

Article

Greedy Algorithms for Sensor Location in Sewer Systems

Bijit K. Banik ¹, Leonardo Alfonso ², Cristiana Di Cristo ^{3,*} and Angelo Leopardi ³

¹ Department of Civil and Environmental Engineering, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh; bijit-cee@sust.edu

² IHE-Delft, P.O. Box 3015, 2601 DA Delft, The Netherlands; l.alfonso@un-ihe.org

³ Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, 03043 Cassino, Italy; a.leopardi@unicas.it

* Correspondence: dicristo@unicas.it; Tel.: +39-0776-299-4340

Received: 4 September 2017; Accepted: 31 October 2017; Published: 4 November 2017

Abstract: Wastewater quality monitoring is receiving growing interest with the necessity of developing new strategies for controlling accidental and intentional illicit intrusions. In designing a monitoring network, a crucial aspect is represented by the sensors' location. In this study, a methodology for the optimal placement of wastewater monitoring sensors in sewer systems is presented. The sensor location is formulated as an optimization problem solved using greedy algorithms (GRs). The Storm Water Management Model (SWMM) was used to perform hydraulic and water-quality simulations. Six different procedures characterized by different fitness functions are presented and compared. The performances of the procedures are tested on a real sewer system, demonstrating the suitability of GRs for the sensor-placement problem. The results show a robustness of the methodology with respect to the detection concentration parameter, and they suggest that procedures with multiple objectives into a single fitness function give better results. A further comparison is performed using previously developed multi-objective procedures with multiple fitness functions solved using a genetic algorithm (GA), indicating better performances of the GR. The existing monitoring network, realized without the application of any sensor design, is always suboptimal.

Keywords: sewer system; sensor location; greedy algorithm; optimization; illicit intrusion

1. Introduction

Wastewater management is receiving growing interest because sewers are not only simple sanitary and flood control systems, but they have an overall environmental management function [1]. Many countries (e.g., the United States and European Union (EU) members) are enforcing new policies for regulating discharges into sewers, but these systems are very vulnerable to illicit intrusions because the collection networks are geographically dispersed and have multiple access points. For this reason, researchers understand the necessity of developing new strategies for wastewater quality monitoring [2] and for controlling accidental and intentional illicit intrusions. In particular, the goal is to develop methods to (1) individuate quickly an illicit intrusion in the system; (2) identify the possible sources; and (3) assess the possible impacts on the treatment plant and/or final receiving water bodies. In all these cases, wastewater quality measurements are necessary. This work investigates the optimal placement of wastewater monitoring sensors in sewer systems for controlling illicit intrusions, solving an optimization problem using greedy algorithms (GRs).

The early studies [3] and [4] presented procedures for individuating illicit injections in a separate storm drainage system using sampling and analytical laboratory analyses. The development of on-line sensors for wastewater quality monitoring [5,6] made possible the implementation of new

methods. For example, using on-line pollutant concentration measurements [7,8] has been proposed as a methodology for the identification of an illicit intrusion source in a sewer system, solving an optimization problem.

In designing a monitoring network, a crucial aspect is represented by the sensors' location. In fact, to contain the number of monitoring stations, reducing the costs in this way, it is important to design the sensors' placement optimally. This problem has been addressed in various fields of water resources engineering, such as river systems (e.g., [9,10]), polder systems (e.g., [11]), water distribution systems (e.g., [12]), and so forth.

For river systems, the sensors' location problem has been addressed using different approaches, such as statistical methods (e.g., [13]), direct surveys (e.g., [14]), optimization methods (e.g., [15]) and information theory applications (e.g., [9]). A comprehensive review is presented by [16].

In the case of water distribution systems, the sensor location has been mainly formulated as an optimization problem. As summarized in the review by [17], many methodologies have been proposed with different objective functions, such as the detection time, the volume of contaminated water consumed, the population exposed, the extent of contamination, the associated risk, the detection likelihood, the probability of failed detection, the sensor response time and the sensor detection redundancy. The different objectives may be applied either separately (single-objective procedure) or simultaneously (multi-objective procedure) in the optimization formulation. Some multi-objective approaches consider different objectives grouped together in a single function (e.g., [12,18,19]), while in other formulations they remain distinct (e.g., [18,20–22]). In the latter procedures, a group of solutions are reported in the form of Pareto front without individuating the single best solution to implement.

In particular, in [12], four design objectives (expected time of detection, expected population affected prior to detection, expected demand of contaminated water prior detection, and reliability) are considered in a single function, to mimic a multi-objective approach and to obtain one final solution. The proposed procedure has been applied to two different case studies of different complexity. Similarly, in [19], the objectives of demand coverage and time-constrained detection likelihood are combined into a single function and different weights are assigned to them depending on the necessity of the supply authorities. The methodology has been applied to a benchmark problem, obtaining several solutions by varying the weights of the two objectives.

Among the methodologies assuming separate objective functions, in [20], network detection likelihood, redundancy and expected detection time are considered, and tradeoff curves are obtained simultaneously for all three objectives and for exploring pairs of objectives. The procedure has been tested on the real water system of Richmond. Moreover, in [21], the sensor location is formulated as a twin-objective optimization problem, and the objectives, the minimization of the number of sensors and the risk of contamination, are considered. The methodology has been tested on the complex distribution system of Almelo, and the estimated Pareto front suggests that a reasonable level of contaminant protection can be achieved using a small number of strategically located sensors. In [22], the authors considered two competitive objectives: the minimization of the delay time and the maximization of sensor redundancy. The study applied on the distribution network of the city of Guelph shows that the evaluation of the Pareto fronts' performance indicates five as the number of sensors needed. Finally, in [18], a methodology entitled Sensor Location Optimal Transformation System (SLOTS) to address both single- and multi-objective sensor location problems is proposed. The SLOTS has been tested on two benchmark water distribution networks, considering as objectives the detection likelihood and the expected population affected prior to detection.

To solve the associated optimization problem, genetic algorithms (GAs), such as the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [23], are usually used. However, among the different available solvers, the greedy algorithms (GRs) represent very interesting and efficient methods, which are usually simpler and computationally less expensive than other heuristics methods. Although little evidence of the application of GRs has been reported in the current literature on water resources, in [9], a promising outcome from a rank-based greedy methodology for designing discharge monitoring in

rivers has been highlighted. Although this method was used for checking the quality of the Pareto optimal solutions derived from a multi-objective approach at its extreme ends, the potentiality of the GR is hinted at.

