

Water Distribution System Leakage Detection and Localization Using Water Demand Automated Meter Reading: An Underexplored and Advantageous Technique

Matheus Pilotto Figueiredo^{1,2*} , Lizandro de Souza Oliveira^{1,2*} , Giancarlo Lucca^{2,4} , Adenauer Correa Yamin^{2,3} , Tiago da Rosa Lopes⁵ 

¹ Catholic University of Pelotas, Pelotas (Brazil)

² Parallel and Distributed Processing Research Group - G3PD, Pelotas (Brazil)

³ Federal University of Pelotas, Pelotas (Brazil)

⁴ Federal University of Rio Grande, Rio Grande (Brazil)

⁵ Autonomous Sanitation Service of Pelotas, Pelotas (Brazil)

* Corresponding author: pilotto@gmail.com (M. P. Figueiredo), lizandro.oliveira@ucpel.edu.br (L. S. Oliveira).

Received 30 April 2025 | Accepted 30 April 2026 | Published 28 August 2026



ABSTRACT

Leakage Detection and Localization (LDL) is one of the greatest challenges to achieving sustainable management in Water Distribution Systems (WDS). As an essential element for the evaluation of LDL approaches, we recently developed LeakG3PD, a simulated dataset that represents WDS behavior in various leakage scenarios, providing all variables and network structure data. Unlike most works in the literature, this article explores the feasibility and advantages of using water demand Automated Meter Reading (AMR) devices, rather than pressure sensors, as the main data source for LDL. To evaluate such an approach, a derived dataset is generated from LeakG3PD using WDS inlet flows and pressures, along with, mainly, consumption demand data. As an introductory work, intuitive and transparent algorithms are applied for LDL. Calculated loss demand is used for detection based on historical thresholds and localization output consists of an ordered list of the pipes in which the insertion of the estimated leak demand causes minimum flow residuals in WDS sources (reservoirs and tanks). A one-step hydraulic simulation is performed to determine the residuals for each candidate pipe. The main difficulties are discussed to provide future improvements. Even though the fundamental step of model calibration is not executed, results indicate promising localization performance for future works. The derived dataset LeakG3PD-OD and the LDL Python programs are made available online for the academic community to contribute with more sophisticated algorithms.

KEYWORDS

Leakage Detection and Localization (LDL), LeakG3PD, Simulated Dataset, Sustainable Management of Water, Water Distribution Systems (WDS).

DOI: 10.9781/ijimai.2026.2245
eLocator: e2245

I. INTRODUCTION

SOCIAL and environmental issues, as well as population growth, have caused the emergence of policies in order to protect the environment and guarantee the rational use of resources. The United Nations (UN) has established the 17 Sustainable Development Goals (SDGs)¹ to its members, a collaborative roadmap for lasting peace and prosperity for both people and the planet. Among the SDGs we highlight the 6th: "Ensure availability and sustainable management of water and sanitation for all". In Brazil, law number 14,026/2020 updates the basic sanitation rules, establishing several obligations to public service delegates, especially the reduction of losses in Water Distribution Systems (WDS) to be achieved until 2033.

¹ <https://sdgs.un.org/goals>

In parallel, technological development has led to the concept of the Internet of Things (IoT), along with an impressive amount of sensors distributed in the most diverse applications and increasingly cheaper.

Once adequate sensors are installed in a WDS, we can acquire their signals and store them in a server as a real dataset, representing the system state over time. Using such a dataset, among other things, it is possible to detect leakages, find their locations and mitigate losses. Despite technological solutions, monitoring of WDS in Brazil and many other countries is still very incipient and, in most cities, not even customer water consumption measurement is automatic yet. Furthermore, this kind of data is usually not public. As a result, real datasets are not available for research. Considering this problem, we recently developed LeakG3PD [1], a simulated dataset, as an alternative to design and evaluate monitoring systems focused on LDL.

The direct hydraulic solution of a WDS model is obtained by inputting node demands, WDS parameters (elevation of nodes, pipe dimensions and roughness, etc.) into a set of non-linear equations related to energy and mass conservation, as well as friction headloss on pipes or headgain on pumps. Such equations are solved by a numerical computation algorithm, such as Newton-Raphson. The output solution is comprised of link flows and node pressures.

As we shall explain, most papers in this research field consider the use of pressure sensors for LDL. In a different direction, this paper evaluates the use of water demand Automated Meter Reading (AMR) devices as an important source for WDS management and LDL.

Here we use a subset of LeakG3PD, consisting of flows and pressures registered from WDS inlets, and, mainly, the consumption demand data, intentionally embedded with missing data, demand fraud and reading errors, as if measured from a real system. This is used to generate a derived dataset incorporating files for calculated flows and pressures along the network. Such dataset, named LeakG3PD-OD, is used to evaluate our LDL approach.

As an introductory work, the LDL algorithms applied are intentionally intuitive and transparent.

Calculated loss demand is used to determine normal condition thresholds, which are used for leak detection.

Localization requires finding an ordered list of pipes in which the insertion of the estimated leak demand causes minimum flow residuals in WDS inlets (reservoirs and tanks). Brute force is applied on every candidate pipe. However the computation involves direct hydraulic simulation to determine the residuals and only evaluates for the leak start time, so computational effort is reduced.

The main difficulties observed are discussed to provide future improvements. Model calibration was skipped, which affected localization performance, revealing the importance of that step and the quality of LeakG3PD concerning model uncertainties. Even so, results are promising for future works.

The main contributions of this paper are as follows: detailed discussion comparing AMR devices and pressure sensors for WDS management and LDL; a step-by-step road map on LeakG3PD usage to obtain the specific derived dataset; simple and transparent LDL algorithms along with thorough discussion on functioning and difficulties observed; and online availability of produced dataset and programs for the academic community.

This paper is organized as follows. Section II positions our work among other related literature contributions. Section III justifies our choice to focus on demand AMR devices comparing them to pressure sensors. In Section IV we explain the method to generate LeakG3PD-OD and the principles behind the LDL algorithms applied to it. Section V presents the results obtained. Section VI discusses the results, limitations observed and solutions. Section VII is the conclusion, where we point out directions for better results in future works and provide the LeakG3PD-OD access link.

II. LITERATURE OVERVIEW

A. AMR for LDL as an Underexplored Approach in the Literature

Literature review paper [2] states that **"Until recently, online demand data using AMR has not usually been considered in the literature, as their installation on real WDNs is not currently extended"**. However, the opposite causal relationship may also be true. It may be argued that if studies are conducted proving the effectiveness of using such data for LDL, the installation of these devices could be increased, considering the various associated advantages, as we shall discuss in Section III.

The same author [2] continues his explanation by saying that **"To remark the usage of this kind of meters in the BattLeDIM2020 articles, the facilitated benchmark included them in one area of the WDN. It is interesting to remark that the BattLeDIM2020 case study did involve a section with AMRs, so that the methods in this contest also use this information"**. However, as observed in paper written by one of the BattLeDIM2020 winner teams [3], demand data was used only to calibrate the network consumption base demand and pattern, while the main signals for leakage localization were from pressure sensors.

Another literature review [4] presents a pie chart of sensor technologies used for LDL in academic papers where 26.4% correspond to pressure sensors while only 2.6% are relative to flow rate sensors.

B. Classification of LDL Approaches

According to [4], **the kind of LDL system proposed here is defined as static (or software-based)**, where sensors' data is transmitted, centralized in a server and used to identify, localize, and pinpoint leaks. On the other hand, dynamic (or hardware-based) LDL systems are those that rely on moving leak detection devices, such as listening sticks or vibration sensors, to the suspected leakage area, pinpointing the leakage *in loco*.

Literature review article [2] indicates that software-based methods are divided according to the underlying information available about the WDS. When a high-fidelity model is available in a simulation software, the methodologies used are named model-based and the LDL process is realized comparing incoming real world data with simulated information. On the other hand, when a model is not available, data-driven methodologies are applied, mining knowledge from real datasets. Finally, when a WDS model exists but is not precise enough, mixed model-based/data-driven approaches are used, where calibration, complexity and modeling errors issues don't affect the results directly since the model is used only for training, whereas a machine learning algorithm is primarily responsible for LDL, using real world measured data. **According to such classification, the work conducted in this article classifies as model-based.**

Model-based LDL methodologies are divided according to the kind of algorithm used. Inverse-problem approaches are the most numerous and apply numerical computation in the contrary direction of direct hydraulic solution, this is, based on a set of measured pressures and/or flows, the hydraulic equations are set, using optimization algorithms, such as Levenberg-Marquardt, to find leak demands and their locations. In Sensitivity-based methods a Fault Sensitivity Matrix (FSM) stores the effects of each possible leak on every node of the network, based on simulation, prior to the operation of the LDL algorithm. There are also Bayesian, Fuzzy logic and, finally, other methodologies, which don't fit exactly any of the previous groups. **Considering the kind of algorithm, our work is classified into other methodologies.**

C. Related Works

Aiming to compare to other previous works in this research field, we use Table I, which is a subset of the one presented in paper [2].

