Article

Pressure Sensor Placement for Leak Localization in Water Distribution Networks using Information Theory

Ildeberto Santos-Ruiz [1], Francisco-Ronay López-Estrada [1,*], Vicenç Puig [2], Guillermo Valencia-Palomo [3], and Héctor-Ricardo Hernández [1]

[1] Tecnológico Nacional de México, I.T. Tuxtla Gutiérrez, TURIX-Dynamics Diagnosis and Control Group. Carretera Panamericana km 1080 S/N, 29050, Tuxtla Gutierrez, Mexico; ildeberto.dr@tuxtla.tecnm.mx (I.S-R.); frlopez@tuxtla.tecnm.mx (F-R.L-E.); hector.hl@tuxtla.tecnm.mx (H-R.H.)

[2] Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Universitat Politècnica de Catalunya, C/. Llorens i Artigas 4-6, 08028, Barcelona, Spain; vicenc.puig@upc.edu (V.P.)

[3] Tecnológico Nacional de México, I.T. Hermosillo. Av. Tecnológico y Periférico Poniente S/N, 83170, Hermosillo, Mexico; gvalencia@hermosillo.tecnm.mx (G.V-P.)

* Correspondence: frlopez@tuxtla.tecnm.mx; Tel.: +52 961 615 0461

Abstract: This paper presents a method for optimal pressure sensor placement in water distribution networks using information theory. The criteria for selecting the network nodes where to place the pressure sensors is that they provide the most useful information for locating leaks in the network. Considering that the node pressures measured by the sensors can be correlated (mutual information), a subset of sensor nodes in the network is chosen. The relevance of information is maximized, and information redundancy is minimized simultaneously. The selection of the nodes where to place the sensors is performed from datasets of pressure changes caused by multiple leak scenarios, which are synthetically generated by simulation using the EPANET software. In order to select the optimal subset of nodes, the candidate nodes are ranked using a heuristic algorithm with quadratic computational cost, which makes it time-efficient compared to other sensor placement algorithms. The sensor placement algorithm was implemented in MATLAB and tested in the Hanoi network. It was verified by exhaustive analysis that the selected nodes are the best combination to place the sensors and detect leaks.

Keywords: Sensor Placement; Pressure Monitoring; Information Theory; Leak Localization; Water Distribution Network.

Citation: Santos-Ruiz, I.; López-Estrada, F.R.; Puig, V.; Valencia-Palomo, G.; Hernández, H.R. Pressure Sensor Placement in Water Distribution Networks Using Information Theory. *Sensors* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Finding a suitable sensor placement is a fundamental problem for monitoring Water Distribution Networks (WDNs) because it is impossible to install sensors at each point of the geographic area covered by the distribution system. A WDN comprises hundreds of nodes; however, only a few sensors can be installed in certain carefully selected nodes. Then, the main question is how to select the optimal sensor placement. Finding an answer to this problem is not trivial because the selected nodes must capture the most relevant information to estimate hydraulic variables at non-measured points and provide essential information for different supervision algorithms, e.g., for leak localization [1,2]. Often there are pressure and flow instruments at the supplying nodes of a WDN and in some cases at critical points (e.g., at the minimum pressure node). However, these measurements are not sufficient for an accurate leak localization, so additional sensors must be installed at other sites [3]. A practical solution is to install more pressure sensors because they are cheaper and easier to install and maintain than flow sensors. In addition, node-pressures are more sensitive to leaks than flow rates, which is why many localization algorithms are based primarily on pressure measurements. The problem of sensor placement is closely related to other WDN management problems, such as the state estimation of the network [4–6], the model calibration [7,8], the water quality monitoring such as detection of contaminants and cyberattacks [9–15], among others. Nevertheless, the present work focuses on the context of leak detection and localization as discussed in [16,17]. Regarding techniques for optimal sensor placement for leak/burst detection and localization in water distribution systems, a comprehensive review can be found at [18].

In a mathematical/computational context, the placement of pressure sensors is a mixed integer programming problem. In this problem, for a network with N nodes, a sensor placement consists of a selection $[s_1, s_2, \ldots, s_N]$ where s_i are binary decision variables such that $s_i = 1$ indicates that a sensor will be placed on the i-th node, whereas $s_i = 0$ indicates that no sensor will be placed on that node.

Combinatorial analysis shows that there are $2^N - 1$ possible sensor placements when non-empty subsets with any number of sensors are considered. If the number of sensors is previously set to a fixed number S, then the number of possible sensor placements is reduced to $\binom{N}{S}$, which is still a very large number. Therefore, in medium-sized and large networks is not feasible to check all possible combinations. For example, in a network containing 500 nodes the number of different placements for 10 sensors is $\binom{500}{10} \approx 2.5 \times 10^{20}$. That is why it is important to find an optimal placement method without analysing all the possible combinations.