Recently, in [24,25], some methodologies for optimally designing a monitoring network in sewer systems have been proposed. The performances of different multi-objective formulations, characterized by distinct objective functions and solved with the NSGA-II algorithm, are evaluated. In [25], a further comparison has been performed with a single-objective rank-based GR procedure, confirming the efficiency of this approach in finding the extreme Pareto solutions.

The main novelty of the presented research is the use of GRs for solving a sensor location problem. New methodologies for locating sensors in sewer systems, formulated as rank-based GR optimization problems, are proposed. The improvement of this work with respect to the previous work by [25] is represented by the comparison of six different GR procedures: three single-objective formulations are compared with three multi-objective formulations. In fact, to the best of our knowledge, no study has been reported that considers multiple objectives in greedy optimization.

The main goals of the research are to test the applicability of GRs to sensor location, showing their potentialities and limitations, and to compare the performances of the different formulations proposed for optimally designing a monitoring network in sewer systems. Four different design objectives are considered, combined in different ways in the six procedures. The procedures are applied to a real case-study represented by the sewer system of Massa Lubrense, a town located near Napoli, Italy.

2. Methodology

The goal of the presented methodology is to individuate the best sensor location of a previously fixed number of monitoring stations to detect any possible contamination scenario in a sewer system. Mathematically, a sewer network has M potential candidate nodes at which to place N sensors, with $M \geq N$. The solution vector Y consists of N monitoring stations, denoted as $Y = [y_1, y_2, \dots, y_i, \dots, y_N]$, where y_i is the original node index of sensor i . It is also assumed that a node can accommodate only one sensor.

The methodology is applied with the six different procedures listed in Table 1, indicated with GR1, GR2, GR3, GR4, GR5, GR6 and detailed described in the following paragraphs, which differ for the used GR and the adopted design objectives. The considered objectives, which are described in detail in the next subsection, are the detection time (D), the reliability (R), the joint entropy (JH) and the total correlation (TC). The first three formulations use a classical single-objective GR, herein indicated with GR_S, while the successive three implement an original multi-objective GR, indicated with GR_M. However, in the GR_M approach, the objectives are grouped into a single fitness function. In all the procedures, the single best solution is individuated; thus it is easier and fairer to make a comparison.

Table 1. The procedures and the required computational time (C-Time) for the test with 14 sensors and detection threshold (0.0001 mg/L).

Procedure	GR1	GR2	GR3	GR4	GR5	GR6	B_IT	B_DR
Algorithm	GR_S	GR_S	GR_S	GR_M	GR_M	GR_M	NSGA-II	NSGA-II
Objectives	D	R	JH	D, R	JH, TC	D, R, JH, TC	JH, TC	D, R
C-Time (s)	1.4	2.6	5200.4	3.8	2460.7	5205.3	143,415.2	1812.0

As more detailed description of the procedures, the data for evaluating the objectives were obtained by performing hydraulic and water-quality simulations using the well-known Storm Water Management Model (SWMM) software by USEPA (Environmental Protection Agency, USA). The proposed procedures are also compared with two procedures presented in [25], indicated as B_IT and B_DR, with the sensor location formulated as a multi-objective optimization problem solved using the Genetic Algorithm NSGA-II.

2.1. Design Objectives

The objectives of detection time (D) and reliability (R) are the two that are more frequently adopted in sensor location problems [17]; joint entropy (JH) and total correlation (TC) are quantities proposed in the information theory framework [26].

2.1.1. Detection Time (D)

The detection time is defined as the time between the beginning of a pollution event and the first non-zero concentration measurement by a sensor. Then, minimizing this objective means detecting the contamination event as quickly as possible with a fixed number of sensors.

For a contamination scenario s , the detection time of the i th monitoring sensor in the solution vector Y , $d_s^i(Y)$, is defined as the elapsed time between the starting time of a contamination event and the time at which the measurable concentration threshold at node y_i is exceeded. The detection time of the monitoring network, $D_s(Y)$, is defined as the shortest time among the detection times of the N monitoring sensors. It is mathematically expressed as

$$D_s(Y) = \min\{d_s^1(Y), d_s^2(Y), \dots, d_s^i(Y), \dots, d_s^N(Y)\} \quad (1)$$

To avoid dispositions with a high number of non-detected cases, a penalty to the non-detected scenarios is applied. For the non-detected scenarios, $D_s(Y)$ is assigned to be equal to the total simulation time, D_{sim} , obtaining

$$D_{sp}(Y) = \begin{cases} \min\{d_s^1(Y), d_s^2(Y), \dots, d_s^i(Y), \dots, d_s^N(Y)\} & \text{if scenario } s \text{ is detected} \\ D_{sim} & \text{otherwise} \end{cases} \quad (2)$$

The average detection time is calculated as the average of $D_{sp}(Y)$ over all possible scenarios:

$$D(Y) = \frac{1}{S} \sum_{s=1}^S D_{sp}(Y) \quad (3)$$

where S is the total number of scenarios considered in the analysis.

2.1.2. Reliability (R)

The reliability, or detection likelihood, of the sensors' network is related to the number of contamination scenarios correctly detected (e.g., [15,27]). Mathematically, the reliability of the solution Y , $R(Y)$, is defined as the ratio of detected contaminated scenarios to the total scenarios considered:

$$R(Y) = \frac{1}{S} \sum_{s=1}^S \delta_s \quad (4)$$

where $\delta_s = 1$ if the contamination scenario s is detected and $\delta_s = 0$ otherwise. A greater reliability corresponds to a greater number of detected scenarios.

2.1.3. Joint Entropy (JH)

In [26], the concept of entropy is introduced to measure the information content of a discrete random variable. Mathematically, for a discrete random variable X , with values x_1, x_2, \dots, x_n and corresponding probabilities of occurrence $p(x_1), p(x_2), \dots, p(x_n)$, the entropy is expressed as

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (5)$$

where n is the number of events of the random variable, which in the considered application is the number of records related to a concentration value x_i at a node X . The amount of information available within two variables (nodes equipped with a sensor) X_1 and X_2 is given by the joint entropy, JH , expressed by

$$JH(X_1, X_2) = - \sum_{i=1}^n \sum_{j=1}^m p(x_{1i}, x_{2j}) \log(p(x_{1i}, x_{2j})) \quad (6)$$

in which $p(x_{1i}, x_{2j})$ is the joint probability of the variables X_1 and X_2 , and n and m are the number of elementary events (measurements) in X_1 and X_2 , respectively. This definition is similarly extended to the N nodes.