At a first glance, Table I indicates that paper [5], has sensor type and densities similar to ours, with few pressure sensors and demand AMR devices in all nodes. However, actually reading that work, we observe that the measurements assumed by the author are actually only 4 pressures and 4 flows from WDS inlets. Demands are indirectly obtained using a Weighted Least Squares Estimation technique. Basically, the demand estimated as higher than others indicates the leak node. Detection is not addressed, but rather considered as a previously realized step. The synthetic network is unrealistic with many simplifications and no uncertainties are considered.

TABLE I. RELATED WORKS

Article	Problem D= Detection, L= Localization, SE= Size Estimation	Multiple Leak?	Sensor Type P= Pressure, F= Flow, D= Demand	Sensor Density A= Analysis of Different Sensor Configurations	Model and Data Types RN=Real-Based Network, SN= Synthetic Network, SD= Synthetic Data	Uncertainties Considered M= Model, S= Sensor, D= Demand, V= Various or Variable leak sizes	Benchmark?
Andersen and Powell (2000) [5]	L	×	(P,D)	(4, 36)/36	SN	-	-
Cantos et al. (2020) [6]	D-L	×	(F,D)	(93/398, 80/412)	RN(SD)	V	-
Rajeswaran et al. (2017) [7]	L	✓	(F,D)	A	RN(SD)	-	EXNET, DTown, ...
Marzola et al. (2022) [8]	D-L(SE)	✓	(P,D)	(33, 82) /785	RN(SD)	M, S, D, V	BattLeDIM2020
Steffelbauer et al. (2022) [9]	D-L(SE)	✓	(P,D)	(33, 82) /785	RN(SD)	M, S, D, V	BattLeDIM2020
This work	D-L(SE)	×	(P,D)	(3, 82)/92	RN(SD)	M, S, D, V	Net3

Once that work was conducted 25 years ago, real customer demand measurements probably weren't considered as a feasible hypothesis, given the technological solutions and prices at that time.

A more recent work [6] was applied on data obtained from a real WDS: a large-scale pilot project conducted at the Scientific Campus of the University of Lille [10]. Data from flow sensors at system inlets and consumption data from AMR devices are subtracted to obtain loss demand, which is analyzed using Machine Learning techniques to define a leak multinomial likelihood indicator of severity for system nodes. The geolocation depends on previously labeled data, as many data-driven LDL methods, and only one case of localization is depicted based on a real leak event, but the resulting indicators don't seem to pinpoint the leak position accurately.

In paper [7] a graph partitioning approach is conducted. It requires that all supply and demand flow rates be measured, similarly to our proposal. However, a quite odd procedure must be executed in that LDL method: portable ultrasonic flow meters are required to be available for various measurements *in loco*, implying the need for sequential excavations.

The algorithms we applied in this work are in fact intuitive for detection and based on brute force for localization. So there is resemblance to the work developed by the 5th place winner of the BattLeDIM2020 [8], where leaks are detected through "an engineering judgment process of the measured inflows and water demands" and localization "is performed by carrying out simulations of every possible leak, and computing the error between each obtained pressures set and the actual ones from the network, selecting the pipe that yields the lowest error".

The last work worth remarking corresponds to the 1st place winner team [9] of the BattLeDIM. Their LDL algorithm involves the generation of a dual model that automatically converts small pressure deviations caused by leaks into sharp and localized signals in the form of virtual leak flows. It's related to our work by the fact that the measured pressures are reproduced in the dual model by means of EPANET reservoir elements, in the same way we shall present in Section IV. The authors state that "...the sensitivity is much higher in the dual model. Inverting the problem acts as an amplifier of leaks. Another advantage is that the imbalances and the value in question (leaks) have the same unit of flow. The sum of all the imbalances normally gives a good first estimate of the size of the leak."

III. ADVANTAGES OF AMR DEVICES OVER PRESSURE SENSORS

In this section, we present Table II, comparing water demand AMR devices and pressure sensors in terms of the advantages provided.

The comparison table indicates that water demand AMR devices would represent a large-scale investment, which could, however, be offset by cost savings from eliminating manual meter reading. Although the number of required devices is massive, the installation process could be quite easy, with no need for original water meter substitution, depending on the adopted solution, which would reduce costs, environmental effects and water losses. When it comes to system operation, an Internet of Things solution would be required for data transmission, but once these devices will be near to each other, in urban spaces, Wi-Fi communication between them would allow synchronized summation of water demands by blocks, reducing throughput. Individual customer data could be directly saved through their home networks, which could perhaps also be used as gateways to reach the water company main server.

Pressure sensors do not reduce billing process costs. They are, however, very important elements to maintain WDS pressure limits and to allow leakage mitigation when used along with pressure reducing valves [19]. The required quantity of these elements for LDL is usually much smaller, then network throughput is not a problem. However, depending on optimal placement analysis, installation points might require excavations and pipe sectioning, representing water loss and temporary service outage.

As stated in the literature [9], "The performance of leak detection and localization depends on how well the system roughness and demand are calibrated". Except for fraud and unmeasured demands, in our approach the use of AMR devices reflects the exact demand values over time, so demand calibration requires less effort, resulting in a much more accurate model.

When a customer has a particular tank in his property, water demand is related to WDS pressure and depends on the pipe section characteristics between the WDS derivation point and the tank, the elevation of the tank and the float-valve orifice area [20]. This means that water demand AMR devices could also be used to provide indirect WDS pressure measurements, requiring, however, the insertion of these particular customer tanks in the WDS model and the execution of a systematic calibration process.

The advantageousness discussion will not be concluded in this article, because it would require a more rigorous and pragmatic approach, perhaps through a case study of a real system. Such a case study should consider the technological water demand AMR solution to be adopted (which could considerably change the analysis framework) and the number of WDS customers. Costs should be carefully considered regarding: billing process dismissal, installation, maintenance and operation.

TABLE II. COMPARING WATER DEMAND AMR DEVICES TO PRESSURE SENSORS

Feature	Water Demand AMR devices.	Pressure Sensors.
Allow real time water loss calculation, improving management	Yes.	No.
Cost reduction relative to human reading of water consumption, representing return over investment	Yes.	No.
Customer property leakage, abnormal use or water meter violation diagnostics with alert messages	Yes.	No.
Statistics for customers, boosting awareness and reduced water consumption	Yes.	No.
Customers' reliance on billing process is increased	Yes.	No.
Installation removes the need for excavation	Yes, devices always placed in easy access overground locations.	No, the optimal sensor location [11] [12] [13] might be underground.
Installation removes the need for pipe sectioning, avoiding water loss	Maybe, some solutions might avoid the need for water meter substitution [4] [5] [6].	No, a Tee fitting is necessary.
Power supply	Batteries and/or customer electrical AC power supply (usually near water meter location).	Batteries and/or water distribution company electrical AC power supply (only for sensors near pumps or system tanks) ²
Quantity of sensors or devices required	One for each customer. For Net 3, in this work, 92 AMR devices were necessary.	Depends on the specific network and leakage location precision. Paper [17] indicates the need of 17 to 21 pressure sensors for Net3.
WDS pressure monitoring for leakage mitigation and operational safety	Limited. The demand of customers with particular tanks is related to WDS pressure, but such data might be useful only while the tank is filled.	Yes.
Sampling period	Depends on technology. Ultrasonic devices can measure up to 4 times in one second ³ .	Down to milliseconds [18].
Requires new device installation as the number of clients increases	Yes, even though LDL algorithms can deal with missing demand data a new AMR should be installed for each new customer to avoid performance reduction.	No, the number of sensors is determined by WDS physical extension.
Communication network and server requirements	Higher, the quantity of data is much bigger. However, demands can be totaled by area before transmission, reducing requirements at the cost of not registering individual customer real time demands.	Lower, small quantity of sensors.
Removes the need for periodic demand calibration	Partially. Fraud and unmeasured parcels should need to be calibrated for better results.	No.

Finally, it's important to note that UNESCO [21] defines the installation of AMR devices as a fundamental step for Smart Water implementation, prior even to LDL.

