization of the contamination impact

The contamination impact is quantified using an extensive number of contaminant injections simulated by tools of hydraulic modeling such as the well-known EPANET or Porteau (available at <http://porteur.irstea.fr/>). For each simulated contamination, a substance is injected from a single node belonging to a hydraulic model and then, the contamination is detected when the substance dose is above a given threshold at a sensor location. We are considering the problem of locating sensors in order to minimize the average impact upon a set of contaminations which lead to the detection of random or accidental contamination events. We are not interested in minimizing the worst impact of possible events because we consider the worst-case scenario subjectives and changing events can imply the variation of the contaminant characteristics or the sensor network setup. Indeed, it can be seen that a sensor placement minimizing the average impact outperforms worst-case designs when contamination conditions change [10].

In this article, the set of simulated contaminations represents the key information in order to optimize a sensor placement upon a distribution network. Contamination scenarios were randomly generated, such as :

- Simulation duration : a few days is enough to avoid the influence of initial conditions on the solution
- Nature of the injected substance : a conservative contaminant is chosen for its stronger impact with respect to the concentration such that the dilution will be induced only by mixing at nodes
- Time and duration of injection : uniformly random between 0 and 24 hours
- Localization of the injection : uniformly random among nodes of the hydraulic model

The estimation of the two following conditions is studied in subsection 3.3 using a sensitive analysis. Nevertheless, we provide some indications to tune these two conditions used in subsection 3.2 :

- Number of simulations : this number is set according to the convergence of the objective function(s) ; as an upper bound, we chose empirically a ratio of about three model nodes for each contamination
- Contamination detection : an event is detected when the concentration of the injected substance is above a given threshold ; we chose empirically a threshold of  $10^{-2}$  with an injection of 100

In this work, contamination injections were simulated using the modeling tool called Synergi Water<sup>®</sup> (DNV GL).

After an extensive number of simulated injections following the previous indications, all the relevant information regarding the contaminations is aggregated into a single *matrix of contaminations* per hydraulic model. It can be seen as a tridimensional matrix  $M_{i,j,k}$  used to compute the objective function(s), where  $i$  denotes a node of the hydraulic model,  $j$  refers to a simulated contamination and  $k$  indicates a certain feature (e.g. duration since the injection time, cumulative sum of affected population at each exposed node over time, etc.) for the computation of certain criterion. Note that undetected contaminations are taken into account : every criterion is still computed as the values by the end of the simulation in order to penalize these incidents that are not detected. Furthermore, this matrix of contaminations is sparse due to the limited number of exposed nodes per simulated contamination. Then, this matrix can be efficiently parsed to perform relevant evaluations of exposed nodes and exploit some structural property of the objective function(s) that is briefly presented in the next subsection.

## 2.2. Incremental methodology for sensor positioning

In addition to the impact estimation, this flexible method can handle certain expert knowledge to promote or oppose some regions of the water distribution network. The first expert priors deal with the specificities of our deployed quality sensors<sup>1</sup> : the sensors are exclusively placed onto nodes considering that its adjacent pipes have a nominal diameter neither too small nor too large, the water velocity cannot be too low (no underestimation of the chlorine concentration) and there is no reversal of the flow direction. Moreover, the goal of the following constraints is to speed up the optimization process : sensors can only be placed onto “super nodes” using some graph decomposition techniques (nodes in the core subgraph with degrees of at least two) [11], and the pre-equipped nodes are not evaluated during the optimization. In addition to these inclusive/exclusive constraints, our methodology can easily penalize (without excluding) some potential locations of nodes by adding a regularization term to the objective function(s).

The sensing optimization problem addressed in this article aims to find a set of sensors that minimizes a single cost function. The formulation of such problem is NP-hard [5,6] and we propose an intuitive greedy algorithm to solve it. This procedure is simply described by selecting iteratively a single location based on the set of sensors previously selected. Such iterative heuristic is fast in practice and finds near-optimal solutions in theory [5] with the optimization of certain submodular cost function. Indeed, the next section of this article presents an optimization study minimizing the expected population exposed to contaminations. Such submodular cost function has a diminishing returns effect [5] which means that adding a sensor to a large set of sensors provides less information than adding it to a smaller set. Without loss of generality, the greedy algorithm would perform well with other submodular functions such as time to detection, etc.

