doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14056

MULTI-LEAK DETECTION AND ISOLATION IN WATER DISTRIBUTION NETWORKS

Débora Alves[1], Joaquim Blesa[2], Eric Duviella[3], Lala Rajaoarisoa[4]

[1] Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politècnica de Catalunya, Campus de Terrassa, Gaia Building, 08222 Terrassa, Barcelona, Spain
[2] Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain
[1,3,4] IMT Nord Europe, Univ. Lille, CERI Digital Systems, F-59000Lille, France
[2] Serra Húnter Fellow, Universitat Politècnica de Catalunya (UPC), Automatic Control Department (ESAII), Eduard Maristany, 16 08019, Barcelona, Spain

[1] *adeboracris@gmail.com*, [2] *joaquim.blesa@upc.edu*, [3] *eric.duviella@imt-nord-europe.fr*, [4] *lala.rajaoarisoa@imt-nord-europe.fr*

Abstract

Water Distribution Networks (WDNs) are complex systems that faces the challenge of detecting and locating water leaks in the system as quickly as possible due to the need for an efficient operation that satisfies the growing world demand for water. This paper introduces an entirely data-driven leak detection and localization method based on flow and pressure analysis. The method can be divided into leak detection when the fusion data of the flow and pressure measurements are studied, thus obtaining the instant where the leak starts and if there is more than one simultaneous leak (multi-leak) occurring in the network. The second part is the leak localization using the fusion of the pressure residues by applying the radial base function (RBF) interpolation to obtain the network zone with the highest leak probability. The method is validated using the L-TOWN benchmark proposed at the Battle of the Leakage Detection and Isolation Methods (BattLeDIM) 2020 challenge.

Keywords
Leakage detection, Leakage localization, Water distribution networks (WDN), Radial base function (RBF), Interpolation.

1 INTRODUCTION

The worldwide growing demand for water generates a constant concern about the proper functioning of the Water Distributions Networks (WDNs). Therefore, the search for new strategies for detecting, estimating, and locating leaks is an important topic, as water leaks are one of the main factors in water loss. In addition, it can produce substantial economic losses, infrastructure damage, and health risks. Because of this, many studies have been carried out to develop WDN leak detection and location methods. Some of the techniques are based on model-based approaches, which provide adequate performance. Still, they rely on the calibration of accurate models and data availability for all possible complex scenarios that some networks are not available. At the same time, data-driven techniques combine standard operation data and topological information to detect and locate the presence of the leak, although they may produce less accurate results.

Works on leak localization applying model-based approaches compare simulated hydraulic information with actual measurements from the WDN; an example, the research in [1] is based on the analysis of pressure residues. In another work in [2], the authors use hydraulic models with AI methods. Moreover, in [3], performing sensitivity analysis and a search space reduction approach to find the leak's location. In [4,5], combine the use of standard operation data and topological information. The particular method in [6] studies the effect of the extra flow when a leak occurs in the pressure sensors presented in the network. It aims at developing a relative incidence of a leak using network topology correlated with the flow and pressure measurement. In [7] has more details about the model-based and data-driven methods.

Another important fact is that in real WDNs, the system can instantly have more than one leak. The Battle of the Leakage Detection and Isolation Methods (BattLeDIM) [8] has raised this concern by presenting the L-Town network representing a small hypothetical town with 782 inner nodes, two reservoirs, and one tank. Several challenges were presented in this challenge. One of them was the rapid detection of leaks and the fact that the system had multiple leaks during the year. The research [9] presented a method of leak detection and estimation using information from flow sensors installed in the reservoir. The technique can give an estimate of the magnitude of the leak, and with a presence of a second leak, the estimation is the sum of these two leaks, being necessary for a human intervention to evaluate the presence of a multi leak.

This work presents a complementary study of leak detection of work [9]. Presenting an entirely data-driven technique to leak detection and localization that tackle multi leaks problems that require minimal topological knowledge of the network and measurements from pressure sensors distributed at a set of inner nodes and flow sensors installed in the inlets. The case study of the L-Town network is analyzed to display the improvement of the method.

The rest of the document is organized as follows: Section 2 presents the leak detection and localization methodology. Section 3 shows the application and the results obtained in the L-TOWN benchmark proposed at the BattLeDIM. Finally, Section 4 concludes this work.

2 METHODS

An overview of the two steps of leak detection and location and the order in which they are applied is illustrated in Fig.1—describing the steps for obtaining the leak initiation time information and calculating the most likely zone to contain a leak. The first leak detection phase descends from the base of sensor fusion theory using the inlet flow and the pressure measurements of the WDN to generate virtual measurements, able to detect the start time of the leaks in a multi-leak scenario. In the second phase, the fused pressure residual of all sensors and the longitude, latitude, and elevation of each node is applied in the radial base function (RBF) interpolation method to determine a network zone with the fault. The two steps of leakage identification and leakage localization are described in detail.

3 LEAK IDENTIFICATION

The fundamental aspect of the detection phases represents the WDN inlet flow and pressure, approximating the current and historical data. Therefore, the demand and pressure forecasts in the WDN are out of the scope of this work. However, it can be assumed that a demand forecast method is calibrated using historical data of the WDN [10] and leak-free pressure estimations that can be computed through available historical data.