IV. METHODOLOGY

To evaluate the proposed approach, based on water demand AMR devices, we used our previous work, LeakG3PD⁴, as starting point. In this section we:

- present LeakG3PD briefly;
- describe the process to generate a derived dataset named LeakG3PD-OD, suited exclusively for water demand AMR based LDL algorithm development and evaluation;
- picture an intuitive leakage detection algorithm;
- describe the application of a brute force algorithm for leakage localization;

A. LeakG3PD

LeakG3PD [1] was introduced as a tool to generate simulated datasets in Water Distribution Systems (WDS), focusing on LDL. It's actually an improved version of LeakDB [22], described by the authors as "a benchmark dataset for leakage diagnosis in water distribution networks". In both works, automatic simulations based on EPANET [23] models were developed using the WNTR [24] [25] [26] Python package to represent the behavior of hydraulic networks under the existence of leakages, providing high temporal variability and plenty of scenarios for the evaluation of LDL algorithms. LeakG3PD is composed by 500 scenarios, each of them with up to 25% random uncertainties in pipe parameters (length, diameter and roughness) to take into account the fact that a real WDS is always different from the WDS model, usually based on implantation projects. Each scenario, except the first, has one or two random pipes embedded with random size leakage nodes. The simulation of these scenarios generates .csv files containing pressures, demands and flows for all nodes and links. Additionally, network EPANET .inp file is provided, along with .csv files for leak demands and labels indicating the presence of a leak for each timestamp. This dataset has the complete states of the network operating ruled by hydraulic laws. LeakG3PD provides datasets for 3 networks. For this paper, we used the Net3 [27] network dataset because it's the biggest and more complex up to now.

² https://www.microsensorcorp.com/Details_Water-Pressure-Sensor-Work-in-Utility-Water-Systems.html

³ <https://www.tisales.com/resources/how-do-ultrasonic-water-meters-work/>

⁴ <https://github.com/matheuspilotto/LeakG3PD>

B. Definition of Real System and System Model

Hereinafter, we will refer to LeakG3PD scenario models as Real System, because in fact it is the source from which we measure as if it were a real system. On the other hand, the original model, with no embedded leakage points and no pipe parameters variation, will be referred to as System Model. Such definitions are illustrated by Fig. 1, which is further explained in Section IV.E.4.

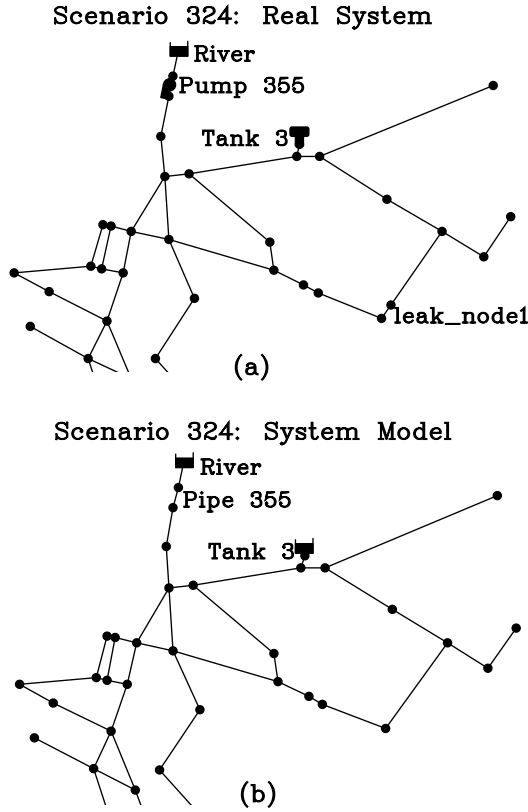


Fig. 1. Graphical representations of Net3 - Real System x System Model.

C. Pipe Equivalent Length

In order to understand the leakage localization results that will be presented in Section V, we shall introduce the empirical Hazen-Williams formula [28], which can be rearranged for SI units as shown in Eq. 1 [29]. It is used to calculate head loss due to friction in pipelines.

$$h_f = \frac{10.65 Q^{1.85} L}{C^{1.85} D^{4.87}}, \quad (1)$$

where:

h_f is the head loss (m)

Q is the volumetric flow rate (m³/s)

L is the length of pipe (m)

C is the pipe roughness unitless coefficient

D is the pipe diameter (m)

As generated by LeakG3PD, maintaining the original idea of LeakDB, Real System has random parametric variations in length, diameter and roughness of each pipe of the network with individual values in aleatory ranges up to $\pm 25\%$. For example, a scenario in which uncertainties ranges were defined as $\pm 20\%$, $\pm 10\%$ and $\pm 15\%$, respectively, a pipe from System Model with unitary parameters could have length = 0.85 m, diameter = 1.1 m, and roughness = 0.9 in Real System.

These variations individually don't seem to be so relevant, but let us analyze the effect on head loss. Inserting uncertainties into Eq. 1 and rearranging as shown in Eq. 2 we obtain the Real System equivalent length parameter based on System Model original length and uncertainty values, considering all the uncertainties concentrated in length variable as if diameter and roughness had their original values unchanged. Note that for the example described in previous paragraph, Real System equivalent length maximum value would occur for $ul=0.2$, $ud=-0.1$ and $ud=-0.15$, so $L_{max} = 3.21$ m. Following the same reasoning, we calculate $L_{min} = 0.34$ m.

$$h_f = \frac{10.65 Q^{1.85} L_0 (1+ul)}{(C_0 (1+uc))^{1.85} (D_0 (1+ud))^{4.87}} = \frac{10.65 Q^{1.85} L_0 \frac{(1+ul)}{(1+uc)^{1.85} (1+ud)^{4.87}}}{C_0^{1.85} D_0^{4.87}} \therefore L = L_0 \frac{(1+ul)}{(1+uc)^{1.85} (1+ud)^{4.87}}, \quad (2)$$

where:

L is the equivalent length of pipe in Real System considering that diameter and roughness had their values maintained (m)

L_0 is the length of pipe in System Model (m)

ul is the unitless length uncertainty

ud is the unitless diameter uncertainty

uc is the unitless roughness uncertainty

D. Leak Demand Modeling

Leak demand is modeled by Eq. 3, which refers to the flow over an orifice whose vertical dimension is small in comparison to the head [28]. In fact it's the same equation used by WNTR simulator to model leakage [25].

$$D_{Leak} = C_d \cdot A \cdot \sqrt{2 \cdot g \cdot h}, \quad (3)$$

where:

D_{Leak} is the leak demand/flow rate (m³/s)

C_d is the discharge coefficient unitless constant

A is the leak area (m²)

g is the gravity acceleration constant (9.80665 m/s²)

h is the pressure (gauge head) at the junction (m)

E. LeakG3PD-OD

Based on LeakG3PD, a derived dataset was developed using basically demand data files, therefore, it was named LeakG3PD-OD, where OD refers to Only Demands. Pressures and flows obtained from simulation on System Model are part of the LeakG3PD-OD dataset. Such variables will be referred to as estimated, because they are just approximations based on a limited set of variables measured from Real System.

In this subsection we describe the steps to generate the LeakG3PD-OD dataset.

1. Selecting Data From LeakG3PD

Keep in mind that the use of any data from LeakG3PD dataset for LDL purpose, such as node pressure, node demand or pump status, for example, presupposes the installation of a sensor in the real world network. Therefore, the feasibility, costs, and operational complexities of sensor installation and maintenance must be taken into account.

Our approach presupposes the use of the following LeakG3PD data files:

- Node demands relative to system consumption points. This corresponds to the installation of water demand AMR devices at the system outlets, i.e., at the customers' water supply input points;

- Node demands relative to system supply points. This implies the installation of high capacity flow sensors, at the system inlets, i.e., at the water treatment plants' output points;
- Node pressures and demands relative to water tanks. This requires the installation of water pressure, level or high capacity flow sensors for tanks. In this case, any of them is enough, because these variables can be inferred from one another;
- Node pressures at the pumps' end nodes. This requires the installation of water pressure sensors at the system inlets, i.e., at the water treatment plants' output points.

2. Addressing Realistic Data Acquisition Issues

In real WDS, especially in developing countries, fraud constitutes an important parcel of apparent water losses. Furthermore the AMR installation process would require time to be completed. Therefore, in order to represent situations related to complete frauds and customers without water demand AMR devices, up to 5 random node demands in each scenario have been set to 0.

Some customers maintain a water meter at the same time they deviate water directly from upstream, as a deception strategy. To represent such partial fraud issue, up to 5 node demands in each scenario have been reduced by a random factor of up to 50%.

Since all measuring devices inevitably produce reading errors, demand values have been changed by a random factor within -5% to +5% in each scenario.

3. System Model Pressure Reproduction Artifices

In this strategy, pressures in Real System water sources are measured. In order to reproduce such pressures, artifices were applied to the System Model.

First, System Model tanks have been replaced by reservoirs, because the latter can be attributed with head patterns according to the pressures measured from Real System over time.

Then, System Model reservoirs have been attributed head patterns, according to the measured pressures. Pump links have been replaced by negligible roughness pipes, avoiding head loss. These pipes have been attributed controls to open or close according to original pumps' behavior.

