Sampling Significant Contamination Events for Optimal Sensor Placement in Water Distribution Systems

Article *in* Journal of Water Resources Planning and Management · June 2017

DOI: 10.1061/(ASCE)WR.1943-5452.0000814.

3 authors:

Silvia Tinelli
University of Pavia

Enrico Creaco
University of Pavia

Carlo Ciaponi
University of Pavia

WR2845

Sampling Significant Contamination Events for Optimal Sensor Placement in Water Distribution Systems

Silvia Tinelli[1], Enrico Creaco[2] and Carlo Ciaponi[3]

Abstract

This work presents a procedure for sampling the most representative contamination events in the framework of the optimal sensor placement with two objective functions to be minimized, that is, sensor redundancy and contaminated population. Compared to other sampling methods present in the scientific literature, it is based on practical considerations on network topology and operation. This aspect confers on the procedure lightness from the computational viewpoint. Sampling was carried out on 4 variables, that is injection location, starting time, mass rate, and duration. The injection location was sampled as a function of the distance from the source based on network connectivity. A single starting time was selected inside each network operating phase, during which pipe water discharges are quite constant. One single mass rate was selected as significant, considering the linearity of the contaminant advection-reaction equation under specific conditions. In fact, thanks to this linearity, the results of quality simulations associated with a generic mass rate can be easily derived from those associated with the selected mass rate. Finally, a single (small) duration was sampled. In fact, a long duration event can be simply regarded as the sum of various small duration events.

The procedure was tested in two case studies of different complexity. As evidence of the sampling effectiveness, the results of the optimal sensor placement did not vary significantly

[1] PhD student, Dipartimento di Ingegneria Civile e Architettura, University of Pavia, Via Ferrata 3, 27100 Pavia (Italy). Email: silvia.tinelli01@universitadipavia.it

[2] Assistant Professor, Dipartimento di Ingegneria Civile e Architettura, University of Pavia, Via Ferrata 3, 27100 Pavia (Italy). Honorary Senior Research Fellow, CEMPS, University of Exeter. Adjunct Senior Lecturer, Faculty of Engineering, Computer and Mathematical Sciences, The University of Adelaide, SA 5005, AUSTRALIA. Email: creaco@unipv.it

[3] Professor, Dipartimento di Ingegneria Civile e Architettura, University of Pavia, Via Ferrata 3, 27100 Pavia (Italy). Email: ciaponi@unipv.it

when the sampled contamination events were used inside the optimization, instead of the totality of possible contamination events.

Keywords: Sensors; Water distribution systems; water pollution; sampling; network design

Introduction

The easy accessibility for intruders and the solubility of many chemical, biological, nuclear agents or toxic substances in water make water distribution systems (WDSs) a vulnerable target of contamination attacks. Furthermore, WDSs are exposed to the accidental ingress of materials and pollutants during maintenance works, due to the lack of hydraulic seal.

Since contamination events can occur at any node and at any time, they are hardly predictable and detectable. Due to these reasons, some countries, especially the most sensitive to the problem of terrorism, such as the U.S. and Israel, have long been involved in research programs to improve the security of WDSs with an appropriate contamination detection system. Spurred by the importance of the problem, many countries of the European Community have recently joined these research programs. In this framework, one of the main research issues lies in determining the optimal location of sensors, able to detect the most common water parameters and thus, to monitor the networks.

The issue of the optimal location of sensors has been faced through both single-objective (e.g., Lee and Deininger, 1992; Kumar et al., 1997; Kessler et al., 1998; Woo et al., 2001; Al-Zahrani and Moied, 2001; Ostfeld and Salomons, 2004; 2005; Berry et al., 2006, 2009; Propato (2006); Propato and Piller, 2006; Shastri and Diwekar, 2006; Cheifetz et al., 2015) and multi-objective (e.g., McKenna et al., 2006; Ostfeld and Salomons, 2006; Dorini et al., 2006; Eliades and Polycarpou, 2006; Gueli, 2006; Huang et al., 2006; Wu and Walski, 2006; Ostfeld et al., 2008; Preis and Ostfeld, 2008) methodologies. Among the single-objective methodologies, Kessler et al. (1998) introduced an algorithm aimed at finding the best combination of sensors capable of providing a given level of service through a set covering algorithm. In the Authors' approach, the term "level of service" indicated the maximum volume of polluted water exposed at a concentration higher than a minimum hazard level and consumed before detecting the contamination. Ostfeld and Salomons (2004) used a similar approach, solving the optimization problem through a Genetic Algorithm while Ostfeld and Salomon (2005) extended their previous work by introducing uncertainties to the demands and to the injected contamination events. Berry et al. (2006) introduced a mixed-integer programming (MIP) for sensor placement. Propato (2006) and Propato and Piller (2006) formulated a linear mixed-integer programming model to identify optimal sensor locations for early warning against accidental and intentional contaminations. Shastri and Diwekar (2006) introduced the study of uncertainties related to contamination location and demand at the time of the intrusion. In fact, since changing water demand can cause changes in flow directions, contaminated nodes may also change; therefore, a change in demand by 25% was introduced by Shastri and Diwekar (2006). Cheifetz et al. (2015) proposed a greedy incremental sensor placement approach, to be used for sensor optimization in large real-world water system.

At the same time, Ostfeld et al. (2008) pointed out the importance of creating multi-objective algorithms to refine the problem of optimal sensor location. In the framework of multi-objective optimization, Ostfeld and Salomons (2006) and Preis and Ostfeld (2008) used the multi-objective genetic algorithm NSGA-II (Deb et al., 2002). As already done by Shastri and Diwekar (2006), McKenna et al. (2006) investigated the problem of the uncertainties. In detail, they assessed the perfect sensor assumption, that is the ability to indicate a positive contamination event as soon as any amount of contamination reaches the sensor. In this context, they evaluated the impact of sensor detection limit and proved that the detection of events is dependent on the detection limit. However, the results of their work showed that a sensor detection limit of 0.01 times the average source concentration is adequate for maximum protection.

During the Battle of the Water Sensor Networks (BWSN) held in the United States many other approaches were evaluated to face the problem of the optimal sensor location: a summary of the results can be found in Ostfeld et al. (2008).

Later, Berry et al. (2009) addressed the problem of the imperfect sensors, accounting for the effect of false sensor readings: the inclusion of false negatives/positives led to a non-linear formulation of the optimization problem and the results showed that the false sensor readings could have a significant impact on network safety. Nowadays, a wide variety of sensors are commercially available. Since the technology is advancing, the sensors can measure simultaneously an increasing amount of physical-chemical water parameters, considered to be crucial for detection (e.g., EPA, 2012 through CANARY; Perelman et al., 2012; Arad et al., 2013; Oliker and Ostfeld, 2014a, 2014b). In the context of event detection, Perelman et al. (2012) utilized the artificial neural networks for studying the trade-off between multivariate water quality parameters and detection of possible outliers: the results consist of alarms indicating a possible contamination event based on single and multiple water quality parameters. Arald et al. (2013) aimed to detect events by exploring the time series-behavior of routine hydraulic and water quality measurements, developing a dynamic threshold scheme. Oliker and Ostfeld (2014a; b) improved the event detection ability including the support vector machine for the detection of outliers and a multivariate analysis for the examination of the relationships between water quality parameters and their mutual patterns.