Usually, sensor placement focused on leak localization has been addressed under an optimization approach from synthetic pressure data obtained by simulation. Some authors have focused on minimizing the number of undetectable leaks [19,20], whereas others reduce the error in the leak location [21,22]. In [23], a min-max optimization algorithm that considers the isolation of the leaks from their signatures obtained through simulation is proposed. In [24], a multi-objective approach to mitigate errors both in detection and localization of leaks, considering minimum night-flow conditions is presented. Regarding the optimization of the objective function, two approaches are usually used: deterministic methods (e.g., branch and bound [25]), and metaheuristic methods, (e.g., genetic algorithms [26–28] and particle swarm optimization [29]). Deterministic approaches guarantee an optimal solution, but the computation time increases exponentially with the number of nodes and possible leak scenarios. On the other hand, metaheuristic methods search for a near-optimal solution that only guarantees optimal when the number of candidate solutions evaluated (named "population size") tends to infinity. Furthermore, optimization-based sensor placement methods are linked to a specific leak localization method because the objective function is expressed in terms of a localization error or isolation index for that method [16,29,30]. Based on this, a sensor placement method may be optimal for one specific leak localization method but not as good for others. Furthermore, the method should be independent of the leak localization method since it is not feasible to change it for every method. Thus, an improved leak localization method could be proposed based on an ensemble of different machine learning algorithms using the information provided by the sensors.

The huge computing time in networks with hundreds and thousands of nodes using optimization-based methods and the high dependence on the selected leak localization method has motivated the present work. In this new proposal, it is not considered how specific leak localization methods will use the information provided by the sensors, but rather that the sensor placement method only focuses on the sensors capturing as much information related to the leaks as possible. The proposed method consists of a heuristic algorithm to select the subset of nodes where to place the sensors, seeking to maximize the relevance of the information captured by the sensors while minimizing the redundancy between the pressures in the selected nodes. Both metrics, relevance and redundancy, will be defined in terms of information theory.

An important contribution of this work is the reduction in computing time for sensor placement, compared to methods based on metaheuristic optimization. Another relevant contribution is the non-dependence of the sensor placement on the leak localization method used, which allows using the same sensor placement with different localization methods. Some aspects not yet covered in this work are the possible heterogeneity of sensors (e.g., different errors and measurement ranges) and the influence of the measurement noise in the optimal placement, but they are considered as future work.

The rest of the document is organized as follows: In Section 2 the concepts of redundancy and relevance will be presented in terms of mutual information, and the information quotient used as the basis of the method will also be defined. In Section 3, the proposed method is formally described and some guidelines for its implementation are given. In Section 4, the results of the proposed method applied in a simplified version of the Hanoi network (case study) are presented. Finally, in Section 5, the conclusions are presented and future related works are proposed.

2. Information Theory Fundamentals

In Shannon's Information Theory (IT), the self-information of a random variable is defined according to the unexpectedness of its values [31]. Thus, the information contained in a constant random variable is zero. Mathematically, if an event E has probability P, its information content is defined by:

$$I(E) \stackrel{\text{def}}{=} -\log_b(P), \tag{1}$$

where the unit of measure of I is defined by the base of the logarithm, b, which in case of being $b = 2$ is called "bit". In a discrete random variable X with probability function $p(x) = \Pr(X = x)$, the self-information of obtaining x as a result when measuring X is given by:

$$I(x) = -\log_b(p(x)) = \log_b(1/p(x)). \tag{2}$$

To quantify the average information that a random variable contains, considering all its possible values, *entropy* is used:

$$H(X) \stackrel{\text{def}}{=} E(I(x)) = \sum_x -p(x)\log_b(p(x)), \tag{3}$$

which is the expected value of the information contained in the measurements of X. That is the sum of the self-information of each of its possible values weighted by its probability of occurrence.

The *mutual information* of two random variables, sometimes called "information gain", measures the amount of information obtained about one of the random variables by observing the other one. For example, in a practical application of WDN monitoring, the mutual information between two node pressures would indicate how much information about the pressure at one node is gained by knowing the pressure at the other one. In probabilistic terms, mutual information determines how different the joint distribution of (X, Y) is from the product of the marginal distributions of X and Y.

For two discrete variables X and Y, defined over the space $\mathcal{X} \times \mathcal{Y}$, the mutual information is computed as the double sum:

$$I(X,Y) = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{4}$$

where $p(x,y) = \Pr(X = x, Y = y)$ is the joint probability function of X and Y, whereas $p(x)$ and $p(y)$ are the marginal probability functions of X and Y, respectively. The mutual information (4) is derived from entropy and conditional probability by the following equivalences:

$$I(X,Y) \equiv H(X) - H(X \mid Y) \equiv H(Y) - H(Y \mid X). \tag{5}$$

Furthermore, $I(X,X) = H(X)$, $I(X,Y) = I(Y,X)$ and $I(X,Y) \geq 0$, where $I(X,Y) = 0$ iff X and Y are independent.

For continuous random variables, the summations in (4) are replaced by integrals and the probability functions by probability densities:

$$I(X,Y) = \int_{\mathcal{Y}} \int_{\mathcal{X}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \, \mathrm{d}x \, \mathrm{d}y. \tag{6}$$

Due to the difficulty in modeling the probability densities and subsequently evaluating the double integrals in (6), a simplification to calculate the mutual information in continuous variables is to discretize the variables with n bits, so that the domain of each variable is reduced to 2^n bins. For example, to compute the mutual information of two node pressures in a hydraulic network, the span of the pressure variables $[P_{\min}, P_{\max}]$ must be divided into a discrete 8-bit grid (256 different values) and then apply (4).