4. Graphical Explanation of System Model Pressure Reproduction Artifices

Observe in Fig. 1 that (a) Real System has Pump 335, Tank 3 and leak_node1; exact pipe parameters are unknown (length, diameter and roughness). In (b) System Model, Pump 335 has been replaced by a negligible roughness pipe, named Pipe 335; Tank 3 has been replaced by a reservoir; both reservoirs have been attributed head patterns in order to reproduce measured pressures; leak_node0 is not in the network; pipe parameters values are known, although they are up to $\pm 25\%$ different from real system ones.

Note in Fig. 2, that estimated pressure in node 1 (Tank 1) tracks the real model pressure perfectly as a result of the artifices realized. However, estimated pressure in node 61 (Pipe 335 end node) tracks real pressure perfectly only in the moments in which Pipe 335 is open and the difference between real and estimated pressure increases when leak starts, near 5605h.

5. Data Generation

LeakG3PD-OD was generated for 500 scenarios by means of a demand driven simulation⁵ of the System Model conditioned by the pressure reproduction artifices and application of realistic consumption demand values derived from Real System measurements.

⁵ <https://usepa.github.io/WNTR/hydraulics.html>

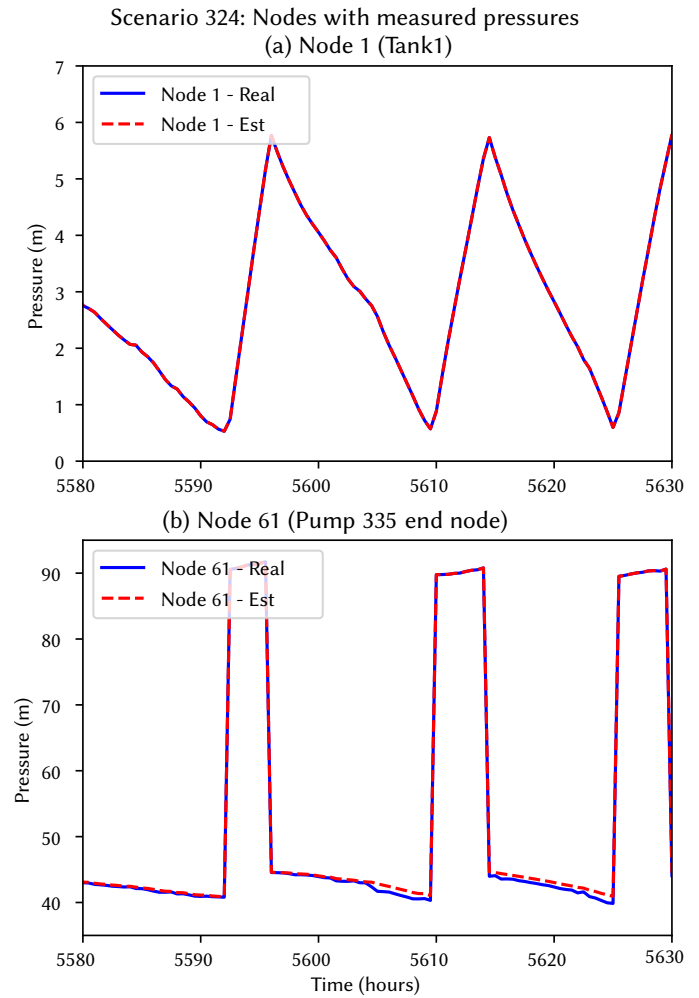


Fig. 2. Scenario 324: Nodes with measured pressures.

The dataset directory structure is shown in Fig. 3. Note that each scenario folder contains the Real System and System Model .inp files, however only the latter shall be used for LDL. Leak folder was copied from LeakG3PD dataset and since leak nodes are not present in the System Model, leak link information files were generated so that each leak point is referred to by its distance to leak link start node. There are files indicating fraud and unmeasured demand junctions. Real System pressure, flow and demand files were deleted and System Model realistic demands, estimated pressure and flow files were added to scenario folder. The dataset folder contains the original System Model .INP file, without the application of pressure reproduction artifices. A summary file was generated to characterize scenarios by their most relevant features. The detection results folder and localization Results.csv file shown are generated by the LDL algorithms, as we shall explain in the following sections.

F. Leak Detection Method

The conservation of mass principle[? seth2019hydraulics] implies that, considering water as a non-compressible fluid, if we calculate the algebraic sum of all node demands in a WDS, the result shall be zero. When reservoirs and tanks have positive output flows their demands are negative; customer node demands are positive when they have positive input flows. In our specific case, the algebraic summation of all available demands won't be zero due to the unmeasured node demands, frauds, leaks and reading errors. If we multiply that algebraic summation by -1, it will actually represent the system loss demand, as expressed by Eq. 4.

```

EPANET Net 3_OD
├── Detection Results
│   ├── leak_det_labels_sc_102.csv
│   └── leak_det_labels_sc_103.csv
├── ...
├── Scenario-99
│   ├── Leaks
│   │   ├── Leak_leak_node0_demand.csv
│   │   ├── Leak_leak_node0_info.csv
│   │   ├── Leak_leak_node1_demand.csv
│   │   ├── Leak_leak_node1_info.csv
│   │   ├── Leak_link0_info.csv
│   │   └── Leak_link1_info.csv
│   ├── EPANET Net 3_OnlyDemands-99.inp
│   ├── EPANET Net 3_Scenario-99.inp
│   ├── Fraud_junctions_info.csv
│   ├── Labels.csv
│   ├── OD_Estimated_Link_flows.csv
│   ├── OD_Estimated_Node_pressures.csv
│   ├── OD_Node_demands.csv
│   ├── Scenario-99_info.csv
│   └── Unmeasured_demand_junctions_info.csv
├── ...
├── EPANET Net 3.inp
├── Results.csv
└── Summary.csv

```

Fig. 3. LeakG3PD-OD Dataset - Directory Structure.

$$D_{loss} = - \sum_n D_n = D_{unmeasured} + D_{fraud} + D_{leak} \quad (4)$$

where:

D_n is the demand for node $n \in \text{System_Model_Nodes}$

To illustrate Eq. 4, in Fig. 4, graph (a), we plot measured loss demand and real leak demand signals together over the one year simulation period of LeakG3PD-OD scenario 88. Note that the real leak demand file was copied from LeakG3PD just for comparison in this figure, but in our strategy framework it is an unknown signal to be inferred based on loss demand.

Graph (b) in Fig. 4 is an amplification, focusing on the moment in which leakage starts, around 6910h. In this scenario there are few unmeasured/fraud nodes and measurement error is 0%, so, when fraudsters are not consuming water, measured loss demand assumes minimum values, which are coincident with real leak demand values.

In this work, we intend to detect and locate leakages based on the loss signal. The leak detector algorithm developed works as illustrated by the flowchart in Fig. 5. UpperDLT is equal to historical maximum loss value, and LowerDLT is set as historical minimum loss plus 20% of the difference between the historical maximum and minimum loss values. Historical maximum and minimum values are calculated every 7h while a leakage is not present nor under suspicion.

G. Leak Localization Method

Analyzing the estimated link flows out of the System Model water sources (tanks and reservoirs), we observed that they were different from the demands measured from Real System, especially after a leakage starts, as illustrated in Fig. 6. Note that in scenario 324 there are two leakages (see Fig. 10), starting around 3000h and 5600h, respectively. This occurs because the water that flows to fraud and leakage nodes in Real System, doesn't have such destination to flow in System Model. Once pressures are fixed in tanks and reservoirs, according to measured values, estimated flows have values so that hydraulic laws and principles are attained. Based on that observation and assuming that leak demand be the biggest parcel in loss demand, our localization strategy consisted of finding the pipe where the