However, despite the large research carried out in the field, a challenge is still unsolved: the potential contamination events in a real WDS with complex network topology are countless, since each of them is characterized by a different injection location, duration, mass rate and starting time. The large number of contamination events to consider makes the problem of the optimal location of sensors intractable in practice. Hence arises the necessity to set up a sampling method able to select the most representative events, which can be considered to make the problem less burdensome to solve. In this context Preis and Ostfeld (2008) developed a heuristic procedure for sampling a set of contamination events. They reduced the initial contamination matrix size by using a statistical approach which selects representative events considering their geographical x/y coordinates, few specific injection mass rates, injection starting times and injection durations. Weickgenannt et al. (2010) introduced an importance-based sampling method to classify effectively the contamination events based on their importance in terms of the total volume of water that is polluted in a certain interval of time. Due to the complexity of WDSs, the prediction of the performance under various conditions such as failure scenarios, detection of contamination intrusion sources and of sensor placement locations is difficult. Thus, Perelman and Ostfeld (2011) developed a graph theory connectivity based algorithm to simplify the system behavior. They introduced the notion of clustering in the context of topological/connectivity analysis and suggested connectivity analysis for topological clustering of nodes, facilitating the sampling of nodes for sensor locations. Chang et al. (2012) established a rule-based expert system where the two rules, accessibility and complexity, converge to a set of nodes for the final sensor locations based on four design objectives, including the expected time of detection, the expected population affected prior to detection, the expected consumption of contaminant water prior to detection, and the detection likelihood. Diao and Rauch (2013) presented a controllability analysis of the network as preprocessing method for sensor placement: it determines the nodes that have an outcome indication over a maximum number of downstream nodes. Rathi and Gupta (2016) formulated a simplified method that simultaneously maximizes two performance objectives, the demand coverage and the time-constrained detection likelihood, which were combined into a single objective by using weights; they also used Genetic

Algorithm to obtain the final optimum sensor locations. Zhao et al. (2016) proposed a sensor placement algorithm based on greedy heuristics and convex relaxation, and demonstrated significant performance by applying it to repeated sampling of random subsets of events. The current paper aims to face the optimal sensor location problem as a bi-objective optimization problem, where the sensor redundancy and the contaminated population are both minimized. Since the solution of the optimization problem requires definition of a set of possible contamination events, a sampling method is developed in order to select a reduced but representative set of events, making the problem solution computationally feasible. Unlike the approaches described above, the proposed methodology considers practical information on network topology and operation for the sampling. In the following sections, the sampling method is first described, followed by the bi-objective optimization of sensor location. The applications to two case studies follow.

Methodology

Sampling

A method for defining a set of contamination events, which is representative for the totality of the events, was set up. Each possible contamination event is characterized by certain values of injection location, starting time, mass rate, and duration. Therefore, the sampling was done for each of the event characteristics, as is explained in the following sub-sections.

Injection location

Though all the network nodes could theoretically be injection locations, an algorithm was developed using the graph theory to select the representative nodes that should be considered for the injection of contaminants. A preliminary step of the algorithm is the opening of network loops and source interconnection paths, which can be carried out through the minimum spanning tree algorithm (Kruskal, 1956) and/or other procedures accounting for pipe diameters and water discharges. Subsequently, the representative nodes of the system are selected as a function of their gradually increasing distance from the source nodes. In fact, the distance from the source node is a variable with more hydraulic meaningfulness than the nodal geographical x/y coordinates of Preis and Ostfeld (2008). Specifically, sampling is done with a prefixed frequency, i.e. one out of two, three, four (and so forth) nodes, along the path outgoing from the network source(s). In each path, the closest nodes to the sources were always accounted for. Subsequently, the sampling is modified through the two following steps, which force selection of the dead-ends at the expense of the close nodes:

1 – the dead-ends are included in the list of selected nodes;

2 – the nodes adjacent to dead-ends, whether previously sampled, are excluded from the list if they are serial nodes.

Inclusion of dead-ends is important because they are the final nodes of the network where the generic contamination events can be detected. In fact, let the generic water path in the network be considered. The dead end is the only node able to detect all the contamination events that have injections along this path.

The network in Figure 1 is provided as an example for the application of the selection with one out of two sampling frequency. The network has 13 nodes with one source node placed at node 1. The application of the minimum spanning tree algorithm leads to the removal of pipe 13 for loop opening. Subsequently, nodes are selected because of their gradually increasing distance from the source node. As shown in Figure 1, the closest nodes to the tank (node 2 and 3) are selected. Therefore, considering a frequency of one out of two nodes, the selected nodes are 2, 3, 5, 6, 7, 12 and 13. In particular, being a dead-end, node 12, which should not have been included as a result of the frequency sampling, is finally included in the list at the expense of node 11, which is a serial node adjacent to node 12.

A remark must be made about the injection location sampling, which may fail to select important crossroad nodes or supernodes, as defined by Deuerlein et al. (2014), which belong to several paths. In this context, it has to be remarked that, even if an important node is missed in the sampling, the information about this node is not lost: in fact, contamination events will always reach it through paths including other nodes, which were sampled by the algorithm. Furthermore, the exclusion from the contamination location sampling does not prevent the generic node from being a good sensor location.

Starting time

Taking as benchmark the typical day of WDS operation, every instant could theoretically be the starting time of contamination. This means that considering the whole day sampled with a 0.5 hr step, there could be 48 potential starting times. The sampling of the starting times is carried out based on the WDS operation phases, detected as a function of pipe water discharges, which can vary based on nodal demand and source head patterns and on the switching on/off of pumps.

In detail, these phases can be identified by detecting the times when the water discharges in network pipes vary significantly. Assuming that the water discharge in a pipe follows the daily trend shown in Figure 2, the instants associated with significant changes in the flow are 0 h, 7 h and 15 h. Three phases are then detected for the pipe, that is, phase 1 from 0 h to 7 h, phase 2 from 7 h to 15 h and phase 3 from 15 h to 24 h. The instants of significant flow variation for the whole network are obtained by putting in a single timeline the instants of significant flow variation in all the network pipes. The generic phase is then detected as the time slot between two successive instants in the timeline. Then, a representative instant can be selected for each phase, that is, either the initial instant or an inner instant in which the pipe water discharge are closest to the average values in the phase. These representative instants are selected as significant starting times for the sampled contamination events. In the assessment of any objective functions, the operating phase durations can be used as weights to be associated with contamination events.

Mass rate

The contaminant advection-reaction equations are linear if the contaminant is conservative or first order reactions are used (as is often the case with the optimal sensor location). The consequences of this aspect are easily shown through the explicative example of the network in Figure 3, with one source node (node 1) and 6 demanding nodes (nodes 2-7). In this network, which features a constant demand of 26 L/s, two contamination events are considered with the same injection location (node 2) and duration (4 h) and differing in the mass rate - only for explicative purposes, equal to 50 gr/min (32.05 mg/L) and 200 gr/min (128.21 mg/L) in the two events, respectively. The separate effects of the two events are shown in the graph in Figure 4. In detail, this graph reports, for node 7, the trends of the contaminant concentration in response to the two events. The results clearly show that the two trends are proportional and one can be obtained from the other through multiplication times a factor equal to the mass rate ratio, that is 4. Considering the two trends in Figure 4, where the contaminant concentration rises almost instantaneously to the highest value (very high variation speed), a sensor placed in node 7 would be able to detect the event in both cases, as long as its sensitivity is small enough (usual assumption in the framework of optimal sensor placement) to detect the contamination. Nevertheless, it must be remarked that, in real networks, the cases in which the contaminant concentration is so low to be under the sensor sensitivity are less dangerous in terms of network safety.

In light of the linearity of the contaminant advection-reaction equations, the mass of injected contaminants does not influence certain variables, such as the number of contaminated nodes or the contaminated population, if the pollutant concentrations are high enough to be detected by the sensors. Therefore, if the focus is on the number of contaminated nodes and/or on the contaminated population, rather than on the contamination concentration, the average of the possible masses can be sampled as representative value.

Furthermore, when the contamination concentration is also relevant, the WDS quality simulation can always be carried out only for one contamination mass rate. The results associated with other values can then be derived by taking advantage of the linearity of the equations, as was highlighted above.

Duration

Under conditions of constant (or slightly variable) pipe water discharges (as it occurs in every WDS operation phase), the nodes reached by the generic contamination do not change as a function of event duration. Furthermore, the long duration events can be regarded as a succession of short duration events. Therefore, a single short duration, shorter than the network operating phase durations, can be sampled. It must be underlined that an event lying astride two consecutive operation phases can always be decomposed into the combination of two events, each of which is fully lying inside a single operation phase. Either composing element is banally decomposable into a series of events equal to the short duration event used for the sampling.

Optimal sensor location

After selecting a set of contamination events, each of which characterized by a certain injection location, starting time, mass rate and duration, the knowledge of the system hydraulics and the solution of the advection-reaction equations enables determination of the fate of the contaminants injected into the network. This can be carried out through such software as EPANET (Rossman, 2000). In EPANET, water quality is solved through a PDE system of 1D advection-reaction pipe equations and perfect mixing is considered at junction nodes. This approach is considered acceptable in the search for optimal sensor locations, in which the contaminant is often considered conservative. Nevertheless, it should be underlined that, under specific hydraulic conditions, a better representation of the water quality processes, above all with regards to the mixing at the junctions, could be obtained by taking account of more accurate modelling, such as the computational fluid dynamics (CFD, see Braun et al., 2015 as an example). CFD and other accurate modelling approaches, which are too burdensome to be considered in the sensor design phase, remain valid and useful tools to be adopted for an *a posteriori* analysis of the results.