In this paper, base 2 is used for the logarithm in Equation (6), and entropy is measured in bits [28]. The probabilities $p(x_i)$ are estimated using a histogram-based method with a given bin size or number of classes [9,16,29,30]. A higher entropy corresponds to a greater amount of information.

2.1.4. Total Correlation (TC)

Natural processes are always influenced by a large number of variables, which may be correlated. The total correlation, TC , concept [31,32] has been introduced to assess the dependencies among N variables. TC represents the amount of information shared by N variables (sensors), taking into account the dependencies between their partial combinations. Mathematically it is given by

$$TC(X_1, X_2, \dots, X_N) = \left(\sum_{i=1}^N H(X_i) \right) - JH(X_1, X_2, \dots, X_N) \quad (7)$$

The total correlation is measured in bits, as for the entropy. Minimizing this objective means reducing the correlated information. The objective of the problem being to maximize the information furnished by the sensors, the TC function is considered always in combination with JH . In fact, TC as a single objective furnishes solutions with less-correlated sensors, for example, terminal nodes, with a poor content of information.

2.2. The Proposed Procedures

The first three procedures (Table 1) implement the classical GR [9,33–35] with a single objective (GR_S). The decision variable that provides the best objective function value is chosen first. In the second step, the decision variable that, in combination with that first selected, gives the largest increment of the objective function is chosen. The procedure continues until the predefined number of decision variables has been chosen. For the case of sensor location, the decision variables are sensors.

As indicated in Table 1, the first three formulations consider the objectives D , R and JH respectively, one at a time, and their objective functions are mathematically expressed as

$$f_1 = \min\{D(Y)\} \quad (8)$$

$$f_2 = \max\{R(Y)\} \quad (9)$$

$$f_3 = \max\{JH(X_1, X_2, \dots, X_N)\} \quad (10)$$

The selection of a single objective being very difficult, multiple objective approaches are often used. However, when distinct fitness functions with different objectives are considered (e.g., [20–22]), many optimal solutions are reported in the form of a Pareto front, and then a further criterion has to be individuated to select which to implement. Differently, to obtain a single optimal solution, procedures 4, 5 and 6 use the GR_M approach with the optimization problem formulated considering one fitness function including different objectives. In these procedures, the fitness functions (detailed description given in the following) are formulated to be minimized and with a score in the range from 0 to 1. Different criteria are adopted for selecting the first sensor.

In procedure 4, the fitness function is composed of the two objectives D and R , and it is formulated as

$$f_4 = \min \left\{ \left[\left(1 - \frac{D_{max} - D}{D_{max} - D_{min}} \right) + \left(1 - \frac{R - R_{min}}{R_{max} - R_{min}} \right) \right] \div 2 \right\} \quad (11)$$

D_{max} and D_{min} are the maximum and the minimum detection time, assumed to be equal to the total simulation time and the reporting time step of the hydraulic simulation, respectively. Similarly, R_{max} and R_{min} are the maximum and minimum reliability of the system, respectively. The first sensor is chosen as that with the maximum reliability.

In the fifth procedure, the fitness function combines the objectives JH and TC and it reads

$$f_5 = \min \left\{ \left[\left(1 - \frac{TC_{max} - TC}{TC_{max} - TC_{min}} \right) + \left(1 - \frac{JH - JH_{min}}{JH_{max} - JH_{min}} \right) \right] \div 2 \right\} \quad (12)$$

TC_{max} and TC_{min} are the maximum and minimum total correlation of the system, respectively, while JH_{max} and JH_{min} are the maximum and minimum joint entropy of the system, respectively. In this formulation, the most informative sensor is chosen as the starting sensor.

The fitness function of procedure 6, considering all four objectives, is formulated as

$$f_6 = \min \left\{ \left[\left(1 - \frac{D_{max} - D}{D_{max} - D_{min}} \right) + \left(1 - \frac{R - R_{min}}{R_{max} - R_{min}} \right) + \left(1 - \frac{TC_{max} - TC}{TC_{max} - TC_{min}} \right) + \left(1 - \frac{JH - JH_{min}}{JH_{max} - JH_{min}} \right) \right] \div 4 \right\} \quad (13)$$

In this case, the starting sensor is that with the highest score in terms of both reliability and information content. In Equations (11)–(13), all the objectives are equally balanced, even if different weights could be assigned to give them a different importance.

2.3. Fitness Function Evaluation

In the proposed methodology, the sensor location is optimized to detect any possible contamination scenario. The required data to evaluate the fitness functions in GR are obtained by performing hydrodynamic and quality simulations through the USEPA's SWMM (<https://www.epa.gov/water-research/storm-water-management-model-swmm>). In this study, the contamination scenario is represented by a continuous injection with a fixed constant concentration of a conservative pollutant in a single node of the system for a fixed duration. The simplifying hypothesis of a conservative contaminant is assumed, because the absence of decay represents the most critical scenario.

For the hydraulic simulation, SWMM uses the equations for conservation of mass and momentum, in which Manning's formula is adopted. For the quality simulation, it is assumed that conduits behave as a continuously stirred tank reactor (CSTR) without considering the dispersion effect, which is assumed to be negligible [36]. Dry weather flow conditions (i.e., without rain) are assumed in the presented applications, as this represents a more impacting situation for the sewer function in the case of illicit intrusion.

To integrate the SWMM simulator within the methodology, the SWMM-Toolkit developed by [37] is used. For computing the fitness functions, the time series of the concentration data are extracted for each node.

For computing the objectives JH and TC through the histogram-based probability calculations, the data are quantized to convert all the records to integer numbers. Quantization [11] is a process to compile a continuous set of data to a discrete set. It rounds a value z to its nearest lowest integer multiple of k , namely, z_q :

$$z_q = \text{floor} \left(kz + \frac{1}{2} \right) \quad (14)$$

The function “*floor*” rounds down a decimal number to its nearest integer. The value of the parameter k is related to the threshold concentration detectable by a sensor, considering that their product has to be equal to 1.

3. Results and Discussion

3.1. Case Study

Massa Lubrense is a small town close to Napoli, Italy. The system, schematically shown in Figure 1, is a combined sewer with 12 subcatchments, covering an area of 19.71 km² and serving a population of 14,087 (2011) with an approximate volume of yearly produced wastewater of 1.13×10^6 m³. The scheme consists of 1909 circular conduits connecting 1902 junctions, 14 pumps, 14 storage units and 1 treatment plant. All geometric data, not reported herein, can be requested from the authors. The calibration of the input file was previously performed using discharge measurements, obtaining a good agreement between simulated results and measured data adopted; for all conduits, Manning’s roughness coefficient is equal to $0.016 \text{ m}^{-1/3} \cdot \text{s}$.