3. Sensor Placement Method

The proposed sensor placement method is based on a dataset of node pressures that collects typical variations due to leaks of different sizes in all network nodes. The pressure dataset is obtained from simulations with the hydraulic model of the network [32]. Each pressure data is labeled with a "leak class" (the node where the leak occurs) so that the proposed method can be classified as supervised.

In the context of machine learning, the placement of pressure sensors is a *feature selection* stage. To select the features (subset of nodes where the sensors will be placed), an algorithm is proposed that seeks to maximize the relevance of the selected features (node pressures) for the response variable (leaky node), while each of them avoids capturing information already contributed by the others, that is, minimizing redundancy.

The following definitions of relevance and redundancy, proposed in [33], are used as a basis for defining the methodology:

Definition 1 (Relevance)**.** A metric of the relevance of the subset of node pressures $\mathcal{S}$ for the response variable y (leak node), is given by

$$\mathrm{Rel}(\mathcal{S}) \stackrel{\text{def}}{=} \frac{1}{S} \sum_{x \in \mathcal{S}} I(x,y), \tag{7}$$

where x is any feature in $\mathcal{S}$, and $S = |\mathcal{S}|$ is the number of features in $\mathcal{S}$ (the cardinality).

Definition 2 (Redundancy)**.** A metric for information redundancy in a feature subset $\mathcal{S}$ is given by:

$$\mathrm{Red}(\mathcal{S}) \stackrel{\text{def}}{=} \frac{1}{S^2} \sum_{x,x' \in \mathcal{S}} I(x,x'), \tag{8}$$

where x and x' are any features in $\mathcal{S}$.

To apply the above definitions to compute a pressure sensor placement, first, a dataset of node pressures is built covering different scenarios that consider leaks of different magnitude in all nodes of the network. Through simulation with the hydraulic model of the network, a series of samples of the node pressures is obtained, one sample for each different leakage scenario. In this way, if M different leakage scenarios are simulated in a network containing N nodes, the result of the simulation is a collection of N M-dimensional vectors, x and x' in (7) and (8), corresponding to the N candidate nodes (initially, it is assumed that all nodes are potential sensing nodes). In addition, an output vector, y in (7), is generated containing integer labels to indicate the leaky node corresponding to each simulated scenario.

The exhaustive search for the optimal subset of sensors, $\mathcal{S}$, requires testing of the $2^N - 1$ different combinations, which would require impractical computation time in networks with many nodes. Therefore, the use of a method proposed in [33] is considered to rank the node pressures through an iterative forward scheme that only requires $O(NS)$

computations. In fact, with this proposal it is possible to rank all the node pressures in order of importance with a computational cost $O(N^2)$.

Next, a heuristic algorithm is proposed that orders the node pressures according to their importance to explain the different leak classes (leaky nodes). The first node pressures in the output list correspond to the nodes with the highest importance in explaining the leak positions according to the information contained in the dataset. The sequential selection of nodes starts from an empty subset and, in each iteration, adds the best-ranked node among those that are still available to be selected. In each iteration, the relevance of each available feature (node pressure) with respect to the output (leaky node) and its redundancy with respect to the variables that have been previously selected is evaluated using the following equations, adapted from (7) and (8):

$$\mathrm{Rel}_y(x) = I(x, y), \tag{9}$$

$$\mathrm{Red}_{\mathcal{S}}(x) = \frac{1}{S} \sum_{x' \in \mathcal{S}} I(x, x'). \tag{10}$$

Since maximizing relevance and simultaneously minimizing redundancy represents a multi-objective problem, a combined relevance/redundancy index (RRI) is defined that increases with increasing relevance and also with decreasing redundancy, so the problem is expressed as a single objective to be maximized:

$$\mathrm{RRI} = \mathrm{Rel}_y(x) / \mathrm{Red}_{\mathcal{S}}(x). \tag{11}$$

The complete node ranking process is formally expressed in Algorithm 1. When the process finishes, the nodes to place the sensors are taken from the first positions in the list $\mathcal{S}$. If it is not necessary to obtain the complete ranking of nodes, but only to know the best-ranked positions, the process may stop prematurely when the subset $\mathcal{S}$ already contains the number of sensors to be placed.

Algorithm 1: Node ranking based on information theory

Data: Set with all node pressures, $\mathcal{A}$. The nodes in $\mathcal{A}$ will be placed in the ordered list $\mathcal{S}$ according to their importance (relevance/redundancy). During the process, $\widetilde{\mathcal{S}}$ denotes the elements of $\mathcal{A}$ not yet added in $\mathcal{S}$.
Result: Set with ordered node pressures, $\mathcal{S}$.
Initialization:
  $\mathcal{S} \leftarrow \left[\arg\max_{x \in \mathcal{A}} \mathrm{Rel}_y(x) \right]$
repeat
  if $\exists x \in \widetilde{\mathcal{S}}, \mathrm{Rel}_y(x) \neq 0, \mathrm{Red}_{\mathcal{S}}(x) = 0$ **then**
    $\mathcal{S} \leftarrow \left[\mathcal{S}, \arg\max_{x \in \widetilde{\mathcal{S}}, \mathrm{Red}_{\mathcal{S}}(x) = 0} \mathrm{Rel}_y(x) \right]$
  else
    break
  end
until $\forall x \in \widetilde{\mathcal{S}}, \mathrm{Red}_{\mathcal{S}}(x) \neq 0$
repeat
  $\mathcal{S} \leftarrow \left[\mathcal{S}, \arg\max_{x \in \widetilde{\mathcal{S}}, \mathrm{Rel}_y(x) \neq 0} \mathrm{Rel}_y(x) / \mathrm{Red}_{\mathcal{S}}(x) \right]$
until $\forall x \in \widetilde{\mathcal{S}}, \mathrm{Rel}_y(x) = 0$
$\mathcal{S} \leftarrow \left[\mathcal{S}, \widetilde{\mathcal{S}} \right]$