Once the fate of the contaminants injected into the network has been determined following hydraulic and water quality simulations, the contamination matrix and the time matrix can be constructed, each of which has as many rows and columns as the number of nodes and contamination events, respectively. The generic element of the contamination matrix is set to 1 or 0 whether the generic node of the network is or is not contaminated during the generic contamination event. The time matrix reports, for the generic contamination event, the time interval after the contamination start when the contaminant reaches the generic node. For nodes which remain uncontaminated, the time interval is set to $+\infty$.

It is evident that the application of the sampling method described above enables significant reduction in the number of contamination events to be considered and, therefore, in the size of the contamination and time matrices.

In network simulations aimed at constructing the contamination and time matrices, the daily hydraulic operation of the network is kept unchanged. In fact, the simulations differ in the settings (injection location, starting time, mass rate and duration) used for the contaminant in the water quality part.

In this work, the optimal sensor location problem is formulated as a bi-objective optimization problem, in which the sensor redundancy, f_1, and the contaminated population, f_2, are the objective functions.

The objective function f_1 is related to the number n_r of sensors that can detect the generic contamination event within a time interval Δt_{red} (to be specified) following the first event detection. For a generic contamination event r, n_r is equal to 0 if no sensor can detect the event; it is equal to 1 if only one sensor can detect the event; it is equal to 2 if two sensors can detect the contamination event in close times, that is, the first detection sensor and an extra sensor that detects the event within a time interval Δt_{red} following the first event detection; generalizing the concept, n_r is equal to x when, besides the first detection sensor, there are other x-1 sensors able to detect it within Δt_{red} following the first detection. After assessing n_r for each contamination event, f_1 is calculated as the weighted average value of n_r, that is:

$$f_1 = \frac{\sum_{r=1}^{S} w_r n_r}{\sum_{r=1}^{S} w_r}, \tag{1}$$

where S is the total number of contamination events and w_r is a weight coefficient associated with the generic contamination event. This coefficient is set to 1 if no event sampling has been carried out. Otherwise, it is set equal to the operating phase duration. A large value of f_1 is associated with a large redundancy in the system. This means that, on average, there are numerous sensors able to detect the generic contamination event in the system in a short time interval between one another. Therefore, should a sensor fail, another sensor would be able to give the warning in its place.

The objective function f_2 is related to the contaminated population pop before the first detection of the generic contamination event. In the generic contamination event r, the nodes contaminated before the first event detection can be evaluated and pop_r can be assessed by summing up the inhabitants served by the contaminated nodes. It is assumed that a warning is given to interrupt network service in a reaction time interval Δt_{react} after the event detection.

Δt_{react} is set to 0 hereinafter for simplifying purposes but can be set to other values without loss of validity of the whole methodology. After assessing *pop* for each contamination event, f_2 is calculated as the average value of pop_r, that is:

$$f_2 = \frac{\sum_{r=1}^{S} w_r pop_r}{\sum_{r=1}^{S} w_r}, \tag{2}$$

For the generic location of sensors in the network, the objective functions can be assessed through simple manipulations on the contamination and time matrices. Functions f_1 and f_2 are minimized simultaneously as mutually contrasting objectives in the bi-objective optimization process. In detail, the minimization of the former yields benefits of system cost while the minimization of the latter impacts positively on the system security. Therefore, the optimization results consist of a Pareto front of compromised solutions.

As for the optimal sensor placement, efficient algorithms can be used to find a global optimum when specific objective functions are used. For example, Kessler et al. (1998) and Ostfeld and Salomons (2004) solved a set-covering problem whereas Propato and Piller (2006) solved a MILP problem. Additionally, it is possible to solve with a greedy algorithm in very efficient manner, even for large networks (see, Cheifetz et al., 2015), with an additional optimal sensor added at each iteration. Nevertheless, though being able to guarantee only the near-optimality of the solutions, genetic algorithms have the advantage of being easily implementable, with whatever objective functions, even in the multi-objective framework. Therefore, for the bi-objective optimization of this work, the NSGAII (Deb et al., 2002) was chosen. In the NSGAII population individuals, the number of genes is equal to the number of network nodes where sensors can be installed. Each gene can take on the two possible values 0 and 1, which stand for absence and presence of the sensor in the node associated with the gene, respectively. At each NSGAII generation starting from the initial population, the parent population is selected based on its fitness. The algorithm then generates the offspring population through crossover and mutation from the parent population. After being obtained as combination of the parent and offspring populations, the new population is sorted according to fitness criteria and the best individuals are chosen in order to keep the total number of population individuals constant during the generations. The process is repeated until the maximum number of generations. To ensure robustness of the final solutions found, which are expected to be close to the global optima, a certain number (n_{par}) of NSGAII runs can be carried out in parallel. The final solutions are then put together and some solutions are sampled on the basis of their fitness. The sampled solutions can be used inside the population of new parallel NSGAII runs. This process can be repeated for a certain number of times (n_{iter}).

Applications

Case studies

The presented method was developed and applied to two WDSs of increasing topological complexity. The first case study is a district of the pipe network model used as benchmark in the Battles of Water Networks of the last WDSA conferences (Marchi et al., 2014). The pipe and node characteristics for this district were reported by Creaco and Pezzinga (2015). The number of inhabitants connected to each network node are reported in Table 1. The second case study is a city in Northern Italy (Guidorzi et al 2009; Creaco and Franchini, 2012). In both case studies, the choice of the NSGAII settings was made based on the results of preliminary simulations unreported here. In particular, the NSGAII run was carried out considering: i) a population of 50 individuals and a maximum number of 50 generations in the first case study; ii) a population of 500 individuals and a maximum number of 500 generations in the second case study. Furthermore, n_{par} and n_{iter} were set to 5 in both cases. In both case studies, Δt_{red}, useful for the evaluation of f_1 (eq. 1), was set to 0.5 hr.

First case study

The network of the first case study has 45 demanding nodes, 52 pipes and 1 tank (Figure 5a). In the lowest node in the layout, the water input from a pumping station is considered as a negative demand, as previously done by Creaco and Pezzinga (2015).

In this case study, the following assumptions were made to define the whole sets of contamination events:

1 - all the nodes except for the tank and the input node, that is 44 nodes, were considered possible injection locations;

2 - possible injections were assumed to occur every 30 minutes, leading to 48 possible values of the starting time in the day;

3 - the mass injection rate comprised 4 possible values, that is 50, 200, 350, 500 gr/min;

4 - the injection duration comprised 5 possible values, that is 60, 220, 380, 500, 600 min.

In detail, assumptions 2, 3 and 4 were taken from the work of Preis and Ostfeld (2008).

Therefore, the S total number of contamination events was 44*48*4*5= 42,240. Once S was set, the contamination and time matrices could be evaluated, as explained in the methodology section.

In network modelling for the construction of the total contamination and time matrices, the multiplying coefficients used for nodal demands were expressed through 1 day long patterns with 24 hourly steps. As injections were assumed to take place during the first day of network operation, the simulations had to be conducted for 3 days, since the highest residence time in the network is of about 24 h. This was done to make sure that even contaminants injected close to the sources at the last instant of the first day had enough time to leave the network. Furthermore, the fact that the simulation duration is superior to the residence time inside the network (Piller et al., 2015) is sufficient to avoid the influence of initial conditions on the numerical concentration solution.

In this case study, sampling for the selection of the most representative contamination events was carried out on all the variables, i.e. location, starting time, mass rate and duration. The optimizations were carried out to search for solutions up to a number of sensors equal to the number of nodes with positive demand, that is 44.

Second case study

The network of the second case study is made up of 536 nodes, 825 pipes and 2 reservoirs (see Figure 5b). The contamination matrix of the whole events would be exceedingly large if the same assumptions as the other case study were made concerning the mass injection rate and the injection duration. Therefore, single values of the mass rate and of the injection duration, equal to 200 gr/min and 60 min respectively, were considered following the hypothesis that terrorist attacks should be massive. Furthermore, only one representative starting time was accounted for at 8 a.m., because the network shows a single operating condition (i.e., no flow inversion at any pipes). All the 536 demanding nodes of the WDS were considered possible injection locations. The overall number of contamination events was then equal to 536.

Like in the first case study, the system water demand was assumed to vary with hourly steps. Therefore 1 day long patterns were used for the demand multiplying coefficients and the simulations were run for 3 days.

In this case study, the sampling was carried out only on the injection location and the optimizations were carried out to search for solutions up to a number of sensors equal to 50.