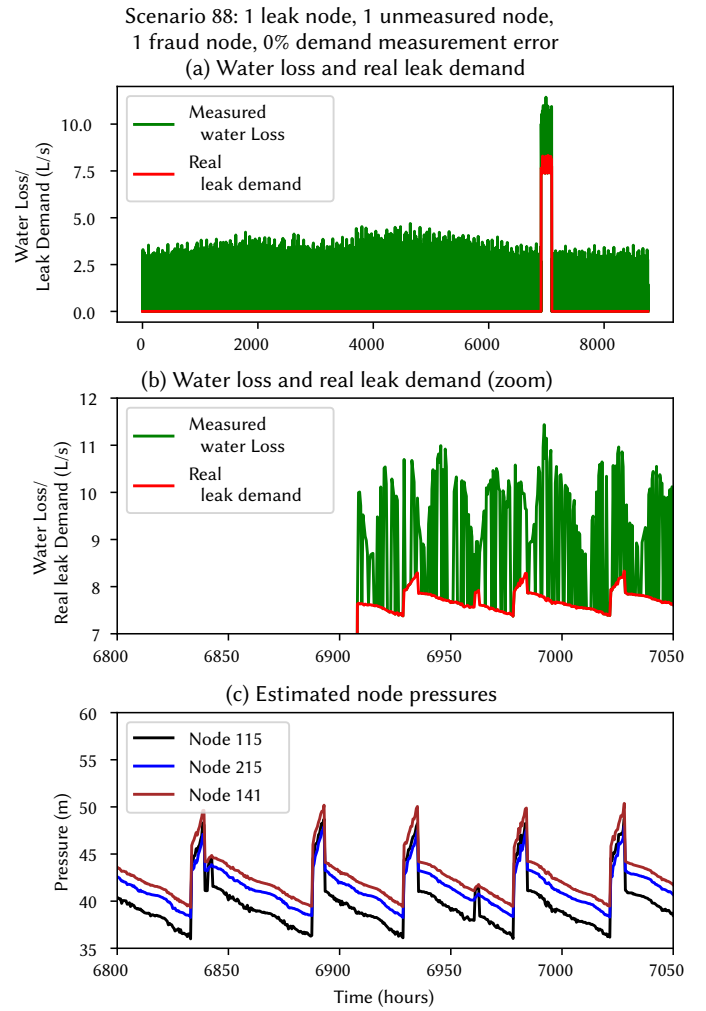


Fig. 4. Scenario 88: Loss x leak x estimated pressure.

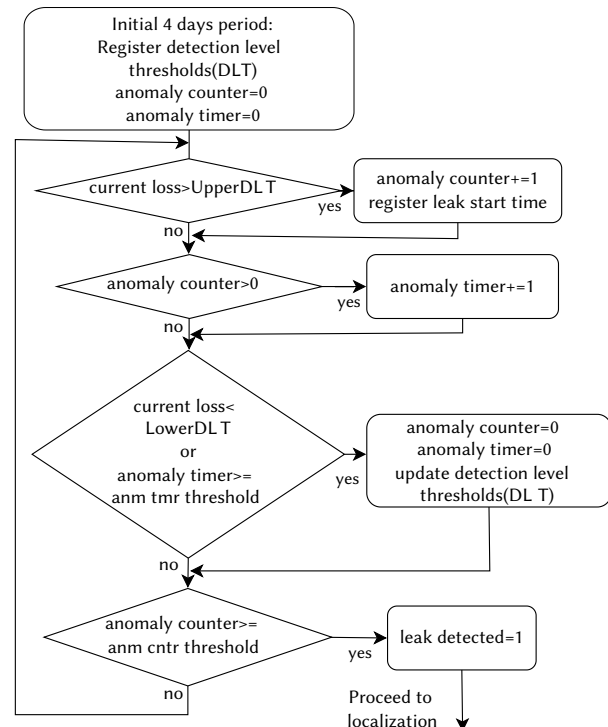


Fig. 5. Leak detection algorithm simplified flowchart.

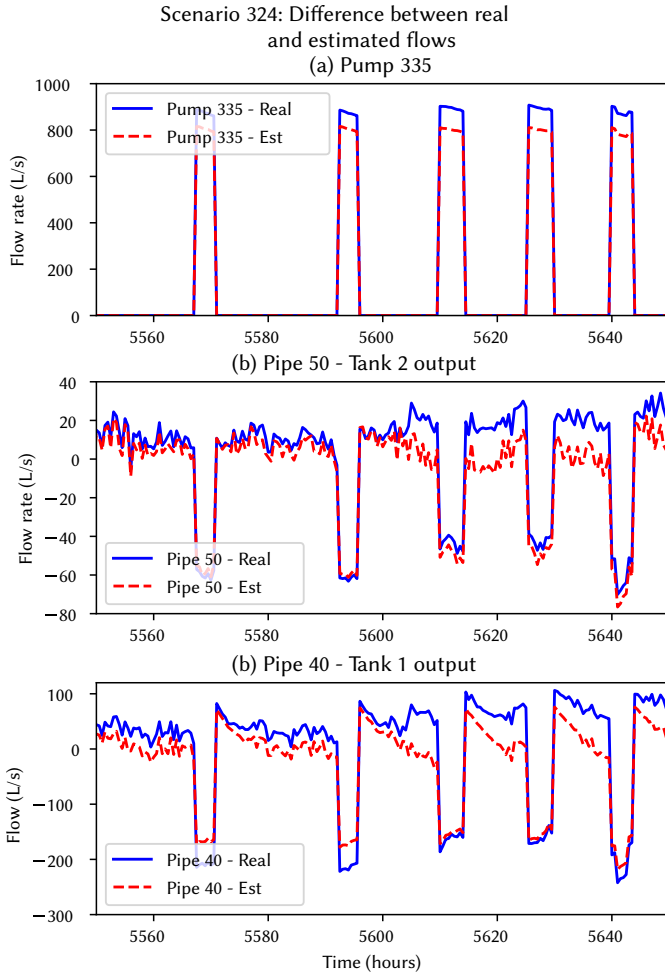


Fig. 6. Scenario 324: Difference between real and estimated flows in water sources. Leakage start time near 5600h.

insertion of a demand equal to the loss demand would minimize the differences between System Model flows and measured node demands for tanks and reservoirs. The algorithm is presented in Fig. 7 as a simplified flowchart.

V. RESULTS

In order to evaluate the detection and localization results, a spreadsheet was automatically created, including:

- Summary characteristics of scenarios, such as number of leaks, leaky pipe names, leak magnitudes, quantities and names of fraud junctions and unmeasured junctions, demand measurement error percentage, scenario parameters uncertainties, total magnitude of fraud and unmeasured junctions demands;
- Evaluation metrics for leak detection, which are Accuracy, Precision, Recall and F1-Score;
- List of guess pipes in ascending order relatively to residuals products, as defined by the localization algorithm;
- Column "True localization found?" that has value 1 when the first pipe on guess list is exactly the leaky one;
- Column "True localization is within the list?" that has value 1 when the leaky pipe is within the list;
- Column "True localization is linked to pipe within the list?" that has value 1 when the leaky pipe is linked to any of the pipes within the guess pipes list.

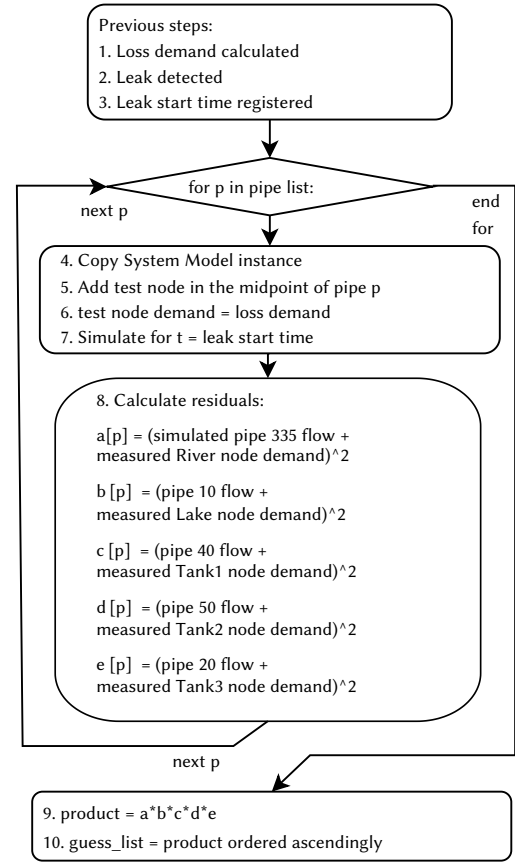


Fig. 7. Leak localization algorithm simplified flowchart. Note that water source node demands are negative when flow out of them is positive, so the residual relatively to estimated flows in their output pipes, is obtained by sum instead of subtraction.

In this work, as an initial approach, we focus only on scenarios with one leakage point. Of the 500 hundred scenarios, 272 have only one leakage.

A. Leakage Detection Results

The results obtained in leakage detection are those shown in Table III. We didn't include Accuracy in the table because it also increases when the number of True Negatives is higher and our algorithm is prone to classify as Negative too easily, so it is not a good metric in this case. On the other hand, F1-Score is a more complete metric combining Precision (out of the instants classified by the detector as leaky, how many were in fact leaky?) and Recall (out of the instants where a leak was present, how many did the detector classify as leaky?). High F1-Score also means early leak detection, which is fundamental to mitigate losses over time. Some detection approaches using exclusively system inlet minimum night flows method may require one entire day to conclude about the existence of leakage.

Leak detection results were registered in the dataset sub-folder named "Detection Results" in the form of .csv files containing the real label from LeakG3PD dataset and the detector label for all timestamps.