Regarding the number of sensors to place for leak localization purposes, this is determined by the equipment available in most cases. The minimum number of sensors for a successful leak localization method will depend on how that method uses the

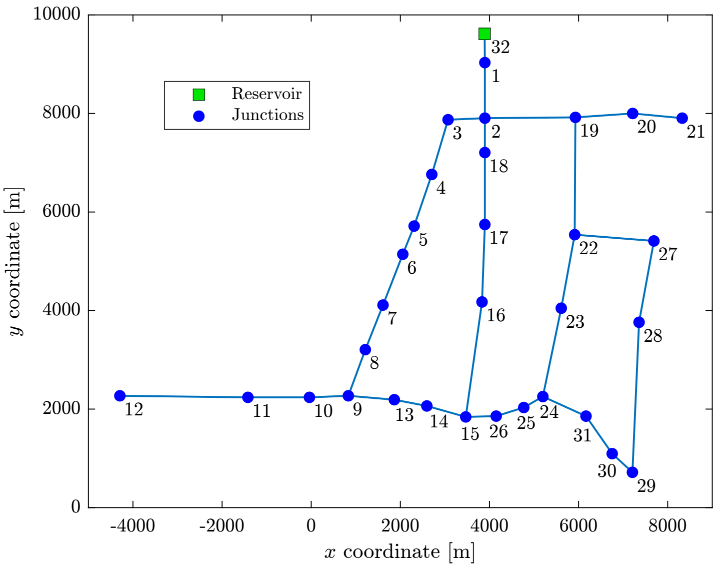

Figure 1. The Hanoi network.

available information, the measurement noise, as well as the quality, resolution and calibration of the sensors. If there are enough resources to intensively instrument the network, it must be taken into account that increasing the number of sensors does not always lead to better performance in locating leaks. To determine up to how many sensors should be placed, it is suggested to start from the ranking obtained by Algorithm 1, and run a marginal analysis with the leak localization method to be used. Starting from one sensor (the best ranked), the number of sensors is progressively increased and the leak localization performance is evaluated for each new set of sensors until adding a new sensor no longer represents a significant benefit in locating leaks.

It should be noted that Algorithm 1 does not take into account the geographical distribution of the nodes, since relevance and redundancy depend only on the mutual information between node pressures. This means that the distance between sensors is not determining, the network topology is what determines the amount of mutual information (i.e., two sensors can be geographically very close but have little mutual information).

4. Results and Discussion

Algorithm 1 was implemented in MATLAB and tested on the Hanoi network [34]. The model of the Hanoi network is composed of one reservoir, 31 consumer nodes, and 34 pipes, as shown in Figure 1. Due to its reduced topology, this network has been used as a standarized benchmark in different works [28,35,36].

In order to build the pressure dataset, leaks of different magnitude were simulated in each junction node using the EPANET 2 simulation program [37] through EPANET/MATLAB Toolkit [38]. The procedure to generate the dataset using EPANET, the training and the predictive use of classifiers in locating leaks have been described in [39]. The dataset generated by simulation for this work considers leaks at all junction nodes with flow rates from 1 l/s to 50 l/s. In order to simulate leaks in a node, the demand assigned to that node in the EPANET hydraulic model was modified increasing this demand by an amount equal to the flow of the simulated leak. Because the Hanoi network contains few nodes, the optimality of the sensor placement calculated by Algorithm 1 was exhaustively verified.

To assess the optimality of the sensor placement obtained from Algorithm 1, leak localization tests were carried out using two machine learning methods that use the pressures in the selected nodes as features (input variables). The methods used were k-Nearest Neighbors (k-NN) and Quadratic Discriminant Analysis (QDA). These leak lo-

Table 1: Better positions to place three sensors in the Hanoi network, obtained by exhaustive analysis. The shaded selection is the one obtained by Algorithm 1.

Rank	Nodes	Location Method	
		k-NN	QDA
1	{12,21,28}	0.9974	0.9948
1	{12,21,27}	0.9974	0.9948
1	{12,21,31}	0.9974	0.9948
2	{7,12,21}	0.9961	0.9936
2	{12,17,21}	0.9961	0.9936
3	{3,12,21}	0.9961	0.9923
3	{4,12,21}	0.9961	0.9923
3	{6,12,21}	0.9961	0.9923
3	{5,12,21}	0.9961	0.9923

(a) Metric: Classification accuracy

Rank	Nodes	Location Method	
		k-NN	QDA
1	{12,21,28}	0.0026	0.0052
1	{12,21,27}	0.0026	0.0052
1	{12,21,31}	0.0026	0.0052
2	{12,13,21}	0.0065	0.0065
3	{7,12,21}	0.0065	0.0090
3	{12,17,21}	0.0065	0.0090
4	{3,12,21}	0.0039	0.0129
4	{4,12,21}	0.0039	0.0129
5	{6,12,21}	0.0065	0.0129

(b) Metric: Average topological distance

calization methods are based on classifiers that recognize directional patterns in pressure residuals using supervised learning techniques, as described in [40].