Results and discussion

This section presents the results for the bi-objective optimal placement of sensors, aimed at minimizing simultaneously the sensor redundancy and the contaminated population in both case studies. As for the first case study, the scenario where the total number of contamination events is considered was indicated as S0. The method proposed for sampling the events was applied considering various scenarios (S1, S1a, S1b, S1c, S1d and S2) to reduce the size of the contamination and time matrices. The sampling was done considering frequencies of one out of two or one out of three for the injection location, the representative starting times of 0 h, 5 h and 18 h, the intermediate mass injection rate of 200 gr/min and the smallest event duration of 60 min. Scenarios S1, S1a, S1b, S1c and S1d were obtained considering the location sampling frequency of one out of two nodes whereas scenario S2 was obtained considering the location sampling frequency of one out of three nodes. Subsequently, 23 and 19 possible injection locations were sampled in S1, S1a, S1b, S1c, S1d on the one hand (see Figure 6a) and S2 on the other hand (see Figure 6b), respectively. Furthermore, in S1 and S2, all the other variables than the injection location, i.e., starting time, mass rate and duration, were sampled at the same time. In S1a, S1b, S1c and S1d, instead, the sampling concerned one variable at a time. The features of all the scenarios of sampling are reported in Table 2. This table shows that, compared to scenario S0, a large reduction in the number of events is obtained through the sampling method, above all in scenarios S1 and S2 operating on all the sampled variables at the same time. In fact, scenarios S1 and S2 are made up of 69 and 57 events, respectively, which are smaller than the number of events in S0 (42,240) by three orders of magnitude.

GA applications enabled deriving the Pareto fronts of optimal solutions in the trade-off between the sensor redundancy and contaminated population in all the scenarios. The graphs in Figure 7 report the Pareto front obtained in scenarios S1a, S1b, S1c and S1d in comparison with that of S0. As expected, each front shows decreasing values of the contaminated population as the sensor redundancy, and therefore the number of installed sensors, increases. Furthermore, the best benefits in terms of contaminated population are obtained up to a redundancy of 2.5 sensors. The analysis of the graphs in Figure 7 points out that the fronts obtained in scenarios S1a, S1b, S1c and S1d are close to the S0 front. As was expected considering the linearity of the advection-reaction equations, the fronts are almost coincident in graph c associated with the mass sampling. However, in the other cases, the differences between the fronts are small. To compare better the results obtained in S1a, S1b, S1c, S1d and S0, the optimal sensor locations obtained in S1a, S1b, S1c and S1d were tested in S0, considering the totality of contamination events. This led to the curve of revalued solutions in the graphs in Figure 7. The closeness of these curves to the Pareto front of S0 attests to the fact that the performance of the optimal sensor location solutions obtained in S1a, S1b, S1c and S1d does not decay when tested against the totality of events of S0. Similar remarks can be made as for the comparison of scenarios S1 and S2, featuring sampling on all the variables, with scenario S0 (see Figure 8).

As an example of the obtained solutions, Figure 9 shows identical locations of 4 sensors in the first case study in scenarios S0 and S1, with values of f_1 and f_2 equal to 0.63 and 772.21, respectively, assessed based on the total number of contamination events (S0). As Figure 9 shows, one of the 4 sensors is located close to the water input whereas the other 3 are halfway between the input and the tank. The 4 locations were selected by the optimizer to promptly detect the generic contamination event, wherever it takes place in the network, and in an attempt to compromise the contaminated population with the sensor redundancy. The results shown in Figure 9 corroborate the previous findings concerning the representativeness of the sampled events and the effectiveness of the sampling method.

Two scenarios (S0 and S1) were considered in the second case study. In particular, S0 is the scenario with the totality of 536 events. Scenario S1, made up of 274 events, was obtained by sampling the network nodes with a one out of two frequency and considering the same values of injection starting time, duration and mass as S0. The Pareto fronts and the curve of revalued solutions of the second case study are reported in the graph in Figure 10. In this graph, f_2 is expressed as percentage over the total population. The closeness of the S0 and S1 Pareto fronts to the curve of revalued solutions attests to the effectiveness of the proposed sampling method. As an additional example of the results of the second case study, Figure 11 shows the optimal location of 10 sensors in the network in scenarios S0 and S1. Though the locations of the sensors in S0 and S1 may seem quite different, they feature similar value of the objective functions re-valued in S0 (f_1=0.13 and f_2=4.79% for the former solution and f_1=0.13 and f_2=4.71% for the latter solution). The results show a clear location of sensors near the water sources and halfway between the sources.

Conclusions

In this work, a procedure for sampling the most representative contamination events was proposed in the framework of the optimal sensor placement. The procedure, which is light from the computational viewpoint, was developed based on practical considerations on network topology and operation.

Applications concerned two case studies of different complexity, in which the procedure was applied. As a proof of its sampling effectiveness, the Pareto fronts of optimal solutions in the trade-off between the sensor redundancy and the contaminated population, both to be minimized, did not vary significantly when the sampled contamination events were used inside the optimization, instead of the totality of possible contamination events. For the sensor locations obtained with and without the sampling, the equifinality and the similar distances from the sources represent further evidence of the sampling effectiveness.

Rather than considering only one contaminant injection for each single node, an extension of the study can involve multiple and simultaneous injections of contaminant at multiple points. Other possible prospects could concern the assessment of the extent to which the sensor placements obtained are effective when unconservative contaminants, such as the real ones, are considered in the analysis. Furthermore, the analysis of spreading contamination during short transients with significant pipe flow changes, which would require use of unsteady flow modelling, is another field to explore.

References

Al-Zahrani, M., and Moied, K. (2001). "Locating optimum water quality monitoring stations in water distribution system." In: Bridging the gap: Meeting the World's Water and Environmental Resources Challenges, *Proc. of the ASCE EWRI annual conference*, Orlando, Florida, published on CD (2011).

Arad, J., Housh, M., Perelman, L., and Ostfeld, A. (2013). "A dynamic thresholds scheme for contaminant event detection in water distribution systems." *Water Res*. 47 (5), 1899-1908.

Berry, J. W., Hart, W. E., Phillips, C. A., Uber, J. G., and Watson, J. P. (2006). "Sensor placement in municipal water networks with temporal integer programming models". *J. Water Resour. Plann. Manage.*, 132(4), 218–224.

Berry, J., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., and Watson, JP. (2009). "Designing contamination warning systems for municipal water networks using imperfect sensors." *J. Water Resour. Plann. and Manage.*, 135(4), 253–263.

Braun M., Bernard T., Ung H., Piller O., and Gilbert D. (2015). "CFD Modeling of Contaminant Mixing at Junctions for an online Security Management Toolkit in Water Distribution Networks." *J. of Water Supply: Research and Technology-AQUA*, IWA, 64(5), 504–515.

Chang, N. B., Pongsanone, N. P., and Ernest, A. (2012). "Optimal sensor deployment in a large-scale complex drinking water distribution network: comparisons between a rule-based decision support system and optimization models." *Comput. Chem. Eng.* 43, 191-199.

Cheifetz, N., Sandraza A. C., Feliers C., Gilbert D., Piller O., and Lang A. (2015). "An Incremental Sensor Placement Optimization in a Large Real-World Water System." *Procedia Engineering* 119(2015), 947 – 952.

Creaco E., and Franchini M. (2012). "Fast network multi-objective design algorithm combined with an a-posteriori procedure for reliability evaluation under various operational scenarios." *Urban Water Journal*, 9(6), 385-399.

Creaco E., and Pezzinga G. (2015). "Embedding Linear Programming in Multi Objective Genetic Algorithms for Reducing the Size of the Search Space with Application to Leakage Minimization in Water Distribution Networks." *Environmental Modelling & Software*, 69, 308-318.

Deb, K., Pratap, A., Agrawal, S., and Meyarivan, T. (2002)."A Fast and Elitist Multiobjective Genetic Algorithm: NGSA-II." *IEEE Transactions on evolutionary computation.* 6 (2), April 2002.

Deuerlein, J., Piller, O., and Montalvo, I. (2014). "Improved Real-time Monitoring and Control of Water Supply Networks by Use of Graph Decomposition." *Procedia Engineering*, 89(0), 1276-1281.

Diao, K., and Rauch, W. (2013). "Controllability analysis as a pre-selection method for sensor placement in water distribution systems." *Water Res.* 47(16), 6097-6108.

Dorini, G., Jonkergouw, P., Kapelan, Z., Di Pierro, F., Khu, S. T., and Savic, D. (2006). "An efficient algorithm for sensor placement inwater distribution systems." In: *Proc., 8th Annual Water Distribution Systems Analysis Symp.*, ASCE, Reston, Va.

EPA, 2012. "CANARY User's Manual Version 4.3.2." United States Environmental Protection Agency (USEPA), Washington, DC. EPA/600/R-08/040B, 2012. (accessed 05.06.15.).