The next section presents the obtained results for the placement of quality sensors on a large scale water distribution network in France.

<sup>1</sup> Sensors KAPTA<sup>™</sup> 3000 measuring active chlorine, temperature, conductivity and pressure ; these probes were developed within the framework of the European project called SecurEau [9].

### 3. Experiments on a large scale water distribution network

#### 3.1. Experimental framework

The SEDIF gather 149 municipalities around Paris (France) and distributes every day about 750,000 m<sup>3</sup> to more than 4 millions of inhabitants based on 3 production plants. The distribution network is more than 8,000 km of pipes and can be represented by 11 hydraulic models with various sizes. The ultimate goal is to deploy about 200 quality sensors (cf. footnote<sup>1</sup>) onto this network and we propose to solve a sensor placement problem in each hydraulic model. A first deployment of about 100 sensors was performed by hydraulic expert knowledge and one hydraulic network was fully equipped as a test-bench within the SMaRT-Online<sup>W<sup>DN</sup></sup> project [8]. Then, the goal of this study is to optimize the placement of the last 100 remaining sensors among more than 100,000 nodes spread over 10 hydraulic models. Table 1 summarizes the deployed sensor network among each hydraulic model independently. It is worth noting that the number of sensors has been set regarding to the number of nodes composing each hydraulic model, and that more sensors are required for models 1, 5, 8 and 9 due to their sizes.

TABLE 1. Description of the 10 hydraulic models and the deployment setup.

Hydraulic model	1	2	3	4	5	6	7	8	9	10	$\Sigma$
Nodes	18,942	4,066	2,843	2,371	16,787	5,514	1,819	33,278	14,250	3,344	103,214
Edges	23,655	4,791	3,195	2,705	19,493	6,137	2,171	39,646	16,395	4,124	122,312
Tanks	8	1	2	2	4	4	1	7	10	3	42
Pumps	1	0	0	0	0	0	0	0	3	0	4
Existing sensors	18	0	3	2	15	9	2	25	20	4	98
Deployed sensors	34	6	8	6	27	12	7	44	45	9	198
Simulated contaminations	5,500	1,500	1,000	1,000	4,500	1,500	500	9,500	4,000	1,000	30,000

In this work, the public health impact of contaminations is evaluated using the two performance criteria that are defined below, given  $\delta$  a set of sensors and the index  $j$  of a contamination simulation :

- Fraction of exposed population (EP)

$$EP(\delta) = \frac{\# \text{ connections at exposed nodes in simulation } j}{\# \text{ connections into the current hydraulic model}} \quad (1)$$

where an exposed node is some node with a significant contaminant concentration occurring at a time before any detection of the current contamination

- Detection delay (DD)

$$DD(\delta) = \min_{d \in \delta} t_j(d) \quad (2)$$

where  $t_j(d)$  is the duration between the injection time and the time of detection by sensor  $d$

Let us recall that we are interested by the averaged impact of contaminations which means that the previous criteria will be evaluated as averages over all the simulated contaminations. And in order to compare the performance improvements for each criteria, these are normalized by the situation with no sensors deployed into the network.

The following subsections present the optimization results by minimizing the average fraction of exposed population. The criteria are normalized such as more the results are good, the more criteria values are high. The next subsections illustrate the deployed sensor network onto the SEDIF network and the performance of the proposed method for incremental sensor placement.

### 3.2. Results on the SEDIF network

We use the greedy algorithm for positioning 100 sensors spread over 10 hydraulic models by minimizing the average fraction of exposed population. Table 2 presents the resulting performance with two normalized criteria (defined in the last subsection) to maximize. The total optimization time was less than half an hour using a standard computer running Win7 with a processor i5-3230M (dual cores, 2.60GHz) and 8 Go RAM. The optimized criterion EP is improved up to 90% with the situation of no deployed sensors for each hydraulic model. The performance of the pre-existing sensor network was between 0% and 70%. Note that the placement of a large number of sensors does not necessarily imply a better performance as illustrated by the model 4. This can be explained by a long residence time inside the studied network.