Through the marginal analysis, suggested at the end of Section 3, it was determined that $S = 3$ is an adequate number of sensors in the Hanoi network, because the addition of the fourth sensor does not produce a statistically significant improvement (with 0.95 confidence level) in leak location. Considering measurement noise may possibly increase the minimum number of sensors, but this discussion has been considered as future work. Because the Hanoi network contains few nodes, it was possible to comprehensively analyze all 4 495 possible combinations of three sensor nodes. For each triplet of nodes (3-sensor placement), 50 leak localization tests were carried out with flow-rates $q_{\text{leak}} = 1, 2, \ldots, 50\,\text{l/s}$ in each node of the network. Finally, the overall performance of both methods was evaluated for each candidate triplet using the classification accuracy (Acc) and the average topological distance (ATD) as performance metrics, as defined in [41]. The Acc is the fraction of exactly located leaks considering all leak scenarios in the test dataset, where Acc $= 1$ means that all leaks were correctly located, whereas Acc $= 0$ means that no leaks were correctly located. The ATD is a measure of how far from the true leaky node the classifier locates the leak, counting the number of separation links between the true leaky node and the estimated leaky node, averaged across all scenarios in the test dataset. So the best sensor placements are the ones that lead to the highest Acc values and the lowest ATD values.

The results in Table 1 show that the node triplet $\{12, 21, 28\}$ computed by Algorithm 1 is among the best ranked, since they present the highest accuracy and the lowest average topological distance.

Figure 2 shows the geographic location of the 3-sensor placement obtained considering the three nodes best ranked by Algorithm 1. Figure 3 shows the complete ranking considering the 31 nodes of the network.

Table 2 shows the sensor placements obtained for 2, 3 and 4 sensors in the Hanoi network, and they are compared with the results obtained by metaheuristic methods reported in the literature [29]. The nodes selected by these methods are quite similar and produce very close results in terms of accuracy in locating leaks based on the pressures of the selected nodes. However, there is an important difference in the computation time of the IT-based method (Algorithm 1) compared to the metaheuristic methods. On a personal computer with an Intel 64-bit processor and 8 GB in RAM, the computation time for the IT-based method was around one second with the synthetic data from the Hanoi network, whereas it was 24 minutes for the genetic algorithm (may be larger, depending on the initial population size) and about one hour for exhaustive analysis.

Further tests were made on larger networks, e.g. in some mid-size sectors of the Madrid network. Figure 4 shows a 10-sensor placement obtained using Algorithm

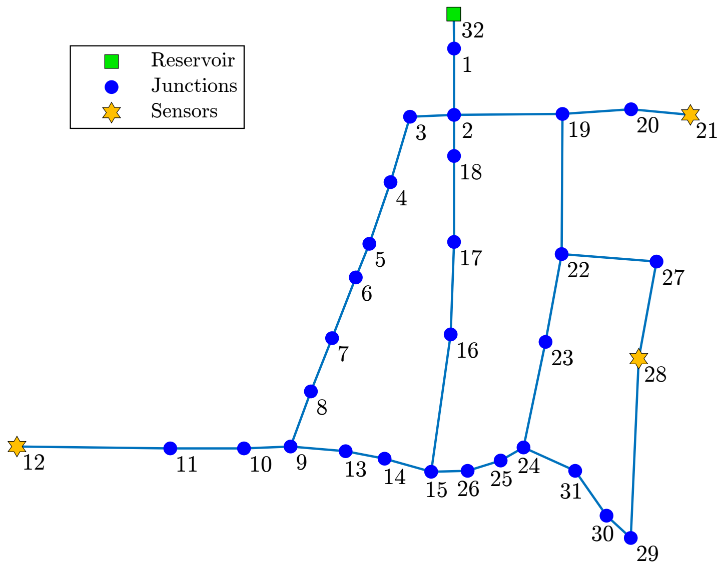

Figure 2. Computed 3-sensor placement in the Hanoi network.

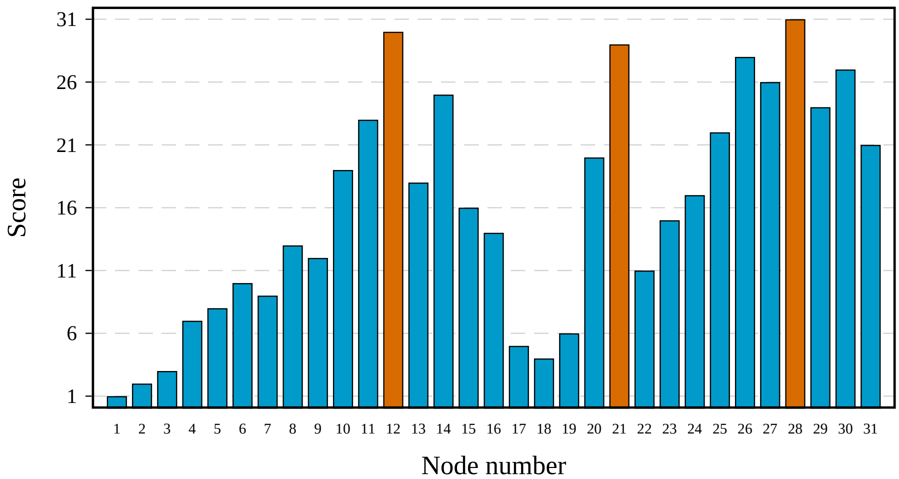

Figure 3. Node ranking in the Hanoi network.