TABLE III. LEAK DETECTION RESULTS

F1-Score range	Number of scenarios within range	Percentage of scenarios within range
99%-100%	201	73.90%
90%-99%	31	11.39%
3%-90%	14	5.15%
0%	26	9.56%
Total	272	100%

TABLE IV. LEAK LOCALIZATION RESULTS

Real System pipes max. equivalent length in relation to System Model (%)	Real System pipes min. equivalent length in relation to System Model (%)	Max. leak demand range (L/s)	Number of scenarios in range	True localization found? (Number of scenarios in range/ Percentage of range)	True localization is within the list? (Number of scenarios in range/ Percentage of range)	True localization is linked to pipe within the list? (Number of scenarios in range/ Percentage of range)
148-250	48-43	100-36518	37	4 / 10.8%	25 / 67.6%	34 / 91.9%
		13.7-100	32	1 / 3.1%	10 / 31.2%	18 / 56.2%
		2.4-13.7	11	0 / 0%	0 / 0%	2 / 18.2%
250-516	43-25	100-36518	63	4 / 6.3%	31 / 49.2%	43 / 68.2%
		13.7-100	49	3 / 6.1%	15 / 30.6%	28 / 57.1%
		2.4-13.7	9	0 / 0%	2 / 22.2%	2 / 22.2%
Total			201	12 / 6%	83 / 41.3%	127 / 63.2%

B. Leakage Localization Results

Leakage localization statistics were taken within the 201 one leak scenarios that had detection F1-Score bigger than 99%, and are summarized in Table IV. Note that the table results were divided by ranges of Real System equivalent length of pipes (as defined in Section IV) and maximum leakage demand. The table layout was organized in that manner to reflect more clearly the impact of those two factors upon algorithm performance. In fact, WDS model parameters calibration is a fundamental step, usually addressed before the application of LDL algorithms. In this work, calibration was intentionally skipped to highlight its importance and to demonstrate that LeakG3PD is a challenging dataset for the evaluation of LDL algorithms, in terms of model parameter uncertainties.

Leak localization results were registered in the dataset Results.csv file, which also contains detection results (Accuracy, Precision, Recall and F1-Score) and all important information on each scenario.

Detection and localization results will be discussed in more detail in Section VI.

VI. DISCUSSION

A. Leak Detection Discussion

Leak detection was not our main focus in this work, so leak detector functioning was very simplistic and it has parameters that were empirically defined through iterative tests. We will discuss the implications of increasing or decreasing these parameters in the next paragraphs.

The detector anomaly counter threshold was set to 4 and the anomaly timer threshold was set to 16. This means that if within a period of 16 consecutive time steps the loss signal has 4 anomalous values, i.e., higher than the upper detection level threshold (UpperDLT), without dropping under the LowerDLT a leak is detected. In Fig. 8 we illustrate scenario 83, where F1-Score was zero with no True Positives. Note that, even though the loss didn't drop below the LowerDLT, its value didn't count 4 consecutive anomalies. The anomaly count mechanism was used to reduce leak detection delay time for high leakage situations, because in such cases, in only 4 time steps leak is detected. On the other hand, leakages with low demands relatively to maximum fraud/unknown demand are not enough to cause the detection. To solve this, we could remove the anomaly count mechanism and use just the anomaly timer, detecting leak after the 16 consecutive time steps without loss demand drop below LowerDLT. The obvious disadvantage would be bigger detection time delay and the advantage would be the possibility of detecting smaller leakages. Setting anomaly timer threshold inferior to 16 could result in False

Positive detections if the loss signal had an occasional time period without dropping below LowerDLT.

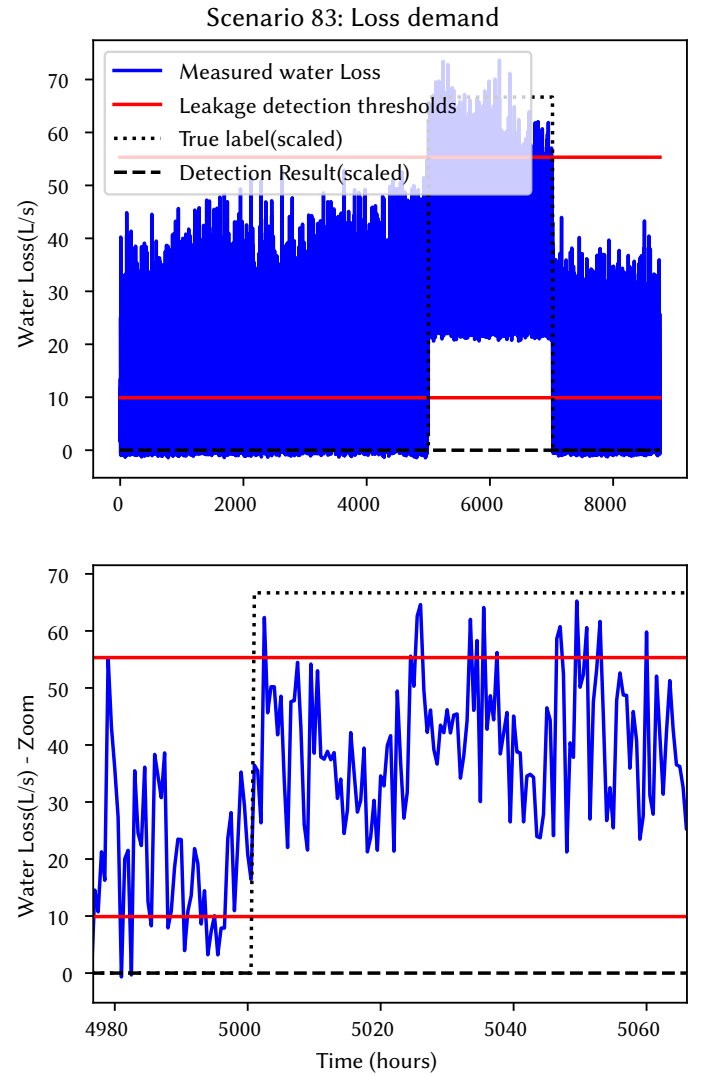


Fig. 8. Scenario 83: Detection F1-Score=0.

After the 16 consecutive time steps relative to anomaly timer threshold, if a leakage is not detected, the detection level thresholds are updated incorporating natural seasonal variations in frauds or even new frauds that started recently. Such mechanism represents an attempt to avoid False Positives related to small loss variations

caused by new frauds over time. However, some adjustments are necessary. Scenario 360 had precision around 48% (see Fig. 9), due to False Positives. The main reason is that LowerDLT was updated based on a singular minimum value around 4810 h (see graph (b), noting that the represented DLT red lines are relative to real leak start time). That event also prevented the detection of leakage end around 8220 h (see graph(c)). To solve that, lower detection level thresholds could be updated based on the mean of historical minimum values, filtering singular extremes.

The detector parameters were clearly set heuristically. However, for future works, adjustments can be done automatically using LeakG3PD dataset and Machine Learning methods for optimization.

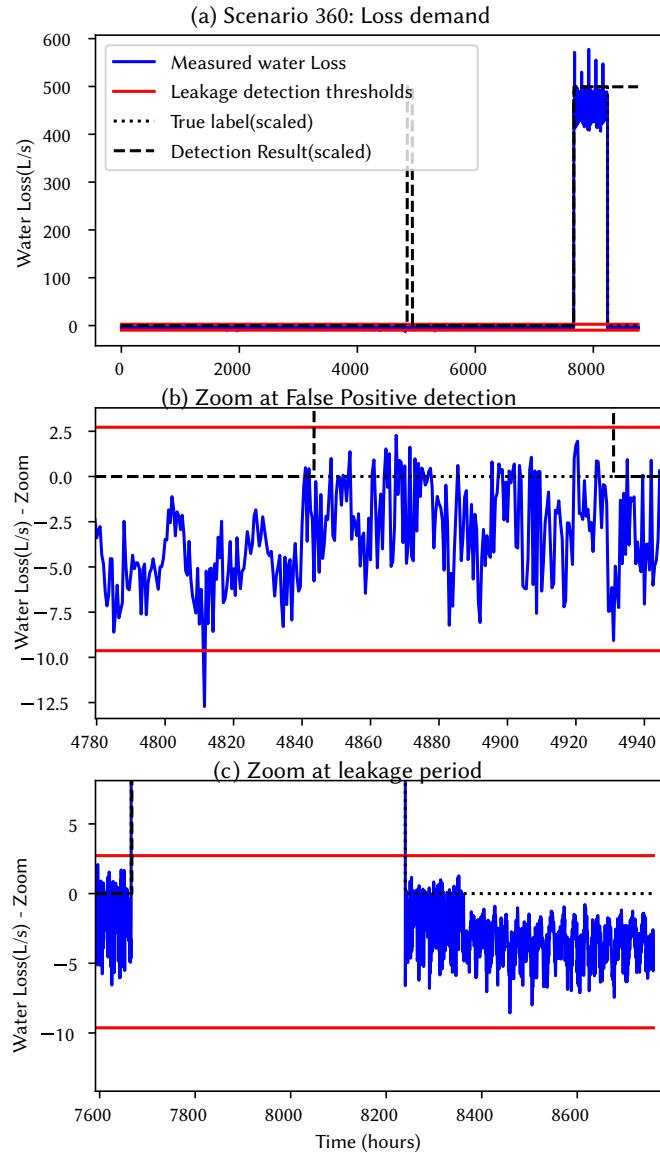


Fig. 9. Scenario 360: Detection F1-Score=65%.