The first step of leak identification, LI-1, is the development of the fusion of flow and pressure data. This step transforms each hour of the day into different features, having 24 features, and

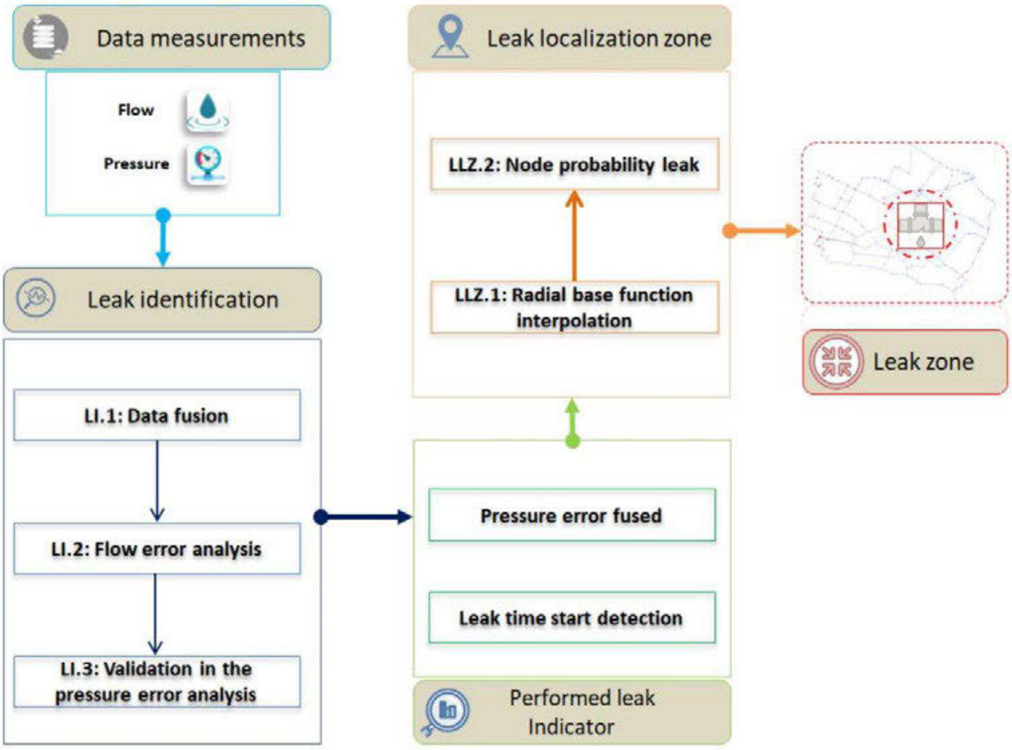

Figure 1. Flowchart of the leak detection and localization proposed method

their fusion improves leak detection thanks to reducing the uncertainties and noise in the measurement.

The first fusion data addressed will be the flow measurement, introduced in [9]. The current inflow y at time k is given as:

$$y(k) = \hat{y}(k) + e(k) \tag{1}$$

where $k = 0,1,2,3, \ldots$ denotes the discrete time corresponding to time $0, T_s, 2T_s, 3T_s, \ldots$, being T_s the sample time of demand forecasting model, $\hat{y}(k)$ is the demand forecast and $e(k)$ is the error that for this study is considered adjusted by a normal distribution (Gaussian) [10], represented by the notation $\mathcal{N}(\mu, \sigma^2(k+T))$ with mean μ and standard deviation $\sigma^2(k+T)$, where T is a periodic variation in time representing the different accuracy of the incoming demand in the periods of the days. In the case of the presence of a leak, i.e.,$l(k) > 0$, equation (1) leads to:

$$y(k) = \hat{y}(k) + e(k) + l(k) \rightarrow \hat{l}(k) = y(k) - \hat{y}(k) = l(k) + e(k) \tag{2}$$

where $\hat{l}(k)$ approximation of the leak size given by the difference between the actual and the estimated inlet flow, with a leak estimation error equal to the demand forecasting error. It is possible to generate different leak estimations using a time window, W considering the current inlet flow value and the previous values:

$$l(k) \approx \bar{l}(k) = \sum_{i=0}^{W-1} \frac{l(k-i)}{W} \tag{3}$$

an average leak estimation $\hat{\bar{l}}(k)$ can be computed at instant k applying the maximum Likelihood estimation method to the joint probability distribution of the W estimations fused in $\bar{l}(k)$

$$\hat{\bar{l}}(k) = \frac{\sum_{i=0}^{W-1} \frac{\hat{l}(k-i)}{\sigma^2(k-i)}}{\sum_{i=0}^{W-1} \frac{1}{\sigma^2(k-i)}} \tag{4}$$

Leak detection can be formulated as a change detection problem, $\hat{\bar{l}}$ will lead to small (but non-zero) values due to demand estimation errors in a no-leakage scenario. In contrast, its value will increase in a leakage scenario. Therefore, a threshold ∇ can be calculated to determine the value of $\hat{\bar{l}}$ above it, which can be assumed to be a leak in the WDN.

The value of ∇ can be calculated by applying equation (4) for leak-free historical data, considering the worst-case scenario ∇ equal to the maximum value of $\hat{\bar{l}}$ calculated for all leak-free historical data, referent the LI-2 step. Furthermore, once $\hat{\bar{l}}$ is above ∇ is considered a disturbance in the system alarming to a probable presence of a leak that needs to be validated with the study of data fusion of pressure measurements, which will be explained in the next topic.