Table 2: Optimal 3-sensor placement in the Hanoi network using different methods.

S	**IT**[a]	**GA**[b]	**PSO**[c]	**SE**[d]
2	$\{12, 28\}$	$\{12, 21\}$	$\{12, 21\}$	$\{12, 21\}$
3	$\{12, 21, 28\}$	$\{12, 21, 27\}$	$\{12, 14, 21\}$	$\{12, 21, 29\}$
4	$\{12, 21, 26, 28\}$	$\{1, 12, 21, 29\}$	$\{1, 12, 21, 24\}$	$\{1, 12, 21, 29\}$

[a] Algorithm 1.
[b] Genetic Algorithm, reported in [29].
[c] Particle Swarm Optimization, reported in [29].
[d] Semi-Exhaustive search, reported in [29].

1 in a sector of the Madrid network containing one reservoir, 312 junction nodes and around 14 km of pipes. In these case, optimality was not exhaustively tested due to the vast number of possible placements to compare. However, it was found that the average accuracy in leak localization with sensor placements obtained by Algorithm 1 was at least better than that obtained with an existing placement (previously obtained by genetic algorithm) for different leak scenarios.

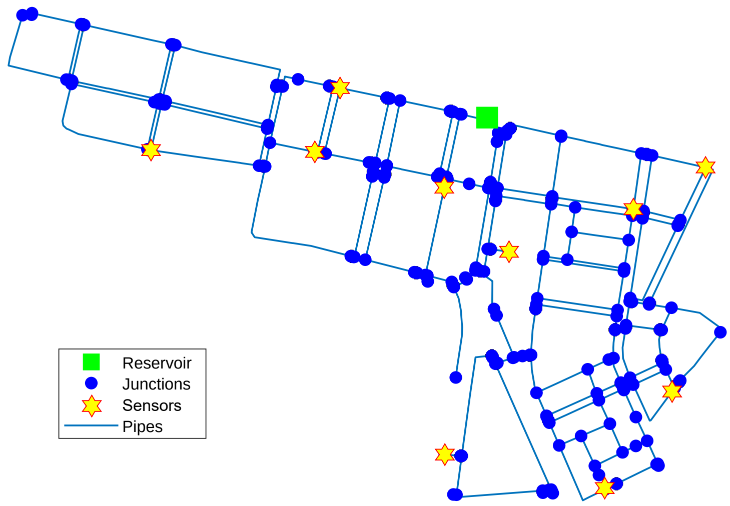

Figure 4. The optimal 10-sensor placement in a sector of the Madrid network.

Figures 2 and 4 show that the computed sensor placements do not show geometric regularity (i.e., the sensors do not appear equally spaced), since geometric or spatial criteria are not used to distribute the sensors in the network. However, regardless of geometric irregularity, leak location tests with these placements demonstrate that pressure measurements at these nodes provide the most useful information for discerning between different leak scenarios. In fact, when the placement of sensors obtained by Algorithm 1 is compared with the results reported by other authors using metaheuristics, sometimes very close performances can be found even though the sensors are distributed in different nodes, because the proposed algorithm does not optimize the position of each sensor individually but the entire set of sensors. This can be explained with an informal analogy: two soccer teams can achieve similar performances using different players.

Although, as noted above, there may be different sensor placements that lead to a good performance in locating leaks, the one obtained by Algorithm 1 has the advantage of being calculated in less time than the methods based on metaheuristics and that it is not linked to a specific leak location method, so changing the leak location method does not imply changing the location of the sensors, which would be impractical.

5. Conclusion

This paper has presented a technique for finding optimal sensor placements from information theory using a sequential forward selection, maximizing the relevance and minimizing the redundancy of the selected node subset. The proposed technique is computationally less expensive than other methods reported in the literature because the proposed technique operates directly on the values of node pressures without performing calculations for leak localization in the implementation of the algorithm. The optimality of the sensor placement obtained with the proposed method was extensively tested by simulation with the Hanoi network. It was found that the selection of nodes to place sensors using information theory produces the best combination of pressure variables to locate leaks using different machine learning methods.

An implicit assumption in the proposed algorithms is that all network nodes have the same availability to place the sensors. However, in practice some specific nodes may have placement priority over others; for example, critical nodes (points of minimum pressure) and nodes that supply essential services (e.g., hospitals) could be monitored as a priority. It may also occur that some nodes already have a sensor installed and that previous partial placement must be held, or that the conditions in a node are physically adverse and instrumentation is avoided. These circumstances warrant adjustments to the proposed sensor placement algorithm that may lead to future work. Another possible working line is the combination of heterogeneous sensors where different sensing specifications are included (e.g., different precision) or where the sensors measure different physical magnitudes (e.g., sensor placements combining pressure and flow sensors).