Eliades, D., and Polycarpou, M. (2006). "Iterative deepening of Pareto solutions in water sensor Networks." *Proc., 8th Annual Water Distribution System Analysis Symp*., Cincinnati.

Gueli, R. (2006). "Predator–prey model for discrete sensor placement." *Proc., 8th Annual Water Distribution System Analysis Symp.*, Cincinnati.

Guidorzi M., Franchini M., and Alvisi S. (2009) "A multi-objective approach for detecting and responding to accidental and intentional contamination events in water distribution systems", *Urban Water J.*, Vol. 6, (2), 115-135.

Huang, J., McBean, E. A., and James, W. (2006). "Multiobjective optimization for monitoring sensor placement in water distribution systems.", *Proc., 8th Annual Water Distribution System Analysis Symp.*, Cincinnati.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks." *J. Water Resour. Plann. Manage.*, 124 (4), 192–198.

Kruskal, J. B. (1956). "On the shortest spanning subtree of a graph and the traveling salesman problem". *Proc. of the American Mathematical Society*, 7: 48–50, JSTOR 2033241.

Kumar, A., Kansal, M. L., and Arora, G. (1997). "Identification of monitoring stations in water distribution system." *J. Environ. Eng*. 123 (8), 746-752.

Lee, B. H., and Deininger, R. A. (1992). "Optimal locations of monitoring stations in water distribution systems." *J. Environ. Eng.* 118 (1), 4-16.

Marchi A., Salomons E., Ostfeld A., Kapelan Z., Simpson A. R., Zecchin A. C., Maier H.R., Wu Z. Y., Elsayed S. M., Song Y., Walski T., Stokes C., Wu W., Dandy G. C., Alvisi S., Creaco E., Franchini M., Saldarriaga J., Páez D., Hernández D., Bohórquez J., Bent R., Coffrin C., Judi D., McPherson T., van Hentenryck P., Matos J.P., Monteiro A.J., Matias N., Yoo D.G., Lee H.M., Kim J.H., Iglesias-Rey P. L., Martínez-Solano F.J., Mora-Meliá D., Ribelles-Aguilar J.V., Guidolin M., Fu G., Reed P., Wang Q., Liu H., McClymont K., Johns M., Keedwell E., Kandiah V., Jasper M.N., Drake K., Shafiee E., Barandouzi M. A., Berglund A. D., Brill D., Mahinthakumar G., Ranjithan R., Zechman E. M., Morley M. S., Tricarico C., De Marinis G., Tolson B. A., Khedr A., and Asadzadeh M. (2014). "The Battle of the Water Networks II (BWN-II)." *J. Water Resour. Plann. Manage.*, 140(7), 04014009-1-14.

McKenna, S. A., Hart, D. B., and Yarrington, L. (2006). "Impact of sensor detection limits on protecting water distribution systems from contamination events." *J. Water Resour. Plann. Manage.*, 132(4), 305–309.

Oliker, N., and Ostfeld, A. (2014a). "A coupled classification-evolutionary optimization model for contamination event detection in water distribution systems." *Water Res.* 51 (15), 234-245.

Oliker, N., and Ostfeld, A. (2014b). "Minimum volume ellipsoid classification model for contamination event detection in water distribution systems." *J. Environ. Model. Softw*. 57, 1-12.

Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *J. Water Resour. Plann. Manage.*, 130(5), 377–385.

Ostfeld, A., and Salomons, E. (2005). "Securing water distribution systems using online contamination monitoring." *J. Water Resour. Plann. Manage.*, 131(5), 402–405.

Ostfeld, A., and Salomons, E. (2006). "Sensor network design proposal for the battle of the water sensor networks (BWSN)." *Proc., 8th Annual Water Distribution System Analysis Symp.*, Cincinnati.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J. P., Dorini, G., Jonkergouw P., Kapelan, Z., Di Pierro, F., Khu, S. T., Savic, D., Eliades, D., Polycarpou, M., Ghimire S. R., Barkdoll B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y., and

Walski, T. (2008). "The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms." *J. Water Resour. Plann. Manag.* 134 (6), 556-568.

Perelman, L., Arad, J., Housh, M., and Ostfeld, A. (2012). "Event detection in water distribution systems from multivariate water quality time series." *Environ. Sci. Technol*. 46 (15), 8212-8219.

Perelman, L., and Ostfeld, A. (2011) "Topological clustering for water distribution systems analysis." *J. Env. Soft.* 26, 969-972.

Piller, O., Deuerlein, J., Gilbert, D., and Weber, J.-M. (2015). "Installing Fixed Sensors for Double Calibration and Early-warning Detection Purposes." *Procedia Engineering,* 119, 564-572.

Preis, A., and Ostfeld, A. (2008). "Multiobjective contaminant sensor network design for water distribution systems." *J. Water Resour. Plann. Manag.* 134 (4), 366

Propato, M. (2006). "Contamination warning in water networks: General mixed-integer linear models for sensor location design." *J. Water Resour. Plann. Manage.*, 132(4), 225–233.

Propato, M., and Piller, O. (2006). "Battle of the Water Sensor Networks." *8th annual Water Distribution System Analysis Symposium*, University of Cincinnatti, Cincinnati, Ohio USA, August 27-30 2006, printed by ASCE, 8.

Rathi, S., and Gupta, R. (2016). "A Simple Sensor Placement Approach for Regular Monitoring and Contamination Detection in Water Distribution Networks." *KSCE J. Civil Eng.*, 20(2), 597-608.

Rossman, L. A. (2000). "EPANET 2.00.10." United States Environmental Protection Agency (USEPA), Washington, D.C., EPA/600/R-00/057, https://www.epa.gov/water-research/epanet

Shastri, Y., and Diwekar, U. (2006). "Sensor placement in water networks: A stochastic programming approach." *J. Water Resour. Plann. Manage.*, 132, 192–203.

Weickgenannt, M., Kapelan, Z., Blokker, M., and Savic, D. A. (2010) "Risk-Based Sensor Placement for Contaminant Detection in Water Distribution System". *J. Water Resour. Plann. Manage.*, 136(6), 629–636.

Woo, H. M., Yoon, J. H., and Choi, D. Y. (2001). "Optimal monitoring sites based on water quality and quantity in water distribution systems." In: *World Water and Environmental Resources Congress*, CD-ROM, ASCE, Reston, Va.

Wu, Z. Y., and Walski, T. (2006). "Multiobjective optimization of sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution System Analysis Symp.*, Cincinnati.

Zhao, Y., Schwartz, R., Salomons, E., Ostfeld, A., and Poor, H. V. (2016) "New formulation and optimization methods for water sensor placement", *Environmental Modelling & Software* 76, 128-136.

Figure captions

Figure 1. Selected nodes considering a sampling frequency for the injection nodes equal to 2 in an explicative WDS. The source node is indicated with a box and the dashed line indicates the pipe that is removed for loop opening. Node numbers close to the nodes. Pipe numbers inside circles and close to the pipes.

Figure 2. Water discharge in a WDS pipe during the typical day.

Figure 3. Explicative water distribution network for sampling contamination mass rates.

Figure 4. Trend of the contaminant concentration at node 7 in response to the injected masses of 50 gr/min and 200 gr/min in the explicative water distribution network shown in Figure 3.

Figure 5. Layout of the network of the a) first and b) second case studies. The arrow indicates water input from a pump.

Figure 6. Selected nodes considering a sampling frequency of the injection nodes equal to a) one out of two (Scenarios S1, S1a, S1b, S1c, S1d) and b) one out of three (Scenario S2) in the first case study.

Figure 7. First case study. Pareto fronts of optimal solutions in the trade-off between the sensor redundancy (f_1) and contaminated population (f_2) in scenarios S0 and a) S1a b) S1b c) S1c d) S1d. Curves of optimal S1a, S1b, S1c and S1d solutions revalued in scenario S0.

Figure 8. First case study. Pareto fronts of optimal solutions in the trade-off between the sensor redundancy (f_1) and contaminated population (f_2) in scenarios S0 and a) S1 b) S2. Curves of S1 and S2 optimal solutions revalued in scenario S0.

Figure 9. First case study. Optimal location of 4 sensors in scenarios S0 and S1.

Figure 10. Second case study. Pareto fronts of optimal solutions in the trade-off between the sensor redundancy (f_1) and contaminated population percentage (f_2) in scenarios S0 and S1. Curves of S1 optimal solutions revalued in scenario S0.

Figure 11. Second case study. Optimal location of 10 sensors in scenarios S0 and S1.