Data fusion of pressure measurements is performed by analyzing pressure residues generated by comparing internal pressure measurements and leak-free pressure for each sensor, installed in the WDN, estimates such as:

$$r_i(k) = \hat{p}_i(c(k)) - p_i(c(k)) \qquad i = 1, \ldots, s \tag{5}$$

where $r_i(k)$, $\hat{p}_i(c(k))$ and $p_i(c(k))$ are the residual, leak-free pressure estimation, and pressure measurement at inner node i, $c(k)$ is the operating condition at given instant k defined by inlet measurements and s is the number of inner sensors installed in the WDN. In the same way as equation (3), it is possible to generate different residuals analyses using a time window, W, (the same value of the leak estimations) considering the current residual pressure value and the previous values. The average pressure residuals $\bar{r}_i$ can be computed at instant k applying the maximum Likelihood estimation method to the joint probability distribution of the W residuals analyses fused in:

$$\bar{r}_i(k) = \frac{\sum_{i=0}^{W-1} \frac{r_i(k-i)}{\sigma^2(k-1)}}{\sum_{i=0}^{W-1} \frac{1}{\sigma^2(k-1)}} \qquad i = 1, \ldots, s \tag{6}$$

The finite difference will be applied to demarcate the beginning and end of a leak in the system to the daily data of residuals fused $\bar{r}_i$. Being analyzed, the maximum value in every 24 hours, k_{day}. The finite difference corresponds to differential operation, an important concept in calculus commonly used to smooth nonstationary time series [12] expression of the form $f(x)$ to $f(x + b) - f(x + a)$. In this study, the difference value $\Delta\bar{r}_i$ is calculated as follows:

$$\Delta\bar{r}_i(k_{day}) = \max(\bar{r}_i(k_{day})) - \max(\bar{r}_i(k_{day} - 24)) \qquad i = 1, \ldots, s \tag{7}$$

When a leak occurs in the WDN, all the measurements of the pressure sensors will be affected; nevertheless, if the sensors closest to the failure show more disturbance. Knowing that the network will be divided into α groups $G = \{g_1, \ldots, g_\alpha\}$, with the region and the neighbouring sensors as a parameter. Moreover, the sum of the $\Delta\bar{r}_i$ of each group will be performed, being normalized in a range of [0,1].

In these analyses, a peak is produced in the signal when has a disturbance in the sensors, for example, when a leak starts or when it is fixed. To proceed with the leak detection method, a threshold, th, for each group $g_1, \ldots, g_\alpha$, is calculated with the number of the sensors of the group divided by 3, the leak detection method can be computed by:

$$\Delta\bar{r}_{gi}(k_{day}) = \begin{cases} \Delta\bar{r}_{gi}(k_{day}) & \Delta\bar{r}_{gi}(k_{day}) > th_{gi} \\ 0 & otherwise \end{cases} \quad \text{i=1}, \dots, \alpha, \tag{8}$$

The $\Delta\bar{r}_{gi}$calculated in equation (8) is set to only present disturbances when a failure is similar to a leak in the system. To set the analysis for disturbances like a leak repair signature, the threshold of the first line must be set to $\Delta\bar{r}_{gi}(k_{day}) < -th_{gi}$.

With the study of equation (4), it is possible to analyse whether the WDN leaks, but it is limited to when there is only one leak in the system or when there are more leaks with time spaces of more than time window W. In other words, if multiple leaks co-occur or with a period smaller than W, the information from equation (4) will only show the sum of the magnitude of all leaks. However, with the validation of the information with the equation (8), it is possible to know when multiple leaks happen because it will present a peak in the analysis data, having a better result if the locations of the leaks are in different groups.

Leak localization zone

The interpolation of data for the WDN has already been studied in other works [2,5]. Still, as questioned in work [5], the interpolation of measured pressure to the nodes that do not have sensors trying to identify the fault at a node-level still has a long way to develop. However, the interpolation of leak indicators to determine the zone close to the sensors that have a fault is of great help for water companies as it will reduce the system zone for the leak's location.

To predict zones with unmeasured nodes the method will use the following information: (i) the average pressure residuals of equation (6) available from the installed sensors, (ii) the topological information of the nodes in the network, and (iii) the Radial basis function (RBF) interpolation technique.

RBF provides a very general and flexible way of interpolation in multidimensional spaces, even for unstructured data, where it is often impossible to apply polynomial or spline interpolation, see for more explanation [14-16]. Due to its good approximation properties, it was chosen in this work.

The method usually works in d dimensional Euclidean space which is $\mathbb{R}^d$ fitted with the Euclidean norm $||\cdot||$. The interpolation space consists of all functions of the form:

$$f(\underline{x}) = \sum_{j=1}^{N} \lambda_j \emptyset(||\underline{x} - \underline{x}_j||) \tag{9}$$

where $\underline{x}$ is a point in $\mathbb{R}^d$, $\underline{x}_j$ are the centre points for the RBFs (equation (6)), λ_j are coefficients to determine, N are points in this space at which the function to be approximated is known, and $\emptyset(r)$ is a radial basis function, set as a multiquadric problem:

$$\emptyset(r) = \sqrt{1 + \varepsilon^2 r^2} \tag{10}$$

where ε is the shape parameter (see [13]). The RBF interpolation can be used in any dimension; in this work, the dimensions used are the latitude, longitude, and elevation of each node in the WDN, and the average pressure residuals of equation (6) are the values to be interpolated.

4 CASE STUDY

The Battle of the Leakage Detection and Isolation Methods is a challenge provided by the organizers of the BattLeDIM [8]. The aim is to detect and locate several leaks in a hypothetical city created with this intent, as depicted in Fig. 2. the city is located in the Northern hemisphere and regroups a population of about 10,000 people. Thus, higher water usage is expected around July/August and lower in December/January. The network is divided into three distinct areas:

Area A is supplied by two reservoirs, each containing flow sensors; Area B that was installed with a pressure reduction valve (PRV) to help reduce background leakages; and Area C was installed with a pump and a water tank, with a flow sensor in this pump to control the flow that enters in the tank. In addition, has been installed in Area C 82 Automated Metered Readings (AMRs), which is a technology used in utility meters for collecting data that does not require physical access or visual inspection. The data can be transmitted to a central database, in this area, only ten regular sensors were distributed. Area C has a significant quantity of AMRS installed in the zone. Because of that, a model-based approach is a good option to solve the leak localization problem in this area.