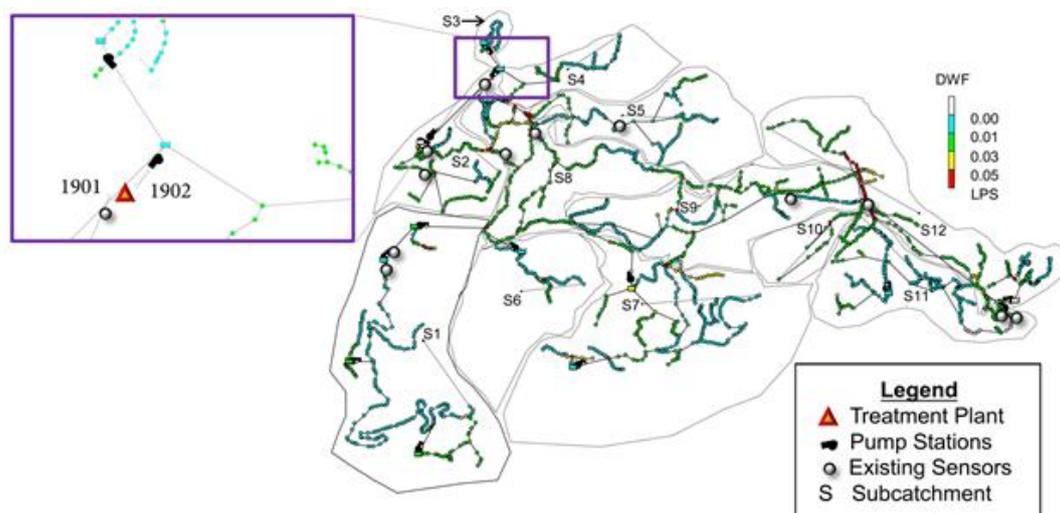


Figure 1. Scheme of the Massa Lubrense system.

The wastewater arrives to the treatment plant from two entry points, nodes 1901 and 1902.

Considering the population connected to each node, Figure 1 also depicts the daily mean dry weather flow (DWF) values, assigned to the 1866 nodes with an inflow.

The system has 12 monitoring stations that were installed as part of the ongoing Sistema Integrato per il Monitoraggio delle infrastrutture idrauliche e dell’Ambiente (SIMonA) project (www.progettosisimona.it). Their locations (Figure 1) have been decided on the basis of practical considerations, mainly related to the availability of the electrical power supply and to the need of the GSM (Global System for Mobile) coverage for transmitting the data.

In this study, the injection duration of the contamination scenario is selected considering the time that the solute takes to move between the two most distant points of the scheme, which is 5 h for the present application. The input concentration is fixed unitarily [38], but the results can be easily scaled for different values. The intrusion point can be any node of the system.

The SWMM hydraulic and quality simulations are run with a time step of 2 s for a duration of 6 h. Considering a reporting time step of 5 min, the size of the extracted time series is 137,952 at each node.

An important parameter to fix in the methodology is the minimum concentration (threshold) detectable by a monitoring station, which depends essentially on the type of sensor used. The values of all the considered objective functions depended on the detection threshold. For studying the effect

of these values on the results, five different threshold values are considered, namely, 0.1, 0.01, 0.001, 0.0001 and 0.00001 mg/L. Moreover, different tests are performed with a varying number of sensors, which is assumed between 1 and 14, and the number of monitoring stations already installed (12) in the range. In summary, as shown in Table 2, for each procedure, 70 tests are performed considering the number of sensors varying between 1 and 14 and five different thresholds.

Table 2. Performed tests.

Procedures	All
Number of sensors	From 1 to 14
Detection threshold (mg/L)	0.1, 0.01, 0.001, 0.0001, 0.00001

In applying procedures 4, 5 and 6, the maximum and minimum values of D , R , JH and TC are required (Equations (11)–(13)). For the Massa Lubrense case study, the maximum reliability is 97.39%, because only 1866 nodes out of the 1916 nodes in the system received the DWF. The minimum value of R is 0. The maximum and minimum D values are assumed to be equal to 360 min (the total simulation time) and 5 min (the reporting time step), respectively. The maximum possible joint entropy and total correlation are the corresponding system's values JH_{sys} and TC_{sys} , and they depend on the detection threshold. Table 3 reports JH_{sys} and TC_{sys} for the different considered thresholds. The minimum JH and TC are assumed to be equal to 1 and 0 bits, respectively. We note that $TC = 0$ means that the locations do not give redundant data from the information theory perspective.

Table 3. JH and TC of the system for different thresholds.

Detection Threshold (mg/L)	0.00001	0.0001	0.001	0.01	0.1
JH_{sys} (bits)	16.74	16.71	16.64	16.40	15.70
TC_{sys} (bits)	1895.16	1601.84	1270.35	948.65	685.80

In Section 3.2, the performances of the presented procedures are evaluated and compared, considering also other two procedures from literature [25]. Successively, the effect of the detection threshold is investigated.

3.2. Procedures' Comparison

As indicated in Table 2, for each procedure and for a fixed threshold, 14 tests were performed with a varying number of sensors, from 1 to 14. In the following comparison, the detection threshold is fixed equal to 0.0001 mg/L. With procedure GR2, using the GR_S algorithm with R as the objective, the maximum R is reached with only six sensors, indicating that additional sensors are not useful for increasing reliability. For the other procedures, the configurations with 8, 12 and 14 sensors are compared, because with a lesser number of sensors, the differences among their performances are negligible.

The presented procedures are also compared with the B_IT and B_DR procedures (Table 1) proposed by [25], considering a multi-objective optimization problem solved using the GA NSGA-II. In these procedures, the multi-objective formulation is expressed by considering more fitness functions, each with a different objective, and the results are expressed in the form of a Pareto front. As explained by the detailed description in [25], the procedure B_IT considers two fitness functions, maximizing and minimizing the objectives JH and TC , respectively. In this case, the nodes with entropy values in the two least-informative quartiles (50%) are filtered prior to the optimization process. The fitness functions of the B_DR procedure is formulated to minimize and maximize D and R , respectively.

Although it is unfair to compare the results of multi- and single-objective approaches, for practical applications, a selection is necessary. To perform this comparison, for a fixed number of sensors for the

procedures involving a multi-objective optimization, one solution has to be selected from the Pareto front. In particular, the solution with the maximum JH value is selected for the B_IT procedure, while for the B_DR procedure, the solutions with maximum R are considered.