TABLE 2. Evaluation of the deployed sensors using performance criteria. The meaning of the normalized criteria is as follows : EP=average fraction of exposed population, and DD=average detection delay.

Normalized criteria	Sensor network	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
EP (%)	Existing	70.25	0.00	42.87	63.87	72.54	60.78	20.91	70.19	49.29	31.93
	Deployed	89.29	84.02	81.33	88.59	90.98	81.36	73.80	88.46	83.44	72.96
DD (%)	Existing	42.40	0.00	20.33	44.88	32.53	38.51	16.81	39.66	32.32	29.24
	Deployed	57.30	58.79	61.97	61.68	59.08	58.93	49.68	53.88	66.18	63.20

### 3.3. Sensitivity analysis

This subsection summarizes a sensitivity analysis for sensor placement on the hydraulic model n°5 (with 15 preexisting sensors) by minimizing the average fraction of exposed population. Figures 1 and 2 represent mean curves of the normalized criteria EP (a) and DD (b) according to various settings : based on 10 matrices of 4,500 random contaminations, the concentration threshold vary from  $10^{-4}$  to 10 (Figure 1), and with a concentration threshold of  $10^{-2}$ , the number of simulations per contamination matrix vary from 10 to 45,000 (Figure 2). We are interested in the shape of the mean trajectories computed for each criteria. A poor precision with high concentration threshold leads to underestimate the criteria, especially with criterion DD and a concentration of  $10^{-2}$  is seen as sufficient (Figure 1). On the other hand, the criteria are overestimated with few simulations and about 1,000 simulations is seen as sufficient (Figure 2). Note that the standard deviation is usually higher with few placed sensors, a high concentration and a small number of simulations. We can observe a certain stability of the optimized criterion and the greedy behavior especially from both Figures with sufficient simulations and a proper detection threshold.

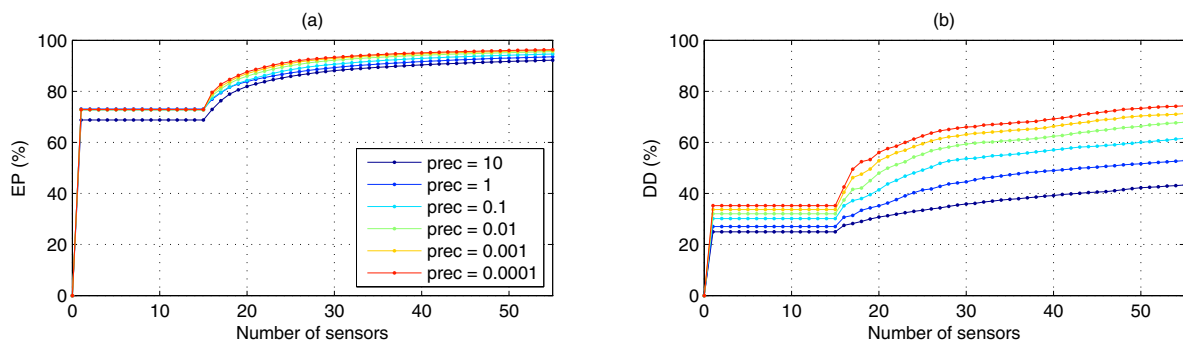


FIGURE 1. Performance criteria comparison of different detection thresholds, so-called *prec*, on hydraulic model n°5. For each sensor placement, the concentration threshold vary from  $10^{-4}$  to 10 with concentration injection of 100.