B. Leak Localization Discussion

The list classified by the algorithm has 10 pipes out of 114, which implies that a random guess about the exact leaky pipe corresponds to 0.88% chance to be right and the chance of guessing 10 pipes in which one of them is the leaky pipe is 8.8%. So in Table IV we observe that for small leakages, with maximum demand inferior to 13.7 L/s, results were worst than random guess, as if the algorithm was induced to error. On the other hand, when maximum demand is superior to 100 L/s

and Real System pipes equivalent lengths in relation to System Model are in ranges nearer to 100%, the algorithm found a list of 10 pipes at least linked to the leaky pipe in 91.9% of scenarios. These results are explained by two main reasons:

1. We assumed loss demand to be a rough approximation of leak demand. Maximum total fraud/unknown demands range is 0.09-54.07 L/s, with mean value equal to 17.30 L/s. For bigger leak demands results are increasingly better, because the first two parcels in Eq. 4 become negligible, so $D_{loss} \approx D_{leak}$;
2. As explained in the leak localization results subsection, the apparently reasonable Real System uncertainties embedded by LeakG3PD are actually excessive in terms of the corresponding equivalent length variations produced.

Proceeding, about the second reason mentioned above, note that four years after developing LeakDB, the authors created the BattLeDIM dataset, in which pipe parameters uncertainties are limited to $\pm 10\%$ instead of $\pm 25\%$. This indicates that they may also have considered such uncertainties excessive. During system implantation, pipe lengths may have slightly bigger or smaller sizes due to cutting processes; diameters and roughness will also have slightly bigger or smaller values relative to fabrication. However, after the implantation, considering network aging, the most common situation is that all network pipes will only have roughness increased (C factor decreases) and diameter decreased due to internal corrosion and deposits [30]. However, as we observe in Table IV, in the same Real System network, one pipe can have equivalent length of 516% at the same time other pipe might have equivalent length of 25%, in relation to System Model. Such extreme and opposite simulated variations are actually unreal and will considerably decrease the performance evaluation of any model-based leak localization algorithm.

In order to solve the problem caused by model uncertainties, following the common practice in related model-based literature, a calibration process must be executed prior to the LDL algorithms, optimizing the parameters of System Model to reflect the ones in Real System, based on the measured variables.

Back to leak demand estimation, it's a very important process to work on in order to detect smaller leakages. The situation shown in Fig. 4 (1 unmeasured node, 1 fraud node, 0% measurement error) will rarely occur. The most common scenario will be similar to the one shown in Fig. 10. To deal with such challenge statistical tools will be necessary.

It is important to register that our first attempt to reach leakage localization was through the correlation between estimated pressures (calculated by simulation) and the estimated water leak demand (inferred from loss signal). To do so, we would rely on Eq. 3.

Note that in Fig. 4, comparing graphs (b) and (c), the estimated pressure for node 115 visibly tracks the real leak demand more closely than the others. Not by coincidence, node 115 is actually much nearer to the leakage point in scenario 88 than the other two. However, in other scenarios (see Fig. 11) considerable differences between estimated and real pressure are observed when leakage starts, which led to poor localization results by correlation method, even if real leak demand is used instead of the one inferred from loss demand.

Once the algorithm is based on brute-force residual minimization, it may face scalability and generalization limitations considering multiple and pre-existing leaks. One insight to address such problem might be the insertion of candidate leak nodes, ruled by Eq. 3, in all pipes, initially positioning them in midpoints and defining leak node areas such that calculated loss demand was somewhat divided between them, representing the algorithm parents, for example, five parents, each one equivalent to single leakage points in well distributed areas of the WDS. After that, a Genetic Algorithm could

Scenario 324: 2 leak nodes, 4 unmeasured nodes, 4 fraud nodes,
-3% demand measurement error

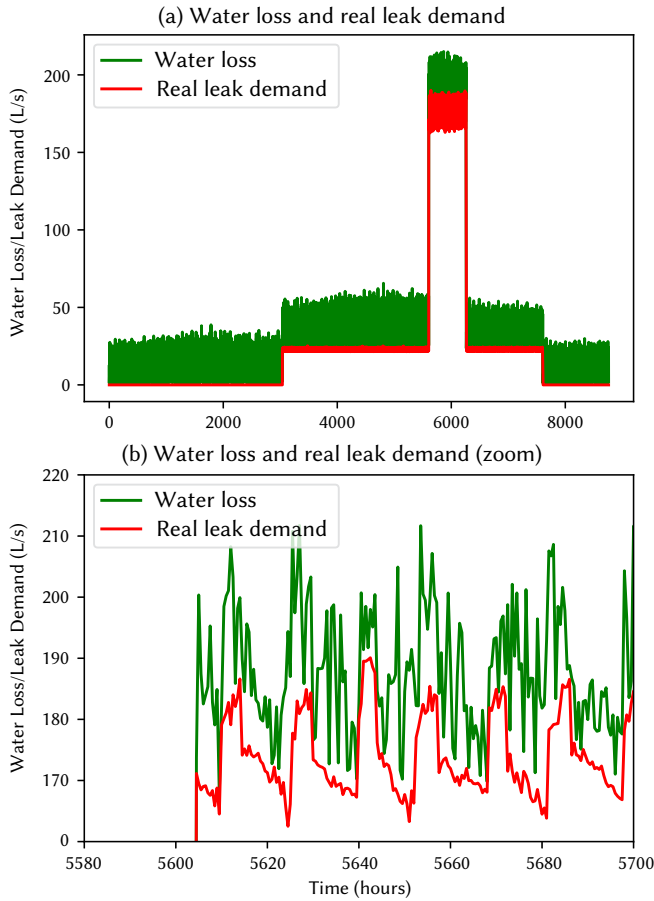


Fig. 10. Scenario 324: Water leak demand is more difficult to infer from water loss than in scenario 88, see Fig. 4.

be applied in two steps. First, leak node areas should be iteratively changed according to the generated sons, towards the combination with smaller flow residuals for System Model inlets over time until a convergence criterion was satisfied. In a second step, maintaining the leak node areas, node positions along the pipes would be worked out similarly in order to pinpoint leakages. The final step would be the insertion of the localized leak nodes in System Model, generating an alert message for maintenance staff. After repair, they should be excluded from System model, restoring the process back to continuous repetition of the algorithm.

VII. CONCLUSIONS, FUTURE WORKS AND DATA AVAILABILITY

Water demand AMR devices are underexplored in the literature about LDL, when compared to pressure sensors. Nonetheless, in this work, we presented them as a promising alternative for model-based LDL. Such an effort is fundamental to boost Smart Water implementation based on AMR devices, which shall be a growing research field in the near future⁶. As an example of that affirmation, we can cite the recent R\$3.8 billion (US\$ 731 million) investment established by the Brazilian public company SABESP to install around 4.4 million AMR devices in São Paulo and São José dos Campos⁷.

⁶ <https://www.bluefieldresearch.com/ns/global-smart-water-metering-market-hits-us6-8-billion-as-digital-transformation-accelerates/>

⁷ <https://www.enlit.world/library/brazil-s-sabesp-set-for-major-iot-smart-water-metering-project>

Scenario 324: Difference between real and estimated pressures

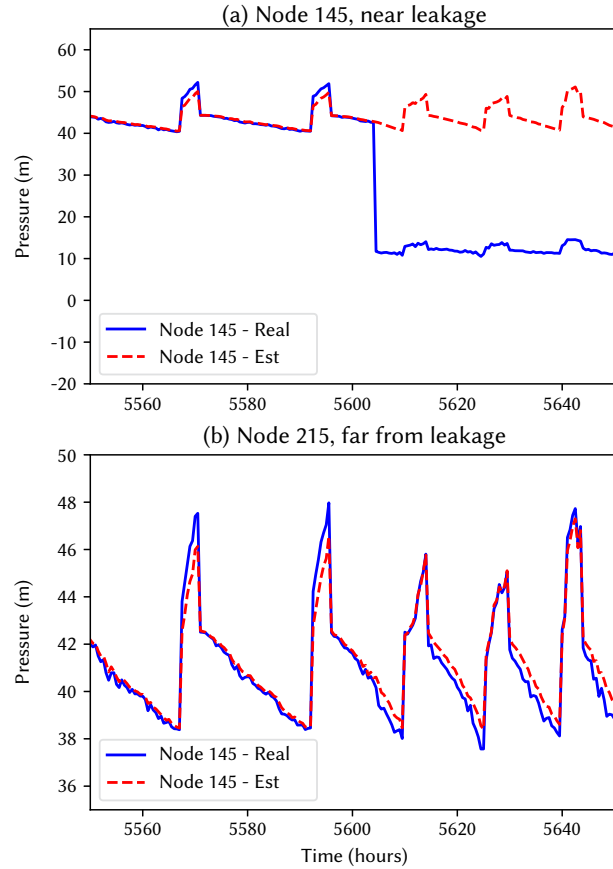


Fig. 11. Scenario 324: estimated node pressures near leak may have considerable difference to real pressures when leakage starts.