Author Contributions: Conceptualization, I.S-R., F-R.L-E. and V.P.; methodology, I.S-R.; software, I.S-R.; validation, F-R.L-E., V.P. and G.V-P.; formal analysis, G.V-P. and H-R.H.; data curation, I.S-R. and H-R.H.; writing–original draft preparation, I.S-R.; writing–review and editing, F-R.L-E., V.P. and G.V-P.; supervision, F-R.L-E. and V.P.; project administration, F-R.L-E. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been co-financed by the European Regional Development Fund of the European Union in the framework of the ERDF Operational Program of Catalonia 2014-2020, under the research project 001-P-001643 Agrupació Looming Factory. Tecnológico Nacional de México (TecNM) has also co-financed this work by granting the research project 11080.21-P.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Sela, L.; Amin, S. Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization. *Advanced Engineering Informatics* **2018**, *36*, 55–63. doi:https://doi.org/10.1016/j.aei.2018.02.004.
2. Puig, V.; Ocampo-Martínez, C.; Pérez, R.; Cembrano, G.; Quevedo, J.; Escobet, T., Eds. *Real-time Monitoring and Operational Control of Drinking-Water Systems*; Springer International Publishing, 2017. doi:10.1007/978-3-319-50751-4.
3. Pérez, R.; Sanz, G.; Puig, V.; Quevedo, J.; Cugueró, M.À.; Nejjari, F.; Meseguer, J.; Cembrano, G.; Mirats, J.M.; Sarrate, R. Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in Barcelona [applications of control]. *IEEE Control Systems Magazine* **2014**, *34*, 24–36.
4. Kang, D.; Lansey, K. Optimal Meter Placement for Water Distribution System State Estimation. *Journal of Water Resources Planning and Management* **2010**, *136*, 337–347. doi:10.1061/(ASCE)WR.1943-5452.0000037.
5. Santos-Ruiz, I.; López-Estrada, F.R.; Puig, V.; Blesa, J. Estimation of Node Pressures in Water Distribution Networks by Gaussian Process Regression. 4th Conference on Control and Fault Tolerant Systems (SysTol'19). IEEE, 2019, pp. 50–55. doi:10.1109/SYSTOL.2019.8864793.
6. Mankad, J.; Natarajan, B.; Srinivasan, B. Integrated approach for optimal sensor placement and state estimation: A case study on water distribution networks. *ISA Transactions* **2021**. doi:https://doi.org/10.1016/j.isatra.2021.06.004.
7. Simone, A.; Giustolisi, O.; Laucelli, D.B. A proposal of optimal sampling design using a modularity strategy. *Water Resources Research* **2016**, *52*, 6171–6185. doi:10.1002/2016WR018944.
8. Wéber, R.; Hős, C. Efficient Technique for Pipe Roughness Calibration and Sensor Placement for Water Distribution Systems. *Journal of Water Resources Planning and Management* **2020**, *146*, 04019070. doi:10.1061/(ASCE)WR.1943-5452.0001150.
9. Kessler, A.; Ostfeld, A.; Sinai, G. Detecting Accidental Contaminations in Municipal Water Networks. *Journal of Water Resources Planning and Management* **1998**, *124*, 192–198. doi:10.1061/(ASCE)0733-9496(1998)124:4(192).
10. Eliades, D.G.; Polycarpou, M.M.; Charalambous, B. A Security-Oriented Manual Quality Sampling Methodology for Water Systems. *Water Resources Management* **2011**, *25*, 1219–1228.
11. Vaidya, U.; Fardad, M. On optimal sensor placement for mitigation of vulnerabilities to cyber attacks in large-scale networks. 2013 European Control Conference (ECC). IEEE, 2013, pp. 3548–3553. doi:10.23919/ECC.2013.6669723.
12. Rathi, S.; Gupta, R.; Kamble, S.; Sargaonkar, A. Risk Based Analysis for Contamination Event Selection and Optimal Sensor Placement for Intermittent Water Distribution Network Security. *Water Resources Management* **2016**, *30*, 2671–2685.
13. Zeng, D.; Gu, L.; Lian, L.; Guo, S.; Yao, H.; Hu, J. On cost-efficient sensor placement for contaminant detection in water distribution systems. *IEEE Transactions on Industrial Informatics* **2016**, *12*, 2177–2185. doi:10.1109/TII.2016.2569413.
14. Hu, C.; Dai, L.; Yan, X.; Gong, W.; Liu, X.; Wang, L. Modified NSGA-III for sensor placement in water distribution system. *Information Sciences* **2020**, *509*, 488–500. doi:10.1016/j.ins.2018.06.055.
15. Hooshmand, F.; Amerehi, F.; MirHassani, S.A. Risk-Based Models for Optimal Sensor Location Problems in Water Networks. *Journal of Water Resources Planning and Management* **2020**, *146*, 04020086. doi:10.1061/(ASCE)WR.1943-5452.0001293.