Table 1 reports the computational time required for running the test with 14 sensors and a detection threshold of 0.0001 mg/L with the different procedures in an Intel(R) Core(TM) i7-6500U CPU @2.50 GHz processor with 12 GB RAM. Comparing the results of the procedures GR5-B_IT and GR4-B_DR with the same objectives, a drastic reduction of the computational time using GRs is evident. It can be also noted that JH as the objective increases the required time.

A further comparison is performed considering 12 sensors and evaluating the values of the 4 objectives for the solution selected for each procedure. Figure 2a reports the values of JH and TC , while Figure 2b shows the R and D values.

With respect to the JH value (Figure 2a), better performances are observed for the procedures GR3, GR5 and GR6 considering the joint entropy as the objective, each characterized by similar results. As expected, the other procedures based on GRs without JH as the objective (GR1, GR4 and GR6) have a slightly lower JH . Finally, the worst performances are registered for the procedures B_IT and B_DR, considering two fitness functions, which are solved using the GA. It can be also noted that the lower values of TC , always considered in combination with JH , do not correspond to the these values of the procedures among the objectives. This means that JH has a stronger effect in the selection of the optimal solution. Additionally, with respect to R and D (Figure 2b), the performances of the GR procedures are similar, while the B_DR and B_IT procedures have a higher D and a lower R . Among the GR procedures, those considering the detection time (GR1, GR4 and GR6) have very similar performances that are slightly better then the others without D as the objective.

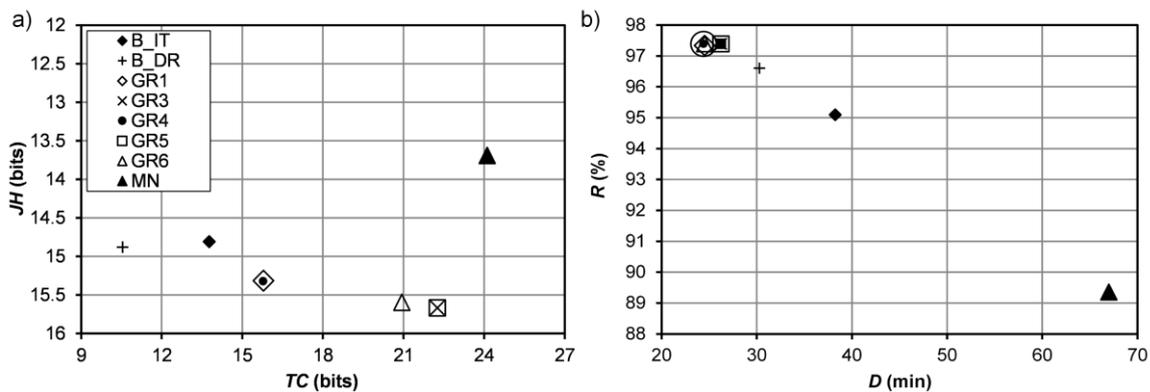


Figure 2. Objective function values obtained with the considered procedures (a) JH and TC objectives; (b) R and D objectives (12 sensors; detection threshold equal to 0.0001 mg/L).

Figure 2a,b also reports the solution corresponding to the existing monitoring network (MN), which is clearly always suboptimal.

A further comparison among the procedures is realized by estimating the overall performance of each approach considering three normalized performance indicators M_1 , M_2 and M_3 , which consider all four objectives. These are estimated as the mean of the parameters W_i (with $i = 1, \dots, 4$) computed for each objective:

$$M_j = \frac{1}{4} \sum_{i=1}^4 W_{j,i} \quad j = 1, \dots, 3 \tag{15}$$

The parameters W_i are evaluated in the three different ways described in the following, and the index $j = 1, \dots, 3$ represents the criterion adopted. The first, used for computing M_1 , is

$$W_{1,i} = \begin{cases} \frac{(O_{i,M} - O_i)}{(O_{i,M} - O_{i,MN})} & \text{if objective } O_i \text{ has to be minimized} \\ \frac{(O_i - O_{i,MN})}{(O_{i,M} - O_{i,MN})} & \text{if objective } O_i \text{ has to be maximized} \end{cases} \quad (16)$$

where $i = 1, \dots, 4$ is the number of considered objectives and $O_{i,M}$ and $O_{i,MN}$ are the maximum and minimum values of the objective among all the selected solutions, respectively.

The second indicator M_2 is calculated considering the following parameter:

$$W_{2,i} = \begin{cases} 1 - \frac{O_i}{O_{i,M}} & \text{if objective } O_i \text{ has to be minimized} \\ \frac{O_i}{O_{i,M}} & \text{if objective } O_i \text{ has to be maximized} \end{cases} \quad (17)$$

The third criterion, used for computing M_3 , considers the following parameter:

$$W_i = \begin{cases} \frac{O_{min}}{O_i} & \text{if objective } O_i \text{ has to be minimized} \\ \frac{O_i}{O_{max}} & \text{if objective } O_i \text{ has to be maximized} \end{cases} \quad (18)$$

O_{max} and O_{min} are the maximum and minimum possible values of the objective i . As mentioned above, for the considered case study, the maximum R value is 97.39%, while the minimum values of D and TC are taken as 5 min and 1 bit, respectively. The tests being realized with 0.0001 mg/L as the detection threshold, the maximum value of JH is 16.71 bits (Table 3).

The indicators M_1 , M_2 and M_3 are in the range $[0, 1]$, and a higher score indicates a better solution. Figure 3 and Table 4 report the values of M_1 , M_2 and M_3 for all the procedures obtained with 8, 12 and 14 sensors.