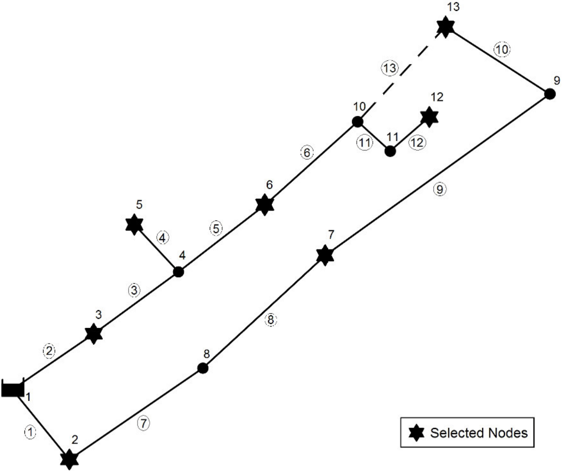
Tables

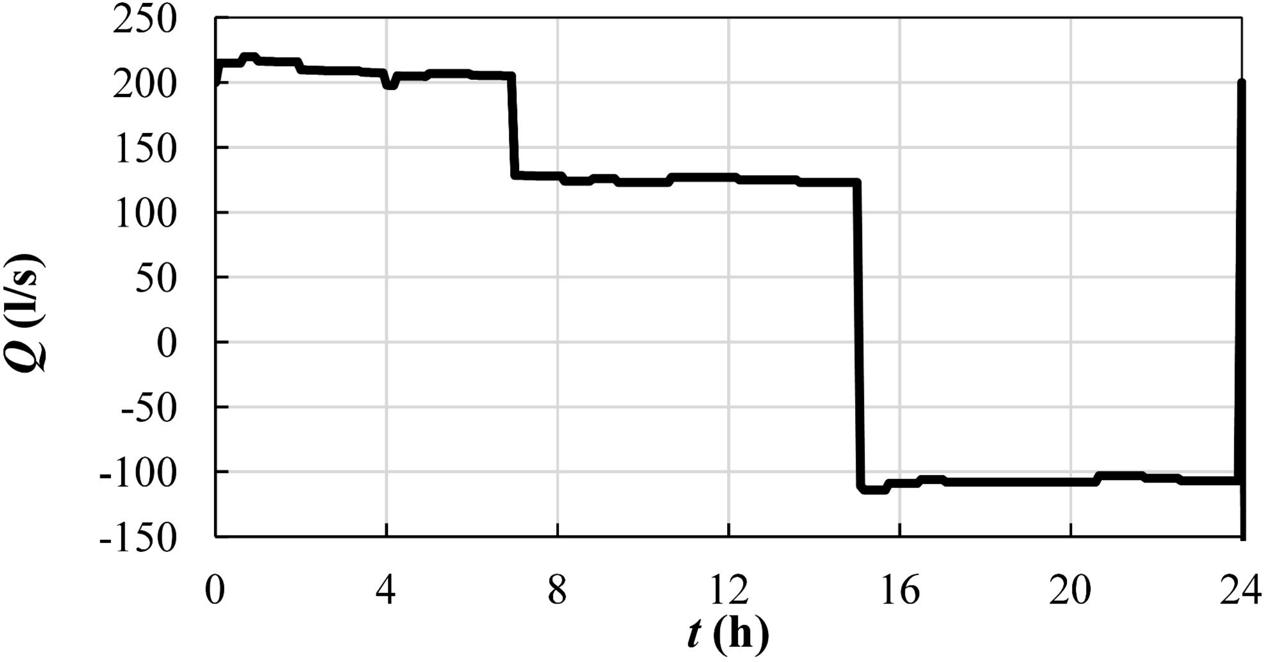
Table 1. Inhabitants connected to the nodes in the first case study.

Node	Inhabitants	Node	Inhabitants
1	0	24	282
2	213	25	165
3	288	26	271
4	341	27	215
5	353	28	300
6	100	29	7
7	59	30	38
8	233	31	46
9	148	32	0
10	149	33	193
11	196	34	237
12	330	35	196
13	167	36	298
14	97	37	32
15	20	38	35
16	88	39	160
17	352	40	314
18	22	41	270
19	141	42	220
20	131	43	135
21	182	44	93
22	141	45	0
23	39	46	0

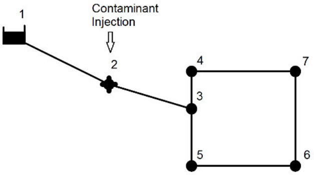
Table 2. Features of the sampling scenarios in the first case study.

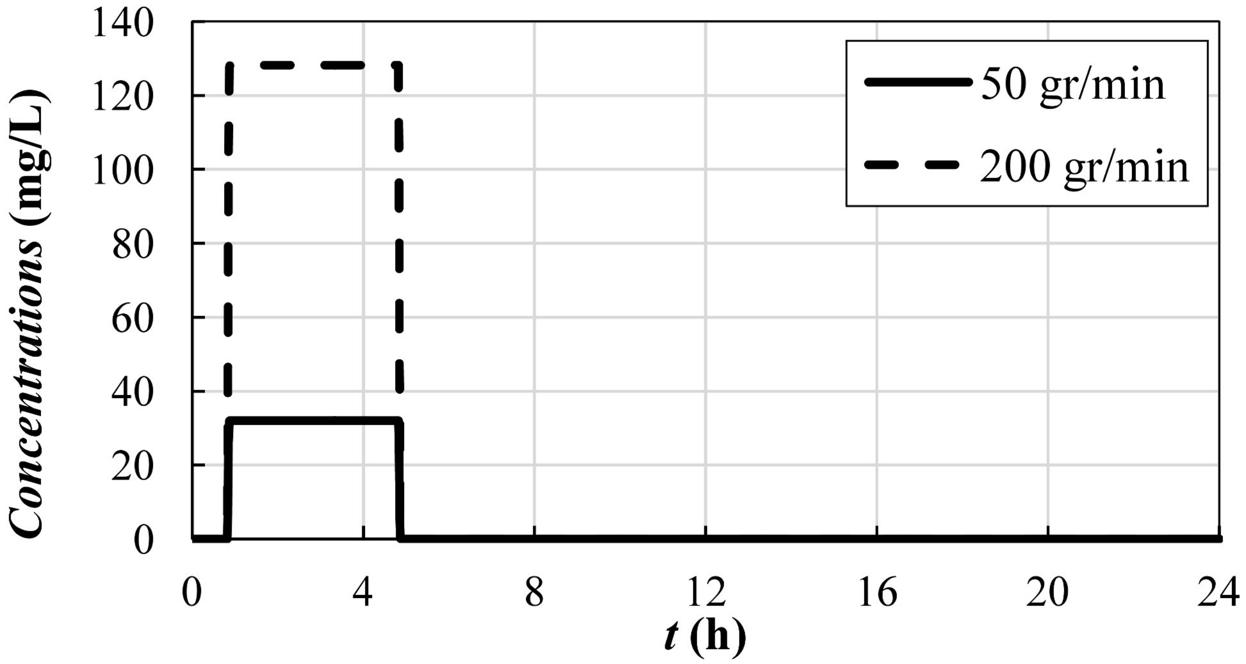
scenario	injection sampling with one out of two frequency	injection sampling with one out of three frequency	starting time sampling	mass rate sampling	duration sampling	number of events
S0						42,240
S1	X		X	X	X	69
S1a	X					22,080
S1b			X			2,640
S1c				X		10,560
S1d					X	8,448
S2		X	X	X	X	57