In this challenge, the network can be divided into two distinct parts with different challenges: the first, Area A and Area C containing simultaneous leakage, and the second, Area C containing the AMR devices.

The leaks in Area A and B of the 2018 year will be addressed in this work. The data set of the BattLeDIM for this year contains the time and repair location of 9 pipe bursts that were fixed. Three types of leaks exist:

- Small background leaks with 1%–5% of the average inflow
- Medium pipe breaks with 5%–10%
- Large pipe bursts with leakage flow of more than 10% of the average system inflow (≈50l/s)

The water utility corrects significant leaks with a flow rate above 4.5 l/s after a reasonable time within two months. The leakages have two different time profiles: either abrupt bursts with

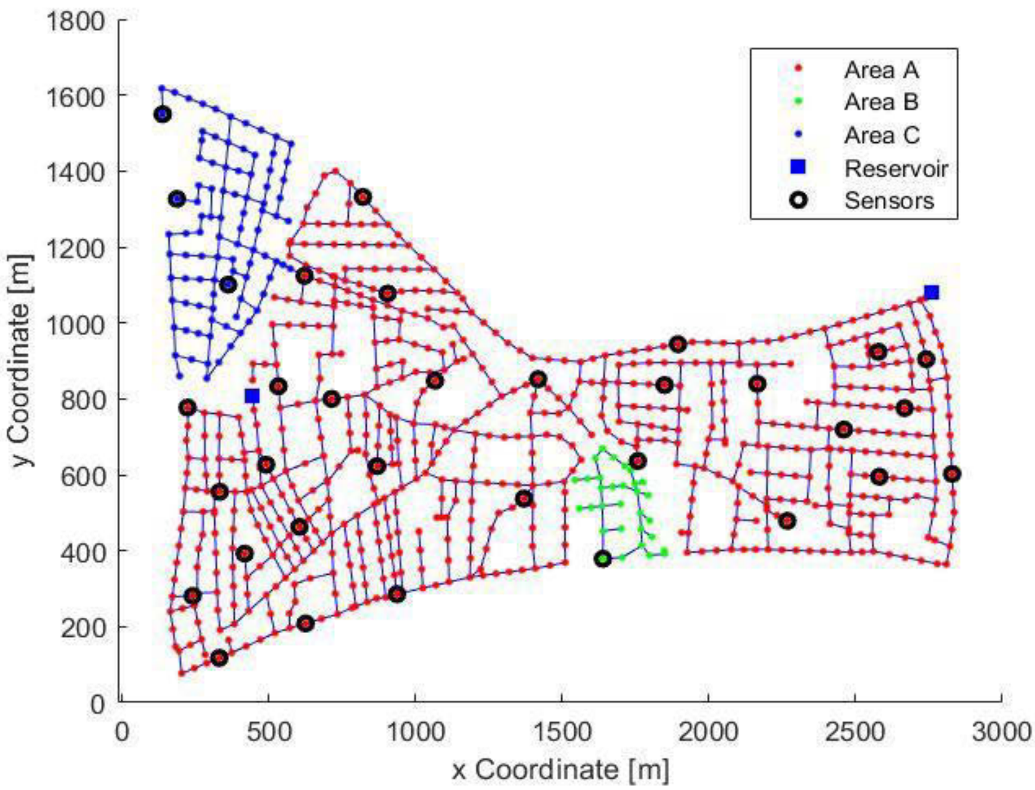

Figure 2. Overview of L-town water distribution network

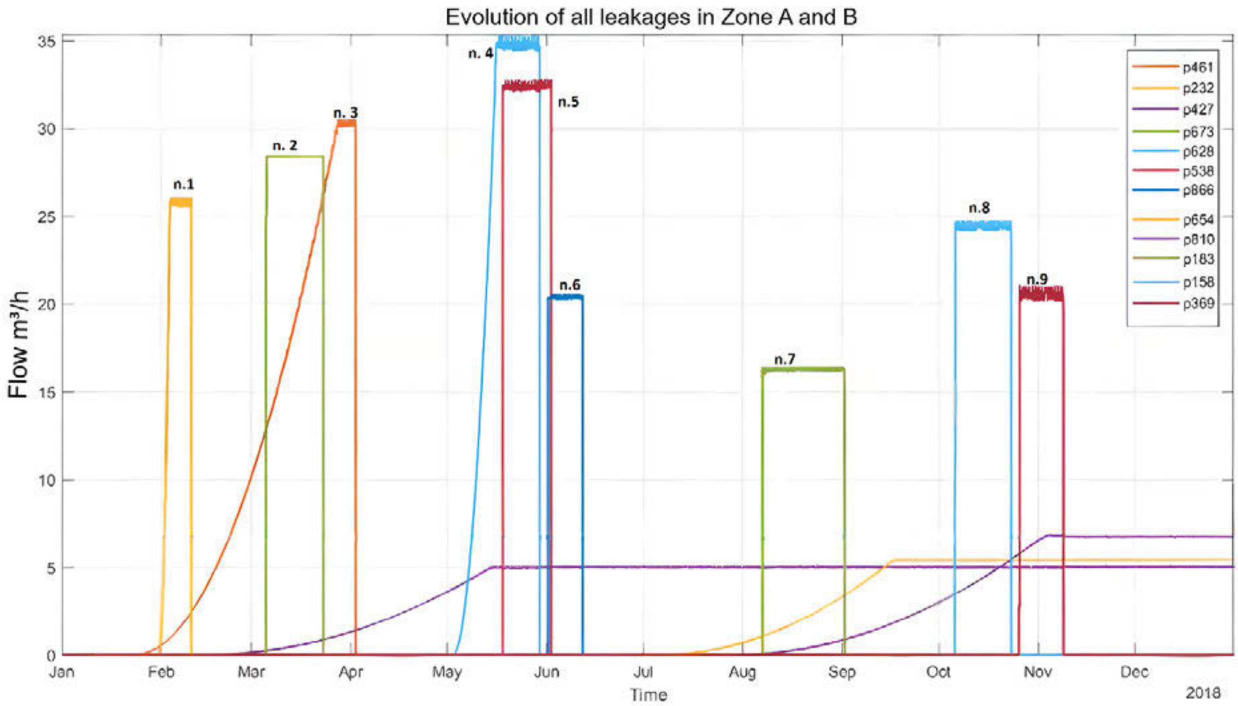

Figure 3. Evolution of leaks (m³/h) in Area A and B during 2018

constant leak flow rates or incipient leaks that evolve until significant outflow rates at which they remain constant. Fig.3 shows the 12 leaks in 2018, with outflow rates between 1.4 and 9.7 l/s (5 and 35 m ^{3}/h). Three leaks are not fixed, and nine leaks are repaired throughout the year that will be analysed in this paper in the highlighted order of n.1 to n.9.

To perform the first step of the proposed approach, it is necessary to define the sensors belonging to group G. In this work, the groups were obtained by the heuristic approach considering the neighboring sensors and the distance between them. The groups do not have