16. Cugueró-Escofet, M.À.; Puig, V.; Quevedo, J. Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: Application to the Barcelona water network. *Control Engineering Practice* **2017**, *63*, 1–12.
17. Soldevila, A.; Blesa, J.; Fernandez-Canti, R.M.; Tornil-Sin, S.; Puig, V. Data-Driven Approach for Leak Localization in Water Distribution Networks Using Pressure Sensors and Spatial Interpolation. *Water* **2019**, *11*.
18. Romano, M., Review of Techniques for Optimal Placement of Pressure and Flow Sensors for Leak/Burst Detection and Localisation in Water Distribution Systems. In *ICT for Smart Water Systems: Measurements and Data Science*; Scozzari, A.; Mounce, S.; Han, D.; Soldovieri, F.; Solomatine, D., Eds.; Springer International Publishing: Cham, 2021; pp. 27–63. doi:10.1007/698_2019_405.
19. Forconi, E.; Kapelan, Z.; Ferrante, M.; Mahmoud, H.; Capponi, C. Risk-based sensor placement methods for burst/leak detection in water distribution systems. *Water Science and Technology: Water Supply* **2017**, *17*, 1663–1672.
20. Venkateswaran, P.; Han, Q.; Eguchi, R.T.; Venkatasubramanian, N. Impact driven sensor placement for leak detection in community water networks. 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS). IEEE, 2018, pp. 77–87.
21. Casillas, M.; Puig, V.; Garza-Castañón, L.; Rosich, A. Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **2013**, *13*, 14984–15005.
22. Cugueró-Escofet, M.A.; Puig, V.; Quevedo, J. Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: Application to the Barcelona water network. *Control Engineering Practice* **2017**, *63*, 1 – 12.
23. Pérez, R.; Puig, V.; Pascual, J.; Quevedo, J.; Landeros, E.; Peralta, A. Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice* **2011**, *19*, 1157–1167.
24. Quiñones-Grueiro, M.; Verde, C.; Llanes-Santiago, O. Multi-objective sensor placement for leakage detection and localization in water distribution networks. 4th Conference on Control and Fault Tolerant Systems (SysToL'19). IEEE, 2019, pp. 129–134. doi:10.1109/SYSTOL.2019.8864746.
25. Sarrate, R.; Blesa, J.; Nejjari, F.; Quevedo, J. Sensor placement for leak detection and location in water distribution networks. *Water Science and Technology: Water Supply* **2014**, *14*, 795–803.
26. Zhao, M.; Zhang, C.; Liu, H.; Fu, G.; Wang, Y. Optimal sensor placement for pipe burst detection in water distribution systems using cost–benefit analysis. *Journal of Hydroinformatics* **2020**, *22*, 606–618. doi:10.2166/hydro.2020.158.
27. Cheng, W.; Chen, Y.; Xu, G. Optimizing Sensor Placement and Quantity for Pipe Burst Detection in a Water Distribution Network. *Journal of Water Resources Planning and Management* **2020**, *146*, 04020088. doi:10.1061/(ASCE)WR.1943-5452.0001298.
28. Ponti, A.; Candelieri, A.; Archetti, F. A New Evolutionary Approach to Optimal Sensor Placement in Water Distribution Networks. *Water* **2021**, *13*. doi:10.3390/w13121625.
29. Casillas, M.V.; Garza-Castañón, L.E.; Puig, V. Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms. *Water* **2015**, *7*, 6496–6515.
30. Blesa, J.; Nejjari, F.; Sarrate, R. Robust sensor placement for leak location: analysis and design. *Journal of Hydroinformatics* **2016**, *18*, 136–148.
31. Thomas M. Cover, J.A.T. *Elements of Information Theory*, 2nd ed.; Wiley Series in Telecommunications and Signal Processing, Wiley-Interscience, 2006.
32. Cooper, J.P.; Robinson, L. *Computer Modeling of Water Distribution Systems*, 4 ed.; Manual of Water Supply Practices — M32, American Water Works Association, AWWA, 2017.
33. Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology* **2005**, *3*, 185–205.
34. Fujiwara, O.; Khang, D.B. A two-phase decomposition method for optimal design of looped water distribution networks. *Water resources research* **1990**, *26*, 539–549.
35. Casillas, M.V.; Puig, V.; Garza-Castanón, L.E.; Rosich, A. Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **2013**, *13*, 14984–15005.
36. Alves, D.; Blesa, J.; Duviella, E.; Rajaoarisoa, L. Robust Data-Driven Leak Localization in Water Distribution Networks Using Pressure Measurements and Topological Information. *Sensors* **2021**, *21*. doi:10.3390/s21227551.
37. Rossman, L.A.; Woo, H.; Tryby, M.; Shang, F.; Janke, R.; Haxton, T. EPANET 2.2 User Manual. Technical Report EPA/600/R-20/133, U.S. Environmental Protection Agency, Washington, DC, 2020.
38. Eliades, D.G.; Kyriakou, M.; Vrachimis, S.; Polycarpou, M.M. EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB. Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI); , 2016; p. 8. doi:10.5281/zenodo.831493.
39. Ferrandez-Gamot, L.; Busson, P.; Blesa, J.; Tornil-Sin, S.; Puig, V.; Duviella, E.; Soldevila, A. Leak localization in water distribution networks using pressure residuals and classifiers. *IFAC-PapersOnLine* **2015**, *48*, 220–225.
40. Santos-Ruiz, I.; Blesa, J.; Puig, V.; López-Estrada, F. Leak localization in water distribution networks using classifiers with cosenoidal features. *IFAC-PapersOnLine* **2020**, *53*, 16697–16702.
41. Soldevila, A.; Blesa, J.; Tornil-Sin, S.; Fernandez-Canti, R.M.; Puig, V. Sensor placement for classifier-based leak localization in water distribution networks using hybrid feature selection. *Computers & Chemical Engineering* **2018**, *108*, 152–162.