Battle of the Leakage Detection and Isolation Methods: An Energy Method Analysis using Genetic Algorithms

Juan Saldarriaga[1], Laura Solarte[2], Camilo Salcedo[3], Carlos Montes[4], Laura Martínez[5], María González[6], María Cuello[7], Andrés Ariza[8], Camilo Galindo[9], Néstor Ortiz[10], Cristian Gómez[11] & Sergio Vanegas[12]

Universidad de los Andes, Civil and Environmental Engineering Department and Water Supply and Sewerage Systems Research Center (CIACUA), Carrera 1 Este # 19A - 40, Bogotá, Colombia

[1]jsaldarr@uniandes.edu.co, [2]lm.solarte@uniandes.edu.co, [3]ca.salcedo959@uniandes.edu.co, [4]cd.montes1256@uniandes.edu.co, [5]ls.martinez10@uniandes.edu.co, [6]ma.gonzalezm1@uniandes.edu.co, [7]mm.cuello547@uniandes.edu.co, [8]ad.ariza@uniandes.edu.co, [9]c.galindo@uniandes.edu.co, [10]nv.ortiz@uniandes.edu.co, [11]cc.gomez@uniandes.edu.co & [12]sm.vanegas@uniandes.edu.co

Keywords: Leakages, emitters, Genetic Algorithms, EPANET, water networks.

1. Context: Problem approach

Leakages in water networks are a common issue in urban water supply infrastructure. These leakages can occur in pipe junctions and sections and in pipes or accessories that are part of storage/compensation tanks. Usually, leakages can be caused by ground movement or settling due to traffic or poor-quality materials. In addition, pressure fluctuations in the network, pressure transitions, or excessive pressures are a common cause of failures leading to leakages.

When the behavior of leaks is analyzed, it can be seen that they are dependent of the demand, something that varies along the day, because during low demand periods, network pressures are higher, and this results in larger leaks discharges. This behavior can be noted in Figure 1 and Figure 2, the first shows fluctuations of flow along a 24h period for a node before a leak event, after it and the case after leak detection using a hydraulic model. Figure 2 shows how a higher pressure generates a higher demand and leakage. The equation from an emitter can be used to simulate leakage flowrate in a particular node in a water distribution network (WDN) (Wu, et al., 2010).

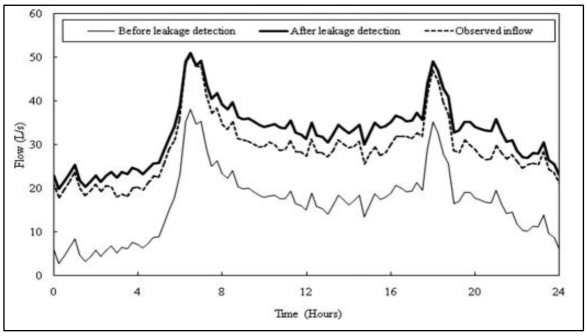

Figure 1.Comparison between flowrate before and after a leak detection throughout a day. Wu et al, 2010.

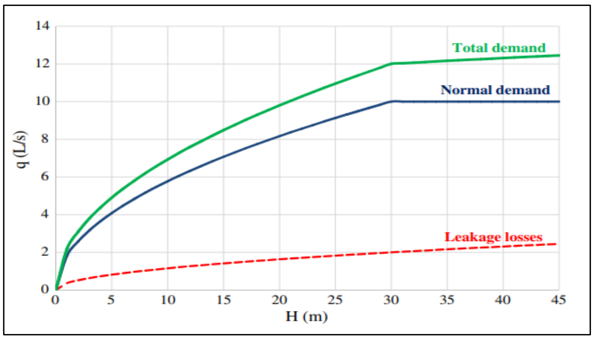

Figure 2.Relation between pressure, demand and leakage in node i. Gupta et al, 2016.