the same number of sensors since group 1 has more sensors concentrated in the same area. Another factor is the use of the pressure sensor data in more than one group because if a leak happens in the border zone between groups, the fault will be identified in more than one group analysis.

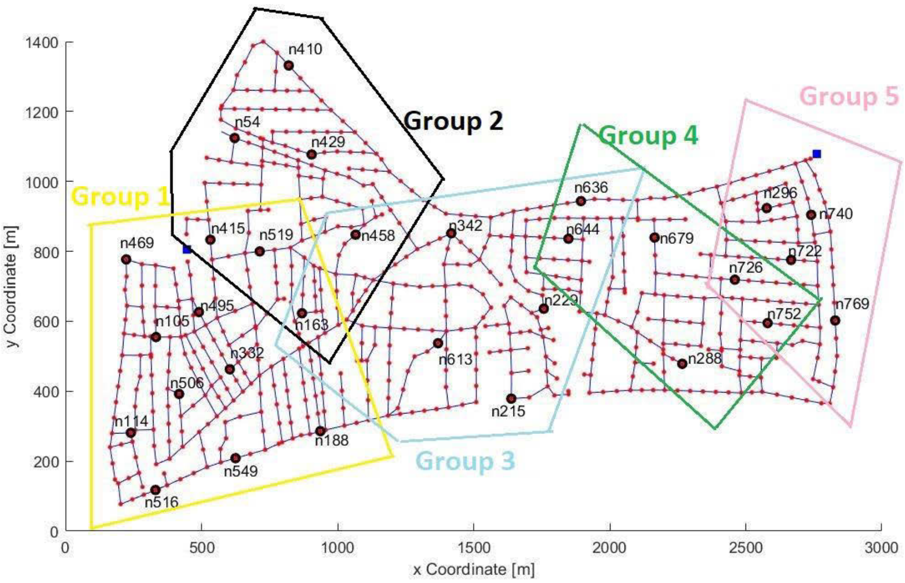

Figure 4. Division of sensors into G groups

Six signals are used to leak detection: first, the $\hat{\bar{l}}$ is calculated with the inner flow measurement, equation (4), and the five-group signals $\Delta\bar{r}_g$ are calculated with the pressure data in equation (8).

Fig.5 presents the result of these six signals, the fig.5 (a) is the $\hat{\bar{l}}$ analysis, which is the first step to detecting a leak. A red circle is highlight for every time k that the threshold is transposit with a red line limited in the pressure analysis $\Delta\bar{r}_{gi}$, fig. 5 (b-f).

When these flow detections happen, it is necessary to validate with the $\Delta\bar{r}_g$ study. When abrupt bursts faults begin in the WDN, it is possible to remark a peak in the $\Delta\bar{r}_g$ analysis in the group more affected by the leak. This is the case of leaks number 1 and 7, and it is possible to point out that leak number 7 started hours before the analysis of the $\hat{\bar{l}}$ alarm the fault. A careful analysis needs to be made in cases where a multi-leak exists, that is leaks number 2-3, leaks number 4-6, e leaks number 8-9.

In leaks number 2-3, and an incipient leak begins in the pipe p427, which was not repaired, but the size magnitude is smaller than the other two, and it is impossible to detect it. The other two