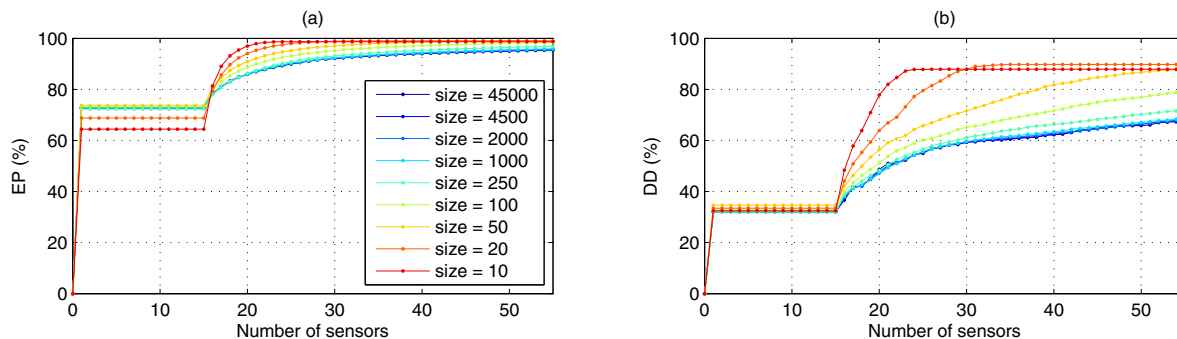


FIGURE 2. Performance criteria comparison of different numbers of contamination simulation, so-called *size*, on hydraulic model n°5. For each sensor placement, the number of simulation vary from 10 to 45,000.

#### 4. Conclusions and prospects

This article presents a global methodology for sensor placement using a greedy algorithm. This approach is flexible integrating various expert knowledge and the optimization is near-optimal with certain objective-function. This efficient method allowed us to perform sensor placements in reasonable time on a large scale water distribution network with about 100,000 nodes. A sensitive analysis is described to choose the concentration threshold for detection and the amount of contaminations to simulate. Further research will be addressed through the upcoming ResiWater project.

#### Acknowledgements

The work presented in this paper is part of the French-German collaborative research project SMaRT-Online<sup>W<sub>DN</sub></sup> [8] that is funded by the French National Research Agency (ANR ; project : ANR-11-SECU-006) and the German Federal Ministry of Education and Research (BMBF ; project : 13N12180).

#### References

- [1] American Society of Civil Engineers (ASCE), Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System, Technical Report, 2004.
- [2] A. Ostfeld, J. Uber, E. Salomons, J. Berry, W. Hart, C. Phillips, J. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, F. di Piero, S. Khu, D. Savic, D. Eliades, M. Polycarpou, S. Ghimire, B. Barkdoll, R. Gueli, J. Huang, E. McBean, W. James, A. Krause, J. Leskovec, S. Isovitich, J. Xu, C. Guestrin, J. VanBriesen, M. Small, P. Fischbeck, A. Preis, M. Propato, O. Piller, G. Trachtman, Z. Wu, T. Walski, The Battle of the Water Sensor Networks (BWSN) : A Design Challenge for Engineers and Algorithms, *Journal of Water Resources Planning and Management* 134 (2008) 556–568.
- [3] W. E. Hart, R. Murray, Review of sensor placement strategies for contamination warning systems in drinking water distribution systems, *Journal of Water Resources Planning and Management* 136 (2010) 611–619.
- [4] S. Rathia, R. Gupta, Sensor placement methods for contamination detection in water distribution networks : a review, in : 16th Conference on Water Distribution System Analysis, WDSA 2014, 2014.
- [5] A. Krause, Optimizing sensing : theory and applications, Ph.D. thesis, Carnegie Mellon University, 2008.
- [6] X. Xu, Y. Lu, S. Huang, Y. Xiao, W. Wang, Incremental sensor placement optimization on water network, in : ECML PKDD 2013, volume 8190 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2013, pp. 467–482.
- [7] R. Clark, S. Hakim, A. Ostfeld, Handbook of Water and Wastewater Systems Protection, Protecting Critical Infrastructure, Springer, 2011.
- [8] SMaRT-Online<sup>W<sub>DN</sub></sup> project, Online Security Management and Reliability Toolkit for Water Distribution Networks, 2012–2015. URL : <http://www.smart-onlinewdn.eu/>.
- [9] SecurEau project, Security and decontamination of drinking water distribution systems following a deliberate contamination, 2010–2013. URL : <http://www.secureau.eu/>.
- [10] M. Davis, R. Janke, C. Phillips, Robustness of designs for drinking water contamination warning systems under uncertain conditions, *Journal of Water Resources Planning and Management* (2013).
- [11] J. Deuerlein, O. Piller, I. Montalvo, Improved real-time monitoring and control of water supply networks by use of graph decomposition, *Procedia Engineering* 89 (2014) 1276–1281.