The methodology used to address the objective of this battle requires a network calibrated hydraulic modeling, which requires detailed information of topological, topographical and hydraulic variables. When calibrating the model, possible leaks are simulated in the nodes using emitters, where the leak flowrate is assumed to be a function of the pressure at that point as indicated in the Equation 1.

$$Q_i = \varepsilon_{i_i} P_i^X \quad \text{[Equation 1]}$$

where Q_i is the aggregated leakage flow in the node i, P_i is the pressure in the node i, and ε_{i_i} is the emitter coefficient (it varies in each node) in the node and X the emitter exponent (constant for all nodes in the case study). Pressure variation in a particular node is different (less) when there is an emitter in it representing a leakage, compared with pressure variation without that leakage. This difference plus the difference in discharge curves can be used to locate leakages. Emitter exponent can vary from 0.5 to 2.5 depending on pipe material; it has been determined that the correct value to determine leakage and failure in metallic pipes is of 0.5 (Lambert, 2002), a value used as constant in our approach. On the other hand, an emitter coefficient larger than zero (0) will result in a leak on the node. That means that ε_i of each node must be calculated as an indicator of possible leaks in order to explain changes in pressure and discharge curves. The basic idea is to use a heuristic to find ε_i to each node (most of them are 0) and those cases with values greater than zero indicate that there is a possibility that the pipes connected to the node could have a leak.

2. Data: Patterns and filtration

The available information that describes the hydraulic behavior of L-Town network is classified into four categories: pressure, flow at the DMA inflows and at the pump, demand in "Area C" and tanks water level. In particular, the pressure data presents hydraulic information at internal points along the network and presents patterns that are highly susceptible to change with the presence of leakages. Therefore, the pressure information was implemented for the analysis of the behavior of the leaks in the network. This analysis had three components.

a. Pressure graphs generation and preliminary analysis.

From the existing data, a graph for pressure vs. time could be obtained for different points. In these graphs it is possible to identify different behavior patterns along a normal day, including maximum and minimum pressures, average pressure, rate of variation at different hours of the day, among other information.

b. Anomaly detection.

The anomaly detection analysis is based on establishing differentiating points of behavior in existing pressure data to identify the possible relation between anomalies and occurrence of leaks in the system. An example of the aforementioned can be seen on Figure 1, where there is a variation of pressure close to 0.5 m in comparison with the behavior previously registered. The selection of parameters, if done in a precise way, will allow the identification of points suitable for study in the leakage occurrence analysis.

c. Screening of available data

The same aforementioned process to obtain pressure graphs was done with discharge and tank level data. An initial analysis took place to pave the way a balancing of the flowrates in the network; nevertheless, there was not a direct relation that could lead to the identifying of a specific anomaly in this pattern. On another note, it can be seen that the existing data for water surface level only shows one site which prevents an extrapolation throughout the entire system. Thus, a decision was made to abandon this approach using discharges and water levels, and our analysis was made using only pressure data on some nodes and the analysis of pipe information for those pipes close to a node with any anomaly in order to find candidates for leakage localization.

3. Function: The emitter as a function of pressure

In a WDN model, simulation of leakages can be approximated by emitters, which can be defined as devices associated with the network nodes. These allow to model outflow through a failure on pipe or accessory discharging directly into the ground. In the model, those leaks are located in the node closer to that pipe or accessory. The hydraulic model uses the emitter (Equation 1) in which the exponent X measures the emitter sensibility to pressure in the node and depends mainly on pipe material (Saldarriaga, 2007). For our approach an exponent X of 0.5 was used as it represents the discharge equation of a rigid circular orifice in turbulent flow regime. The coefficient ε_i aims to consider the water loss in the pipe connected to the node in the given period of time and can be determined through GA.

4. Genetic Algorithms: Operation, data recombination and mutation

The use of GA was proposed to solve the battle problem, given that they are greatly used in nonlinear problems as identification of leaks in WDN (Vítkovský, et al., 2000; Wu & Sage, 2006). Its potential relies on the optimization of a function through the combination and mutation of the data, which allows the method to skip local optimal and approach a global optimal. However, if there is a significant part of randomness in the problem, the optimal solution is not always reached, so these methods are part of a heuristic optimization (Keedwell & Khu, 2006). This means that GA performance is highly linked to the number of generations modelled, because it allows the consideration of more alternatives and in return generates a higher probability to achieve an optimal.