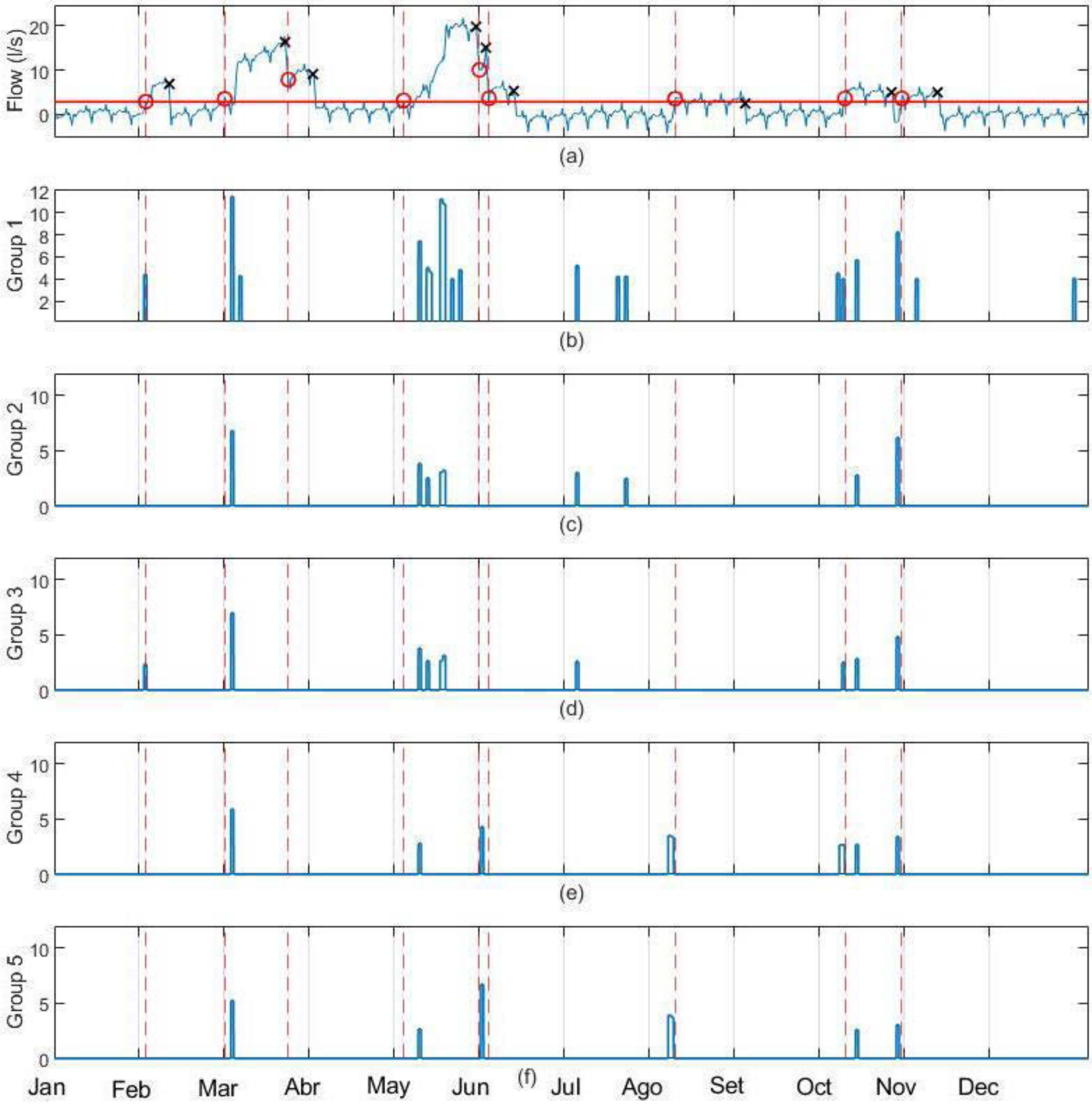

Figure 5. Result of leak detection, red line is the time when $\hat{\bar{l}}$ exceeds the defined threshold

are the types of incipient and bursts. The bursts occur in Area B of the network and affect all Δr_g signal groups. However, this zone is an isolated area with just one sensor, and a study of it can be done, see [9]. A second peak can be detected that happens only in Group 1, indicating a probable second fault in this area: the incipient leak.

Leaks number 4-6 have an extra incipient leak in the pipe p427 that saturates in the meanwhile. The study of $\Delta\bar{r}_g$ of these times instant needs to have more attention because the leaks 4 and 5 are situated near each other in groups 1 and 3, and the leak in the pipe p427 is in group 2. Group 1 has five peaks at this period, with the two most prominent peaks identifying leaks 4 and 5. The other peaks are due to saturation in the pipe p427 and the proximity in the time when leaks start. Leak number 6 is in group 5, and it is easy to identify the start time because it only affects groups 5 and 4.

The leaks 8-9 are not occurring together. However, the system has three saturated leaks in pipes p427, p654, and p610 that achieve the saturation moment during the leak 9. In the analysis of $\hat{\bar{l}}$ in this instant is possible only to identify the leaks 8 and 9. In the Δr_g examination, group 1 is the