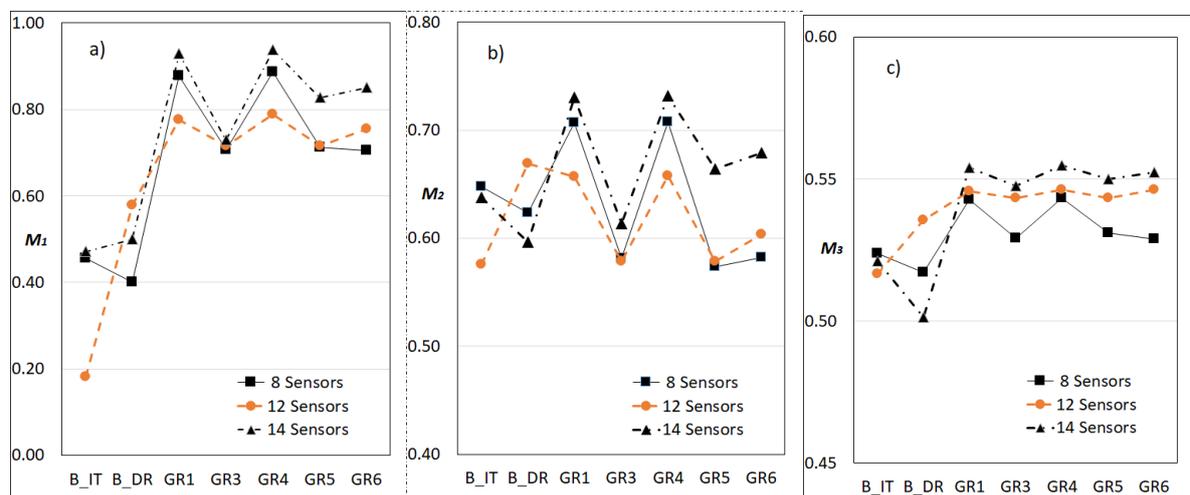


Figure 3. Indicator values of the different procedures with 8, 12 and 14 sensors. (a) M_1 : Equations (15) and (16); (b) M_2 : Equations (15)–(17); (c) M_3 : Equations (15)–(18).

Procedures GR4, with D and R as the objectives, and GR1, with the detection time as a single objective, rank first and second, respectively, in all cases except for that in which the M_2 indicator is estimated with 12 sensors. Procedures GR3, GR5 and GR6 have similar performances with eight sensors, while procedure GR6, which considers all objectives, is third in the list with a higher number of monitoring stations. These results indicate detection time to be the more suitable objective.

Table 4. Performance indicator values.

Procedure	8 Sensors			12 Sensors			14 Sensors		
	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
B_IT	0.4552	0.6476	0.5237	0.1811	0.5757	0.5165	0.4715	0.6380	0.5212
B_DR	0.4000	0.6236	0.5172	0.5795	0.6690	0.5356	0.5000	0.5960	0.5013
GR1	0.8777	0.7067	0.5429	0.7772	0.6566	0.5458	0.9293	0.7303	0.5540
GR3	0.7073	0.5816	0.5292	0.7167	0.5785	0.5433	0.7301	0.6134	0.5476
GR4	0.8867	0.7077	0.5433	0.7883	0.6578	0.5463	0.9378	0.7314	0.5547
GR5	0.7132	0.5732	0.5309	0.7167	0.5785	0.5433	0.8277	0.6640	0.5501
GR6	0.7057	0.5819	0.5289	0.7550	0.6033	0.5462	0.8509	0.6793	0.5523

Comparing procedures GR5, with JH and TC as the objectives, and GR3, with the joint entropy as the single objective, these have the same score considering 12 sensors, while with 14 stations, procedure GR5 has a better performance. The comparisons GR4–GR1 and GR5–GR3 suggest that the adoption of the GR_M algorithm, which incorporates multiple objectives into a single fitness function, improves the solution.

The performance indicators confirm also that the procedures using the GR perform better with respect to the methods B_IT and B_DT, using multi-fitness functions and a GA solver.

It is important to remark that, as for any other heuristic method, GRs have limitations. In fact, when some nodes have the same objective value, they select the node first in the list and the other candidates are not considered. Moreover, the GRs consider the best situation in the current state, and once a sensor is selected, it is fixed during the successive selections. Thus, in this way, only a subset of the search space is investigated.

3.3. Detection Threshold Influence

To investigate the influence of the threshold values on the results obtained with the proposed procedures, the five considered detection thresholds are 0.1, 0.01, 0.001, 0.0001 and 0.00001 mg/L (Table 2).

For procedure GR1, Figure 4a reports the D values as a function of the number of sensors for the different thresholds. As expected, the D values increase when the threshold increases, even if in the range 0.001–0.00001 the differences are small. For analyzing the differences in terms of placement, Figure 4b shows the optimal location obtained with 14 sensors for the detection thresholds of 0.001, 0.0001 and 0.00001 mg/L, showing that 13 out of 14 monitoring locations are in the same site or are very close. This confirms that the variation of the detection limit in the range 0.001–0.00001 mg/L does not have a significant influence on the optimization process. For the threshold values of 0.1 and 0.01 mg/L, not shown herein, slightly different placements are observed.

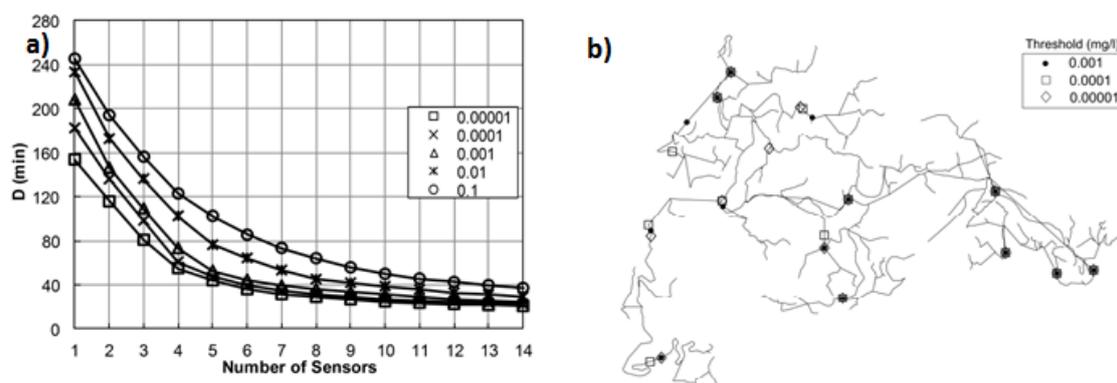


Figure 4. Procedure GR1 for different detection thresholds: (a) D values as a function of the number of sensors; (b) sensor placement.

For procedure GR2, which considers as the objective the reliability (R), the performed tests show that the maximum R is achieved with 11, 8, 6, 6 and 5 sensors for the thresholds of 0.1, 0.01, 0.001, 0.0001 and 0.00001 mg/L, respectively, revealing an effect on the results for values larger than 0.001 mg/L.