GA provides a way of evaluating possible solutions and then determining which are more suitable. Two key elements in GA are the combination and mutation: the combination is the process in which a new generation is produced. Although a variety of methods exist for this procedure (Sivanandam & Deepa, 2007), most methods search for a common trait or traits within the selected solutions, in order to improve the studied population. Mutation adds randomness to the next generations which in turn reduces the probability of encountering local optimums as a final solution. Thus, the GA is part of a collection of plausible solutions in a defined space (i.e. the first generation), and it then selects the best available solutions from the pool using the function. Using it as starting point generates a new set of solutions implementing combination and mutation.

The problem requires the implementation of GA for the exploration of the solution space, assigning different emitter coefficients in different nodes. Then, it uses a hydraulic model to calculate pressure behavior on those nodes in which the pressure data was given. By using the objective function applied to those pressure graphs, the GA proceeds to find a better solution changing nodes location and coefficients. The function must be expressed in terms of hydraulic and topological variables (Saldarriaga, 2007). Subsequently, the genes that make up each individual are the topographic coordinates of the leaking node, the coefficient of variation, and the time at which the leakage occurs.

5. Objective function: RMSE

A method is required to evaluate the precision and accuracy of the obtained results in a specific model before implementing it for the final result calculation. For the model, it is necessary to classify the veracity of the results obtained by the different individuals with the use of heuristics, so the identification of a mechanism that best adapts to the problem can be found. Our GA approach thus used the root of the quadratic mean error or RMSE between the pressure signals calculated and recorded in the database (Mulligan & Wainwright, 2004), due the fact that it is an evaluation criterion used in a variety of cases of autonomous supervised learning.

The RMSE is mainly effective because it quantifies the quality of the prediction obtain compared to the real results. It also represents a low computational cost, compared to other methods. Nonetheless, since the analysis only considers the variation of pressure in the studied node, a decision was made to implement the RMSE with the measured (given) pressure signal and the estimated one for each individual and generation on that node. With the criterion, it is possible to identify the leak present in the data base, as it approaches the real site of the leak both in time and location. With this, it is possible to evaluate precisely the performance of the individuals generated carrying out the Implemented GA, obtaining a better approximation to the function of leak detection implemented.

6. Application: Implementation of GA

For the implementation of the previously described model, an intuitive programming language that allowed the interaction with the network simulated in EPANET was used. MATLAB was thus selected as the numerical computation system for this implementation. The Toolkit "EPANET-MATLAB" operates through the programming language in MATLAB, the hydraulic modelling and water quality software EPANET. Modelling with the EPANET software was conducted in order to obtain an approximation to the real network that allows to establish flowrates on each pipe, pressure in each node and to detect where the hydraulic gradient line is located (HGL). This interaction allows to add the parameters from the emitter equation and analyze the data from the hydraulic network for the different individuals and generations required by the GA.

The process is as follows; initially, the difference between signals from EPANET and SCADA 2019 was quantified using RMSE; if this measurement was greater than 0.2 for any instant, then for that instant a leakage is identified. In order to identify the same leakage twice, a restriction was implemented so that 2 leakages must be separated at least two weeks. After all, 8 leakages were identified for 2019. For each leakage location the GA was implemented with 300 generations and 60 individuals. Figure 3 shows the results of the implementation.

Figure 3. Identified Leakages

The leakages detected are located in "Zone A" and are mostly concentrated in two groups. The first group is located near the tank that supplies the "Zone C", i.e. northwest of the network, for the period between April and June. The above indicates a correlation between the leakages and the time of installation of the pipes. Additionally, the concentration of leakages reflects a significant deterioration in the sector. The second group occurs in the area near reservoir # 1, in the northeast of the network. For February and September respectively. Therefore, the sector is progressively affected by degradation throughout the year.