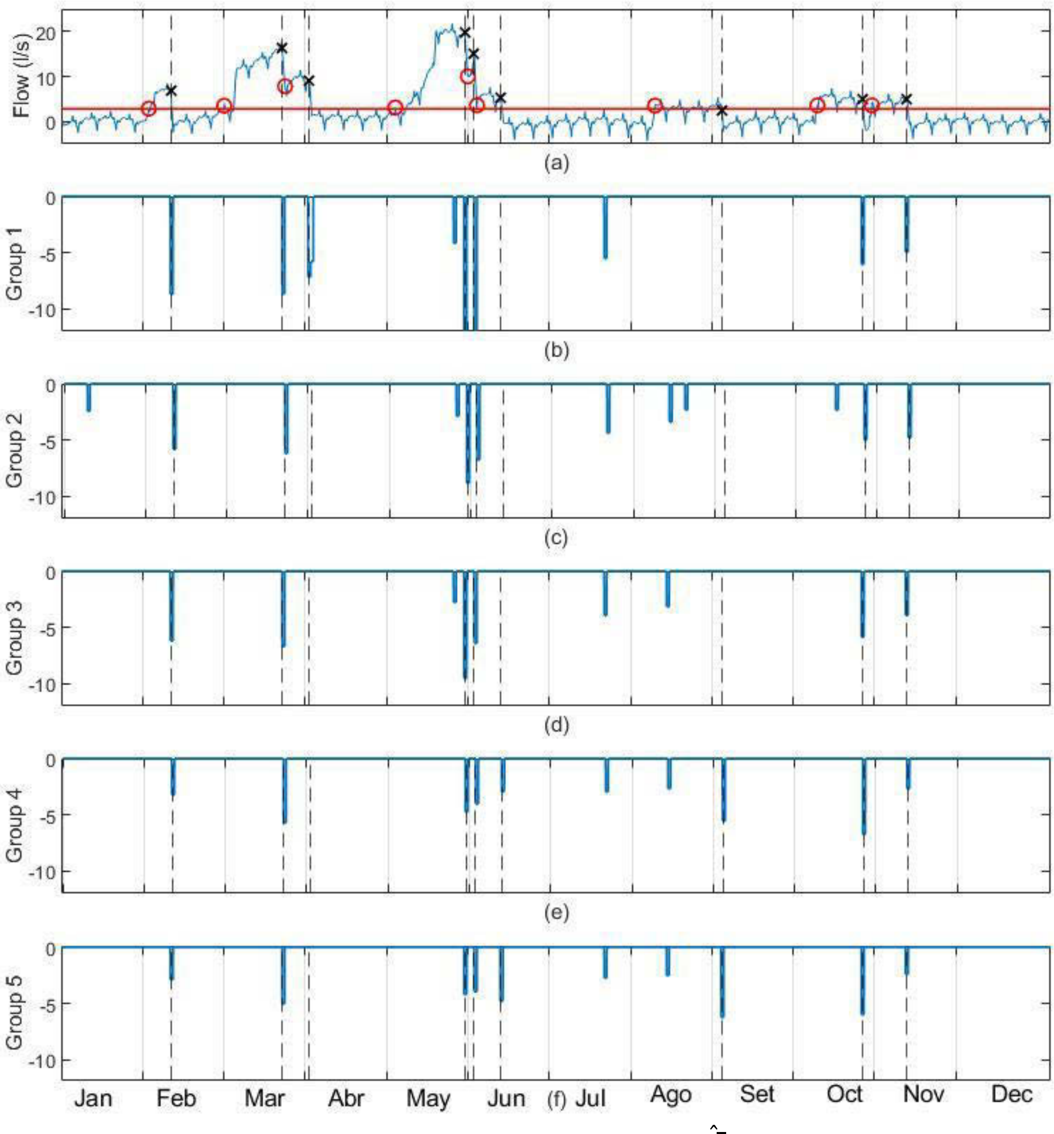

Figure 6. Result of leak detection, red line is the time when $\hat{\bar{l}}$ exceeds the defined threshold

more affected, having five peaks, not making it clear at which time leaks 8 and 9 started but indicating a fault in the WDN.

The same analysis can be done when a leak is fixed. Figure 6 shows these results. Fig. 6 (a) is the same study as $\hat{\bar{l}}$ of Fig. 5(a) but the black line that propagation to the other $\Delta\bar{r}_g$ signal is when a leak is fixed in the zone. In all $\Delta\bar{r}_g$ analyses, a peak negative occurs due to a leak repair; the signal has more than 4 negative peaks caused by some uncertainties of measurements and their estimations.

To perform the second step of the proposed approach, the time instant of each leak begins more than the time they are repaired was used to calculate an average of the residues in equation (5) to apply the RBF interpolation method. Fig. (7) shows the results of the nine fixed leaks. The zones quoted to have a leak vary according to the location of the fault and how it affects the surrounding sensors, but for all leaks retaining the apex in red in the region of the leak.