13
10
13
9
12
10
11
11
12
6
6
9
5
4
5
7
4
3
3
8
2
8
1
1
7
2
Selected Nodes

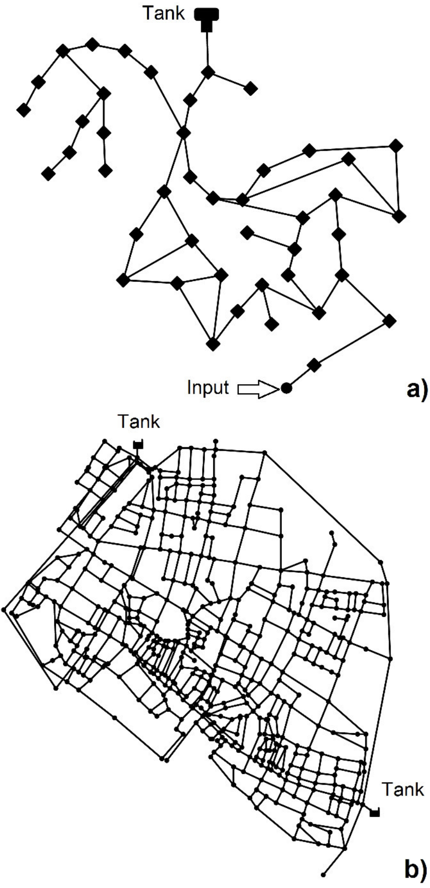


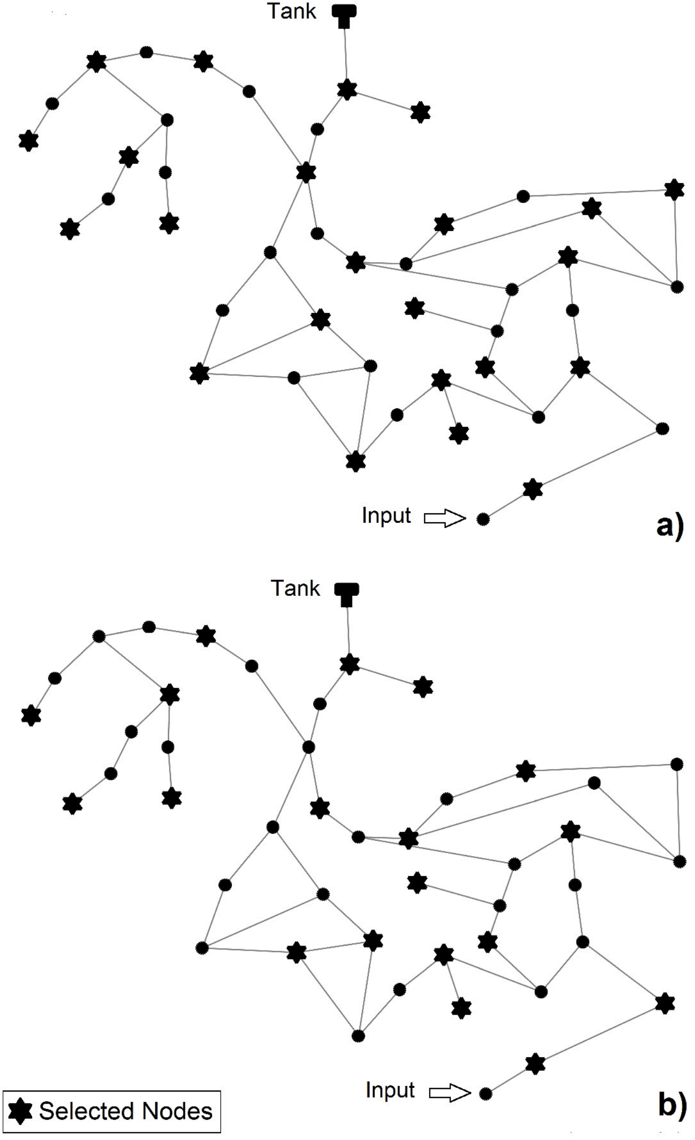
250
200
150
100
50
0
-50
-100
-150
Q (l/s)
0
4
8
12
16
20
24
t (h)