The performance of procedure GR3 is estimated by computing the percentage of the system's JH achieved with the selected optimal placement. Considering the joint entropy value of the system reported in Table 3, with 14 sensors, the percentages achieved are 60.69%, 70.08%, 88.56%, 94.84% and 97.50% for the detection thresholds of 0.1, 0.01, 0.001, 0.0001 and 0.00001 mg/L, respectively. Additionally, in this case (results not shown herein), the optimal placement for the detection thresholds of 0.001, 0.0001 and 0.00001 mg/L show 12 out of 14 sensors placed at exactly the same location or very close to each other.

For procedure GR4, Figure 5a reports the D and R values as a function of the number of sensors for the five considered threshold values, while Figure 5b shows the optimal placements of 14 sensors obtained considering thresholds of 0.001, 0.0001 and 0.00001 mg/L. Although there are some differences among the D and R values corresponding to the different detection thresholds, the obtained sensor placement are very similar, as 13 out of 14 sensors are located at exactly the same position or are very close.

Additionally, for procedures GR5 and GR6 (results not reported herein), the optimal placement of 14 sensors for the detection thresholds of 0.001, 0.0001 and 0.00001 mg/L show that 13 out of 14 are placed at exactly the same location or are very close.

Similar results are also obtained with a lesser number of sensors. In conclusion, for all procedures, detectable concentrations lower than 0.001 mg/L do not influence the optimal sensor placement, and small differences are observed for larger values.

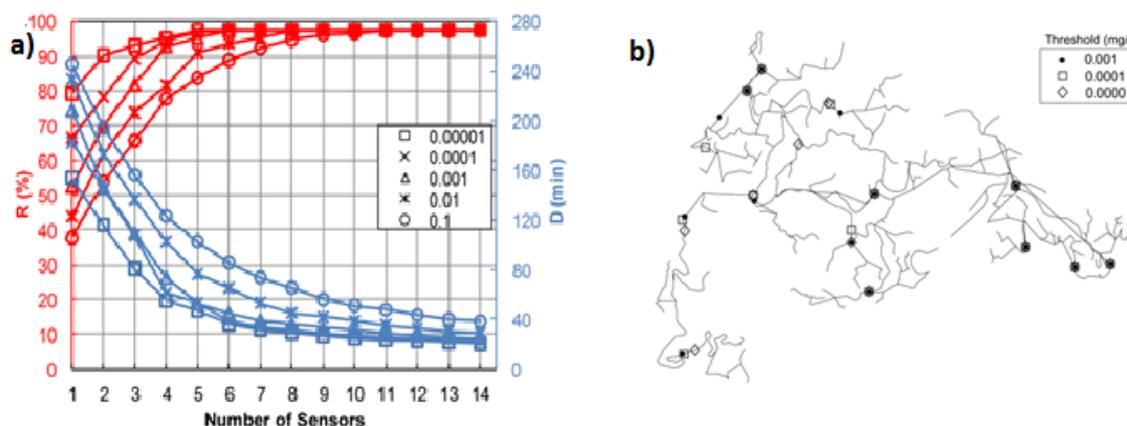


Figure 5. Procedure GR4 for different detection thresholds: (a) D and R values as a function of the number of sensors; (b) sensor placement.

4. Conclusions

GRs are usually simpler and computationally less expensive than other techniques for the solution of optimization problems. In this paper, six different GR-based procedures to evaluate the optimal placement of sensors in a sewer system are proposed. They differ for the adopted design objectives (JH , TC , D and R) and the GR used (GR_S or GR_M). The proposed sensor location procedures are tested on the real case study of the sewer system of Massa Lumbrese, Italy, showing promising results.

Usually, an important parameter to consider in solving the sensor location problem is the minimum concentration (threshold) detectable by a monitoring station. The investigation reveals that detectable concentrations lower than 0.001 mg/L do not influence the optimal sensor placement, and small differences are observed for larger values.

The comparison among the GR procedures indicates that the detection time is the more suitable objective and that the GR_M algorithm, which incorporates multiple objectives into a single fitness function, gives better results.

Greedy approaches use some heuristics to guide the searching process that produces close-to-optimal solutions, but it is not possible to attain the “real” optimal solution because of the size of the search space. However, a relative comparison with respect to some previously developed multi-objective approaches using NSGA-II shows the effectiveness and quality of the GR approaches in the optimal sensor design. The existing monitoring network, realized without applying any methodology, is always suboptimal, showing the importance of the sensor location design.

Acknowledgments: The work described in the present paper was partially realized in the framework of the project SIMonA, financed by the Campania Region (Italy) in the Campus Campania Program. The first author would like to thank the EU for the financial support through the Erasmus Mundus Joint Doctorate Programme ETeCoS3, grant agreement FPA n° 2010-0009. The second author has been partially supported by the EC-FP7 WeSenseIt project, grant agreement 308429.

Author Contributions: The authors contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Falconer, R.A. Global water security: An introduction. *Sci. Parliam.* **2011**, *68*, 34–36.
2. Pouet, M.F.; Thomas, O.; Marcoux, G. Quality survey of wastewater discharges. In *Wastewater Quality Monitoring and Treatment*; Quevauviller, P., Thomas, O., Van der Beken, A., Eds.; J. Wiley & Sons: Chichester, UK, 2006; pp. 275–288.
3. Field, R.; Pitt, R.; Lalor, M.; Brown, M.; Vilkelis, W.; Phackston, E. Investigation of dry-weather pollutant entries into storm-drainage systems. *J. Environ. Eng.* **1994**, *120*, 1044–1066. [[CrossRef](#)]
4. Irvine, K.; Rossi, M.C.; Vermette, S.; Bakert, J.; Kleinfelder, K. Illicit discharge detection and elimination: Low cost options for source identification and track down in stormwater systems. *Urban Water J.* **2011**, *8*, 379–395. [[CrossRef](#)]
5. Bourgeois, W.; Burgess, J.E.; Stuetz, R.M. On-line monitoring of wastewater quality: A review. *J. Chem. Technol. Biotechnol.* **2001**, *76*, 337–348. [[CrossRef](#)]
6. Llopert-Mascaró, A.; Gil, A.; Cros, J.; Alarcón, F. Guidelines for on-line monitoring of wastewater and stormwater quality. In Proceedings of the 11th International Conference on Urban Drainage, Edinburgh, Scotland, UK, 31 August–5 September 2008.
7. Banik, B.K.; Di Cristo, C.; Leopardi, A. A pre-screening procedure for pollution source identification in sewer systems. *Procedia Eng.* **2015**, *119C*, 360–369. [[CrossRef](#)]
8. Banik, B.K.; Di Cristo, C.; Leopardi, A.; de Marinis, G. Illicit intrusion characterization in sewer systems. *Urban Water J.* **2017**, *14*, 416–426. [[CrossRef](#)]
9. Alfonso, L.; He, L.; Lobbrecht, A.; Price, R. Information theory applied to evaluate the discharge monitoring network of the Magdalena River. *J. Hydroinform.* **2013**, *15*, 211–228. [[CrossRef](#)]
10. Ridolfi, E.; Alfonso, L.; Di Baldassarre, G.; Dottori, F.; Russo, F.; Napolitano, F. An entropy approach for the optimization of cross-section spacing for river modelling. *Hydrol. Sci. J.* **2013**, *59*, 126–137. [[CrossRef](#)]
11. Alfonso, L.; Lobbrecht, A.; Price, R. Information theory based approach for location of monitoring water level gauges in polders. *Water Resour. Res.* **2010**, *46*, W03528. [[CrossRef](#)]
12. Aral, M.M.; Guan, J.; Maslia, M.L. Optimal Design of Sensor Placement in Water Distribution Networks. *J. Water Resour. Plan. Manag.* **2010**, *136*, 5–18. [[CrossRef](#)]
13. Moss, M.E.; Tasker, G.D. Intercomparison of hydrological network-design technologies. *Hydrol. Sci. J.* **1991**, *36*, 209–221. [[CrossRef](#)]
14. Davar, Z.K.; Brimley, W.A. Hydrometric network evaluation: Audit approach. *J. Water Resour. Plan. Manag.* **1990**, *116*, 134–146. [[CrossRef](#)]
15. Telci, I.T.; Nam, K.; Guan, J.; Aral, M.M. Optimal water quality monitoring network design for river systems. *J. Environ. Manag.* **2009**, *90*, 2987–2998. [[CrossRef](#)] [[PubMed](#)]