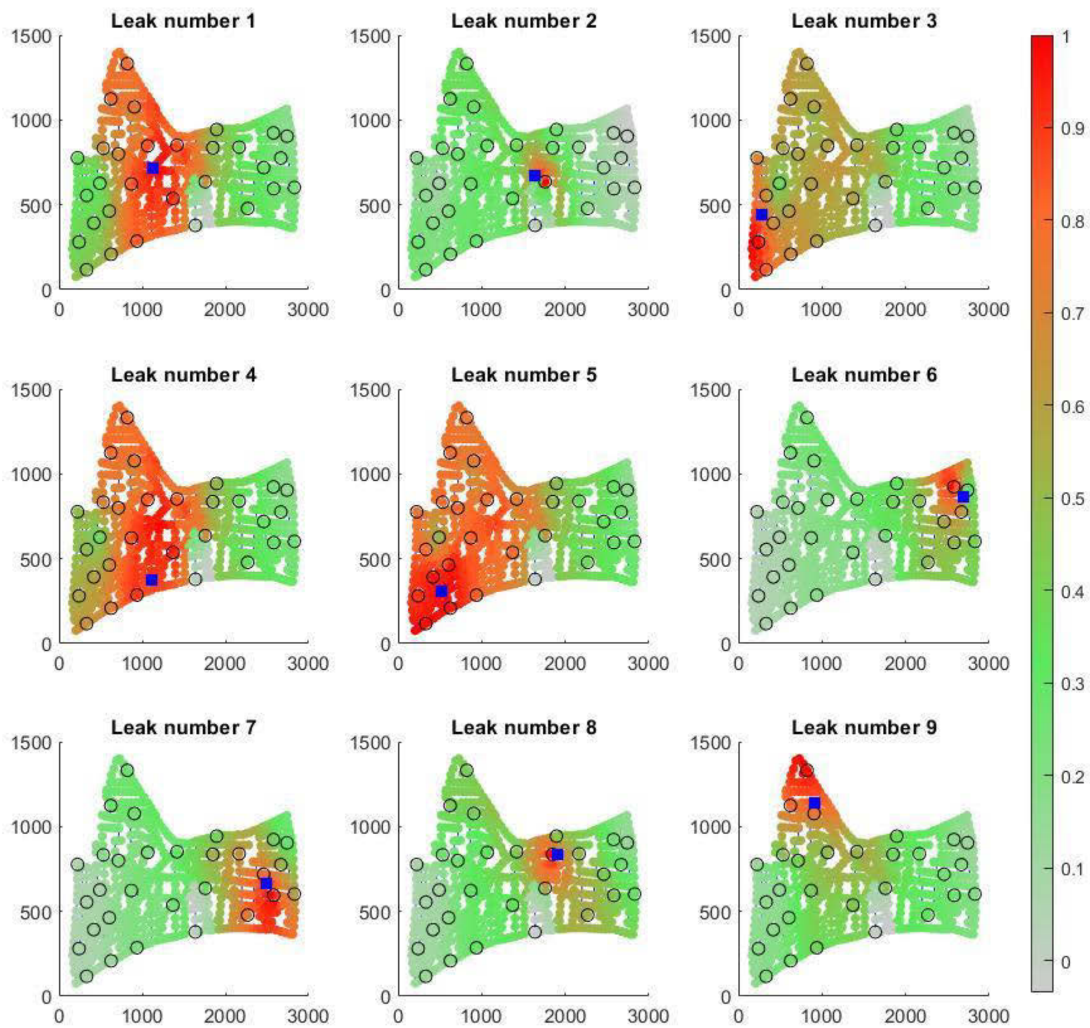

Figure 7. Graphical comparison of the interpolated states for the nine leaks in the WDN

5 CONCLUSIONS

In this work, we present a new complete data-driven method utilizing flow and pressure measurements and the information of longitude, latitude, and elevation of all the nodes in the WDN to leak detection and location of overlapping leakages purposes. The methodology has been explained. It has mainly two phases: first, the leak detection, which converts every hour of the day into features and fuses them to obtain an average flow and pressure measurement signal. The leak detection method is a multi-validate problem that starts with a study of the fused average flow and validates with the analysis of the fused average residual pressure divided by groups made by neighbour sensor and the area of the WDN. The second phase is the leak localization zone that applies the Radial basis function to interpolate the average residual pressure for each sensor to all the nodes in the network, resulting in the zone most likely to have the fault.

The L-town network utilized in the Battle of the Leakage Detection and Isolation Methods has been used as a case study. The data studied were from the year 2018 with 12 leaks and only 9 repaired, having two different temporal profiles: burst pipe and incipient leaks that stature in some instant. The result of the leak detection demonstrates a good result when the leak is of the bursts type leak. On the other hand, detecting when the leak is incipient with a low growth rate is difficult because the method evolves with the data. Moreover, the method can detect simultaneous leaks.

The “leak localization zone” phase is satisfactory, even using only data information and without resorting to hydraulic models. Also, it was possible to locate the leakage area, limiting it to a single leak at a time in the WDN. If simultaneous leaks happen, the leak location zone will be the region closest to the leak, thus increasing the result area. Future work will investigate a clustering method to obtain the groups in the leak detection phase to define the most appropriate leak detection and location procedure.

6 ACKNOWLEDGEMENTS

This work has been partially funded by the project PID2020-115905RB-C21 (L-BEST) funded by MCIN/ AEI /10.13039/501100011033.

7 REFERENCES

[1] Soldevila, A., Fernandez-Canti, R.M., Blesa, J., Tornil Sin, S., and Puig, V. (2016). Leak localization in water distribution networks using model-based bayesian reasoning. In 2016 European Control Conference (ECC),1758–1763. IEEE.

[2] Javadiha, M., Blesa, J., Soldevila, A., and Puig, V. (2019). Leak localization in water distribution networks using deep learning. In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), 1426–1431. IEEE.

[3] Sophocleous, S., D. Savic, and Z. Kapelan. (2019). “Leak localization in a real water distribution network based on search-space nreduction.” J. Water Resour. Plann. Manage.145

[4] Soldevila, A., Boracchi, G., Roveri, M. et al. (2021), “Leak detection and localization in water distribution networks by combining expert knowledge and data-driven models”. Neural Comput & Applic

[5] Sun, Congcong, et al. 1 (2020) "Leak localization in water distribution networks using pressure and data-driven classifier approach." Water 12.: 54

[6] Alves, D., Blesa, J., Duviella, E., & Rajaoarisoa, L. (2021). Robust Data-Driven Leak Localization in Water Distribution Networks Using Pressure Measurements and Topological Information. Sensors, 21(22), 7551

[7] Hu, Z., Chen, B., Chen, W., Tan, D., & Shen, D. (2021). Review of model-based and data-driven approaches for leak detection and location in water distribution systems. Water Supply, 21(7), 3282-3306.

[8] Vrachimis, S. G., Eliades, D. G., Taormina, R., Ostfeld, A., Kapelan, Z., Liu, S., ... & Polycarpou, M. M. (2020). BattLeDIM: Battle of the Leakage Detection and Isolation Methods.

[9] Romero-Ben, L., Alves, D., Blesa, J., Cembrano, G., Puig, V., & Duviella, E. (2022). Leak Localization in Water Distribution Networks Using Data-Driven and Model-Based Approaches. Journal of Water Resources Planning and Management, 148(5), 04022016.

[10] Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban water demand forecasting: review of methods and models. Journal of Water Resources Planning and Management, 140(2), 146-159.

[11] Malik, O. (2016). Probabilistic leak detection and quantification using multi-output Gaussian processes (Doctoral dissertation, University of Southampton).

[12] Flajolet, P., & Sedgewick, R. (1995). Mellin transforms and asymptotics: Finite differences and Rice's integrals. Theoretical Computer Science, 144(1-2), 101-124.

[13] Mongillo, Michael. "Choosing basis functions and shape parameters for radial basis function methods." SIAM undergraduate research online 4.190-209 (2011): 2-6.

[14] Buhmann, M. D. (2003). Radial basis functions: theory and implementations (Vol. 12). Cambridge university press.

[15] Fasshauer, G. E. (2007). Meshfree approximation methods with Matlab. vol. 6 World Scientific Publishing Co Inc.

[16] Wendland, H. (2004). Scattered data approximation (Vol. 17). Cambridge university press.