SUMMARY

Solving the problem of this battle required the use of a calibrated hydraulic model for the network. Within the hydraulic modelling, the nodes represent possible water entrances (from tanks, for example) and exits (demands and/or leakages) in the distribution network. When calibrating the model, the leakages are simulated in the nodes through emitters, where the flowrate from leakage is assumed to be a function of the pressure at that point as previously indicated in the emitter. In the emitter equation, our methodology assumed a constant exponent of 0.5 and a variable coefficient to be determined with the GA.

The use of GA was proposed due to its wide use in nonlinear problems as identification of leaks in a Water Distribution Network WDN (Vítkovský, et al., 2000; Wu & Sage, 2006). Its potential relies on the optimization of a function through the combination and mutation of the data which allows the method to skip local optimal and approach a global optimal. The performance of the GA revolves around the function which varies according to the RMSE between the pressure signals calculated and recorded in the database. The performance of the GA is highly linked to the number of generations modelled because more alternatives can be considered, which generates a higher probability of achieving an optimal. Additionally, the genes that make up each individual are the topographic coordinates of the leaking node, the coefficient of variation, and the time at which the leakage occurs, in order to summarize the hydraulic behavior of leakages in the network. The steps to develop the proposed methodology are explained in detail in the extended abstract above.

Bibliography

Gupta, R., Nair, A. G. R., & Ormsbee, L. (2016). Leakage as Pressure-Driven Demand in Design of Water Distribution Networks. *Journal of Water Resources Planning and Management*, *142*(6), 04016005. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000629

Jimenez-Carrion M. (2018). Simple genetic algorithm to solve the Job Shop Scheduling Problem. *Informacion Tecnologica*, *29*(5), 299–313. https://doi.org/10.4067/S0718-07642018000500299

Keedwell, E., & Khu, S.-T. (2006). A novel evolutionary meta-heuristic for the multi-objective optimization of real-world water distribution networks. *Engineering Optimization*, *38*(3), 319–333. https://doi.org/10.1080/03052150500476308

Lambert, A. O. (2002). International Report: Water losses management and techniques. *Water Supply*, *2*(4), 1–20. https://doi.org/10.2166/ws.2002.0115

Mulligan, M., & Wainwright, J. (2013). Modelling and model building. *Environmental Modelling: Finding Simplicity in Complexity*, 7–26.

OpenWaterAnalytics/EPANET-Matlab-Toolkit. (n.d.). Retrieved May 29, 2020, from https://la.mathworks.com/matlabcentral/fileexchange/25100-openwateranalytics-epanet-matlab-toolkit

Saldarriaga, J. (2016). *Hidráulica de tuberías* (Tercera edición.). Alfaomega. ISBN: 9789586829717

Saldarriaga, J. et al., 2014. An Energy Based Methodology Applied to C-Town. *Procedia Engineering*, pp. 78-86.

Sivanandam, S. N., & Deepa, S. N. (2007). *Introduction to genetic algorithms* (Vol. 1–xix, 442 p. : il). Springer; https://doi.org/10.1007/978-3-540-73190-0

Vítkovský, J. P., Simpson, A. R., & Lambert, M. F. (2000). Leak Detection and Calibration Using Transients and Genetic Algorithms. *Journal of Water Resources Planning and Management*, *126*(4), 262–265. https://doi.org/10.1061/(ASCE)0733-9496(2000)126:4(262)

Wu, Z. Y., & Sage, P. (2008). Water loss detection via genetic algorithm optimization-based model calibration. *Water Distribution Systems Analysis Symposium 2006*, 1–11.

Wu Z.Y., Sage P., & Turtle D. (2010). Pressure-dependent leak detection model and its application to a district water system. *Journal of Water Resources Planning and Management*, *136*(1), 116–128. https://doi.org/10.1061/(ASCE)0733-9496(2010)136:1(116)