16. Chacon-Hurtado, J.C.; Alfonso, L.; Solomatine, D.P. Rainfall and streamflow sensor network design: A review of applications, classification, and a proposed framework. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 3071–3091. [[CrossRef](#)]
17. Rathi, S.; Gupta, R. Sensor placement methods for contamination detection in water distribution networks: A review. *Procedia Eng.* **2014**, *89*, 181–188. [[CrossRef](#)]
18. Dorini, G.; Jonkergouw, P.; Kapelan, Z.; Savic, D. SLOTS: Effective algorithm for sensor placement in water distribution systems. *J. Water Resour. Plan. Manag.* **2010**, *136*, 620–628. [[CrossRef](#)]
19. Rathi, S.; Gupta, R. A simple sensor placement approach for regular monitoring and contamination detection in in water distribution networks. *KSCE J. Civ. Eng.* **2016**, *20*, 597–608. [[CrossRef](#)]
20. Preis, A.; Ostfeld, A. Multiobjective contaminant sensor network design for water distribution systems. *J. Water Resour. Plan. Manag.* **2008**, *134*, 366–377. [[CrossRef](#)]
21. Weickgenannt, M.; Kapelan, Z.; Blokker, M.; Savic, D.A. Risk based sensor placement for contaminant detection in water distribution systems. *J. Water Resour. Plan. Manag.* **2010**, *136*, 629–636. [[CrossRef](#)]
22. Shen, H.; McBean, E. Pareto optimality for sensor placements in a water distribution system. *J. Water Resour. Plan. Manag.* **2011**, *137*, 243–248. [[CrossRef](#)]
23. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
24. Banik, B.K.; Alfonso, L.; Torres, A.S.; Mynett, A.; Di Cristo, C.; Leopardi, A. Optimal placement of water quality monitoring stations in sewer systems: An information theory approach. *Procedia Eng.* **2015**, *119*, 1308–1317. [[CrossRef](#)]
25. Banik, B.K.; Alfonso, L.; Di Cristo, C.; Leopardi, A.L.; Mynett, A. Evaluation of different formulations to optimally locate pollution sensors in sewer systems. *J. Water Resour. Plan. Manag.* **2017**, *143*. Available online: <http://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0000778> (accessed on 31 March 2017). [[CrossRef](#)]
26. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
27. Ostfeld, A.; Salomons, E. Optimal layout of early warning detection stations for water distribution systems security. *J. Water Resour. Plan. Manag.* **2004**, *130*, 377–385. [[CrossRef](#)]
28. International Electrotechnical Commission. *International Electrotechnical Commission*; IEC 80000–13:2008; IEC: Geneva, Switzerland, 2013.
29. Markus, M.; Vernon, K.H.; Tasker, G.D. Entropy and generalized least square methods in assessment of the regional value of stream gages. *J. Hydrol.* **2003**, *283*, 107–121. [[CrossRef](#)]
30. Alfonso, L.; Lobbrecht, A.; Price, R. Optimization of water level monitoring network in polder systems using information theory. *Water Resour. Res.* **2010**, *46*, W12553. [[CrossRef](#)]
31. McGill, W.J. Multivariate information transmission. *Psychometrika* **1954**, *19*, 97–116. [[CrossRef](#)]
32. Watanabe, S. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* **1960**, *4*, 66–82. [[CrossRef](#)]
33. Greco, S.; Zaniolo, C. Greedy algorithms in Datalog. *Theor. Pract. Log. Prog.* **2001**, *1*, 381–407. [[CrossRef](#)]
34. Tallam, S.; Gupta, N. A concept analysis inspired greedy algorithm for test suite minimization. In *CM SIGSOFT Software Engineering Notes*; ACM: New York, NY, USA, 2005; Volume 31, pp. 35–42.
35. Kumar, R.; Moseley, B.; Vassilvitskii, S.; Vattani, A. Fast greedy algorithms in mapreduce and streaming. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, Montreal, QC, Canada, 23–25 July 2013; pp. 1–10.
36. Rieckermann, J.L.; Neumann, M.; Ort, C.; Huisman, J.L.; Gujer, W. Dispersion coefficients of sewers from tracer experiments. *Water Sci. Technol.* **2005**, *52*, 123–133. [[PubMed](#)]
37. Banik, B.K.; Di Cristo, C.; Leopardi, A. SWMM5 toolkit development for pollution source identification in sewer systems. *Procedia Eng.* **2014**, *89*, 750–757. [[CrossRef](#)]
38. Cozzolino, L.; Della Morte, R.; Palumbo, A.; Pianese, D. Stochastic approaches for sensors placement against intentional contaminations in water distribution systems. *Civ. Eng. Environ. Syst.* **2011**, *28*, 75–98. [[CrossRef](#)]