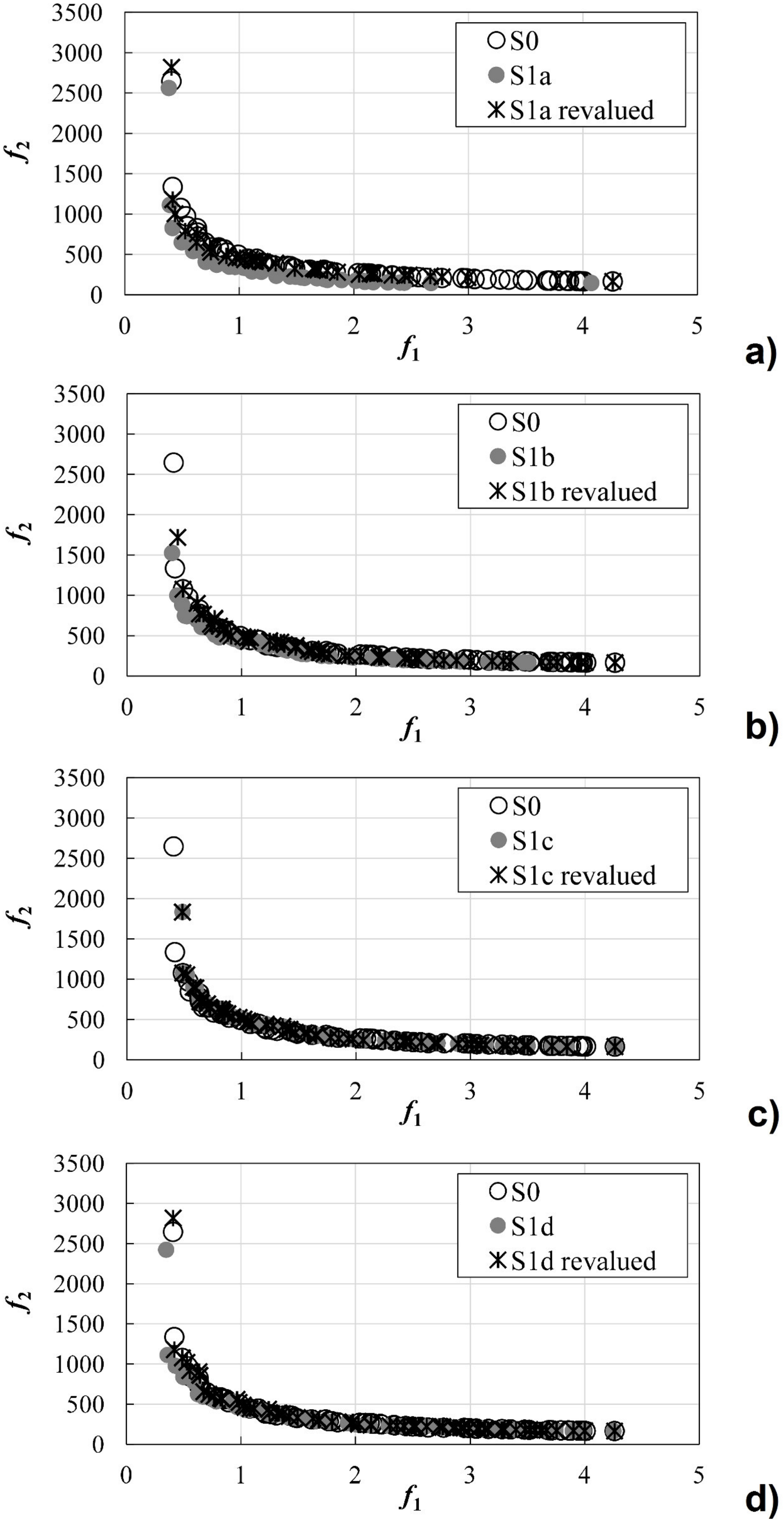

Contaminant
Injection
1
2
3
4
5
6
7

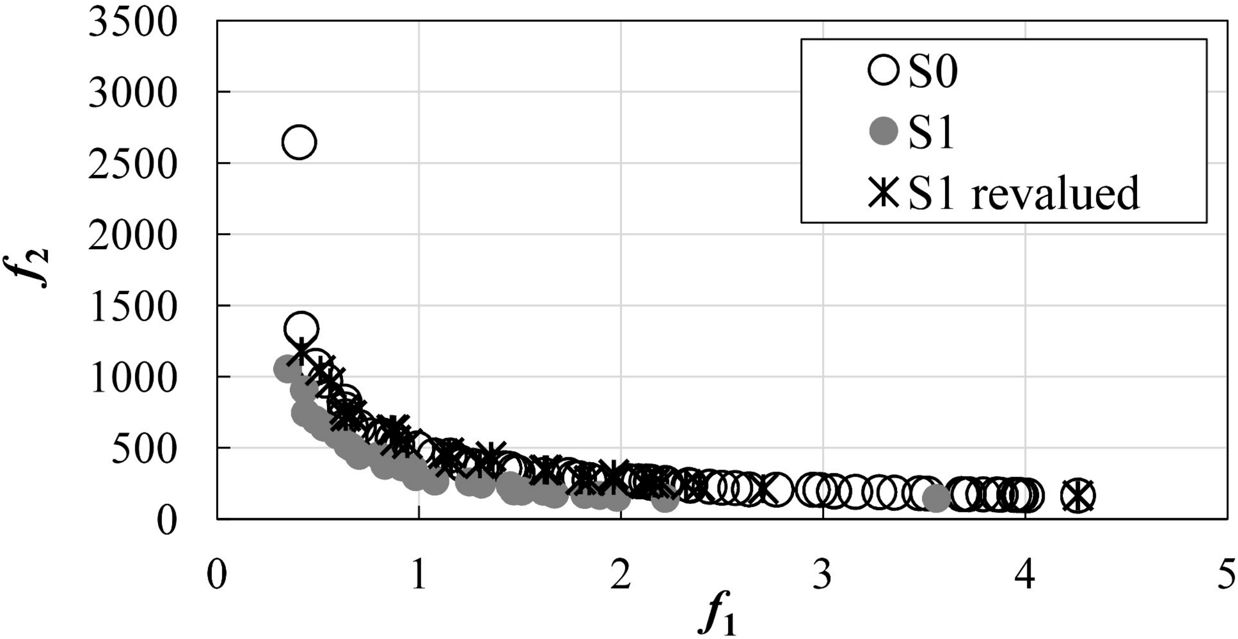


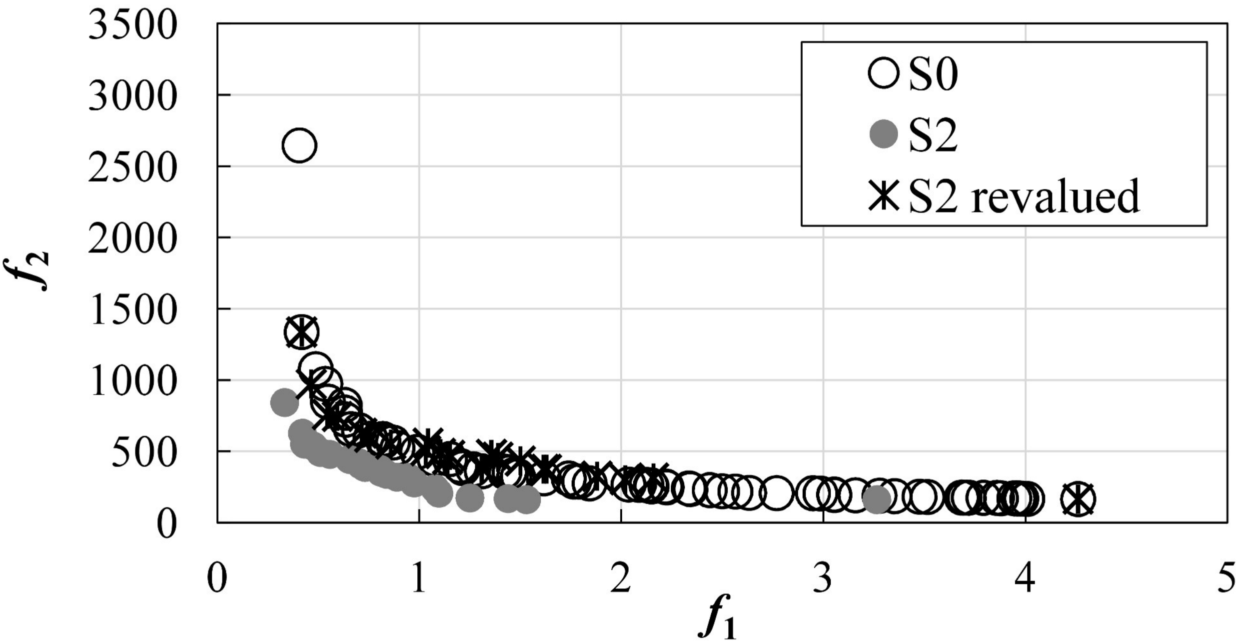
140
120
100
80
60
40
20
0
Concentrations (mg/L)
0
4
8
12
16
20
24
t (h)
50 gr/min
200 gr/min

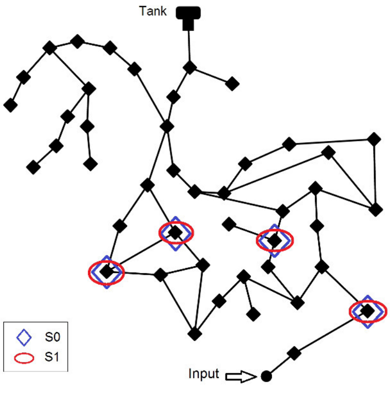

Tank
Input
a)
Tank
Tank
b)

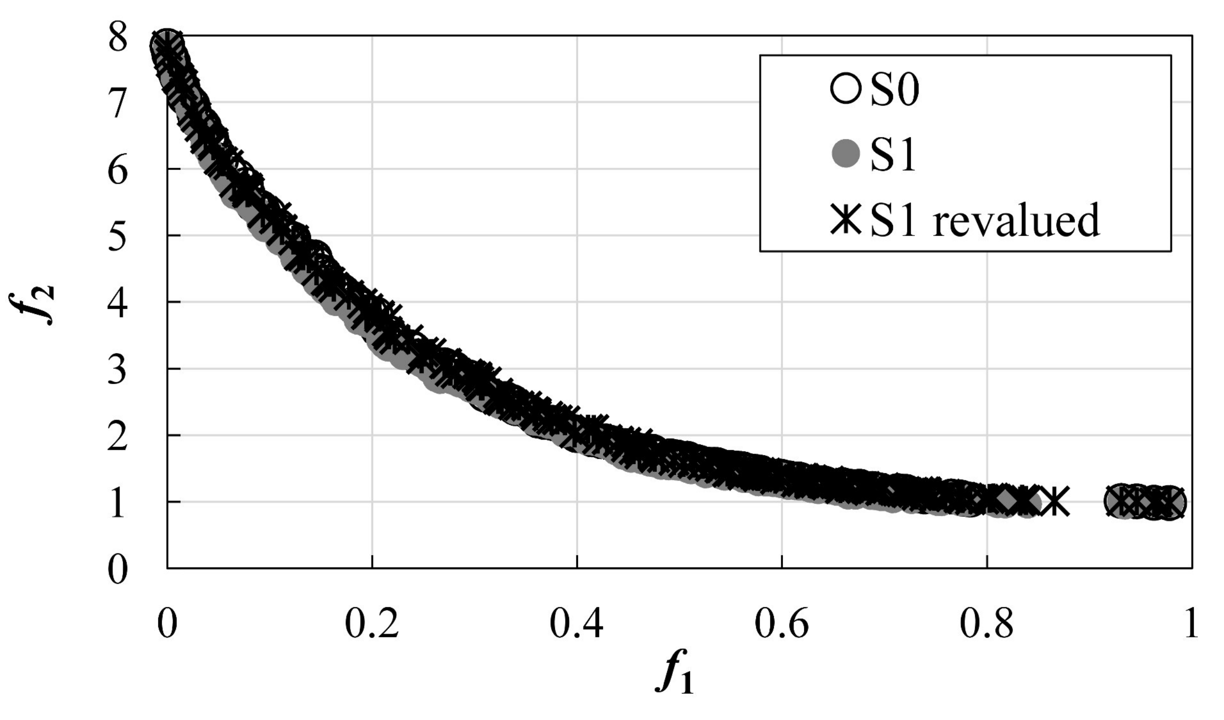

Tank
Input
a)
Tank
Input
b)
Selected Nodes


3500
3000
2500
2000
1500
1000
500
0
f_2
0
1
2
3
4
5
f_1
S0
S1a
S1a revalued
a)
S0
S1b
S1b revalued
b)
S0
S1c
S1c revalued
c)
S0
S1d
S1d revalued
d)

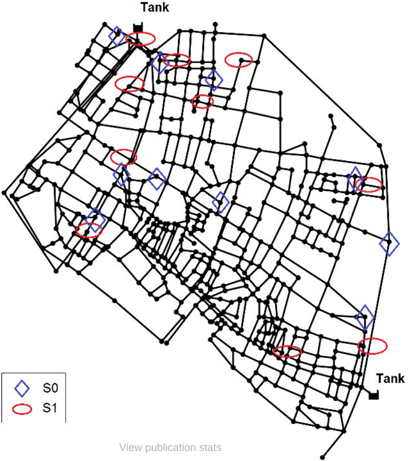

S0
S1
S1 revalued
f_2
f_1
3500
3000
2500
2000
1500
1000
500
0
0
1
2
3
4
5
a)


S0
S2
S2 revalued
f_2
f_1
3500
3000
2500
2000
1500
1000
500
0
0
1
2
3
4
5
b)


Tank
S0
S1
Input



S0
S1
S1 revalued
f_2
f_1
0
1
2
3
4
5
6
7
8
0
0.2
0.4
0.6
0.8
1


Tank
Tank
S0
S1