The proposed LDL approach is promising, however, several related limitations were addressed in Section VI. Therefore, we intend to work on the following future improvements:

- Generate a new dataset based on a bigger network, but limiting parameter uncertainties to reasonable ranges according to the literature;
- Include particular tanks for random customers to explore indirect pressure inference based on the consumption demand measurements, as a strategy to increase data spectrum using the same demand AMR devices;
- Apply a calibration process to System Model prior to LDL;
- Explore statistical methods to characterize unmeasured/fraud demands and to infer leak demand out of loss demand more precisely;
- Evaluate the use of Genetic Algorithms to address multiple and pre-existing leakages as discussed in the last paragraph of Section VI;
- Study about methods to localize unmeasured/fraud demands, under the proposed framework.

The approach presented in this work should be further considered in the literature. We, therefore, made LeakG3PD-OD and the LDL Python algorithms accessible in a GitHub repository⁸ for the academic community.

⁸ <https://github.com/matheuspilotto/LeakG3PD-OD>

CREDiT AUTHORSHIP CONTRIBUTION STATEMENT

Matheus Pilotto Figueiredo: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing (original draft), writing (review & editing).

Lizandro de Souza Oliveira: funding acquisition, methodology, project administration, supervision, writing (review & editing).

Giancarlo Lucca: funding acquisition, writing (review & editing).

Adenauer Corrêa Yamin: writing (review & editing).

Tiago da Rosa Lopes: funding acquisition.

DATA STATEMENT

Both LeakDB and LeakG3PD datasets and respective generator software, used as basis for this study, are available through their respective GitHub pages:

- <https://github.com/KIOS-Research/LeakDB> and
- <https://github.com/matheuspilotto/LeakG3PD>

DECLARATION OF CONFLICTS OF INTEREST

We have no conflict of interest to declare.

ACKNOWLEDGMENTS

The authors would like to thank the Catholic University of Pelotas (UCPEL), the Autonomous Sanitation Service of Pelotas (SANEP) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - FAPERGS (24/2551-00013962).

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES

- [1] M. Pilotto Figueiredo, L. de Souza Oliveira, G. Lucca, A. Correa Yamin, W. Huckembeck dos Santos, T. da Rosa Lopes, "LeakG3PD: A Python Generator and Simulated Water Distribution System Dataset," in *Intelligent Data Engineering and Automated Learning – IDEAL 2024*, Cham, 2025, pp. 71–82, Springer Nature Switzerland, doi: https://doi.org/10.1007/978-3-031-77738-7_7
- [2] L. Romero-Ben, D. Alves, J. Blesa, G. Cembrano, V. Puig, E. Duviella, "Leak detection and localization in water distribution networks: Review and perspective," *Annual Reviews in Control*, vol. 55, pp. 392–419, 2023, doi: <https://doi.org/10.1016/j.arcontrol.2023.03.012>
- [3] I. Daniel, J. Pesantez, S. Letzgus, M. A. Khaksar Fasaee, E. Berglund, G. Mahinthakumar, A. Cominola, "A Sequential Pressure-Based Algorithm for Data-Driven Leakage Identification and Model-Based Localization in Water Distribution Networks," *Journal of Water Resources Planning and Management*, vol. 148, 2022, doi: [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001535](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001535)
- [4] S. El-Zahab, T. Zayed, "Leak detection in water distribution networks: an introductory overview," *Smart Water*, vol. 4, no. 1, p. 5, 2019, doi: <https://doi.org/10.1186/s40713-019-0017-x>
- [5] J. H. Andersen, R. S. Powell, "Implicit state-estimation technique for water network monitoring," *Urban Water*, vol. 2, no. 2, pp. 123–130, 2000, doi: [https://doi.org/10.1016/S1462-0758\(00\)00050-9](https://doi.org/10.1016/S1462-0758(00)00050-9)
- [6] W. P. Cantos, I. Juran, S. Tinelli, "Machinelearning–based risk assessment method for leak detection and geolocation in a water distribution system," *Journal of Infrastructure Systems*, vol. 26, no. 1, p. 04019039, 2020, doi: [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000517](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000517)
- [7] A. Rajeswaran, S. Narasimhan, S. Narasimhan, "A graph partitioning algorithm for leak detection in water distribution networks," *Computers Chemical Engineering*, vol. 108, pp. 11–23, 2018, doi: <https://doi.org/10.1016/j.compchemeng.2017.08.007>
- [8] I. Marzola, F. Mazzoni, S. Alvisi, M. Franchini, "Leakage detection and localization in a water distribution network through comparison of observed and simulated pressure data," *Journal of Water Resources Planning and Management*, vol. 148, no. 1, p. 04021096, 2022, doi: [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001503](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001503)
- [9] D. B. Steffebauer, J. Deuerlein, D. Gilbert, E. Abraham, O. Piller, "Pressure-leak duality for leak detection and localization in water distribution systems," *Journal of Water Resources Planning and Management*, vol. 148, no. 3, p. 04021106, 2022, doi: [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001515](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001515)
- [10] E. Farah, A. Abdallah, I. Shahrour, "Sunrise: Large scale demonstrator of the smart water system," *International Journal of Sustainable Development and Planning*, vol. 12, no. 01, pp. 112–121, 2017, doi: <https://doi.org/10.2495/SDP-V12-N1-112-121>
- [11] M. Cheng, J. Li, "Optimal sensor placement for leak location in water distribution networks: A feature selection method combined with graph signal processing," *Water Research*, vol. 242, p. 120313, 2023, doi: <https://doi.org/10.1016/j.watres.2023.120313>
- [12] I. Santos-Ruiz, F.-R. López-Estrada, V. Puig, G. Valencia-Palomo, H.-R. Hernández, "Pressure Sensor Placement for Leak Localization in Water Distribution Networks Using Information Theory," *Sensors*, vol. 22, no. 2, 2022, doi: <https://doi.org/10.3390/s22020443>
- [13] S. Peng, J. Cheng, X. Wu, X. Fang, Q. Wu, "Pressure Sensor Placement in Water Supply Network Based on Graph Neural Network Clustering Method," *Water*, vol. 14, no. 2, 2022, doi: <https://doi.org/10.3390/w14020150>
- [14] L. M. Ramadhan, R. P. Astuti, H. Fakhurroja, "Compact Smart Water Meter Development for Smart City," in *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2023, pp. 677–682, <https://doi.org/10.1109/COMNETSAT59769.2023.10420528>
- [15] F. Martinelli, F. Mercaldo, A. Santone, "Water Meter Reading for Smart Grid Monitoring," *Sensors*, vol. 23, no. 1, 2023, doi: <https://doi.org/10.3390/s23010075>
- [16] M. Pilotto Figueiredo, L. de Souza Oliveira, "AquOculus: A Cost-effective Advanced Metering Infrastructure for Urban Water Distribution Systems," *IEEE Latin America Transactions*, vol. 24, no. 2, pp. 174–186, 2026, doi: <https://doi.org/10.1109/TLA.2026.11369472>
- [17] G. Yang, H. Wang, "Optimal Pressure Sensor Deployment for Leak Identification in Water Distribution Networks," *Sensors*, vol. 23, no. 12, 2023, doi: <https://doi.org/10.3390/s23125691>
- [18] M. Jentzsch, W. Lechner, R. Woszdlo, C. N. Nayeri, C. O. Paschereit, "The dynamic response of a pressure transducer for measurements in water," *Experiments in Fluids*, vol. 65, no. 8, 2024, doi: <https://doi.org/10.1007/s00348-024-03854-5>
- [19] A. V. Serafeim, N. Th. Fourniotis, R. Deidda, G. Kokosalakis, A. Langousis, "Leakages in Water Distribution Networks: Estimation Methods, Influential Factors, and Mitigation Strategies—A Comprehensive Review," *Water*, vol. 16, no. 11, 2024, doi: <https://doi.org/10.3390/w16111534>
- [20] T. AL-Washali, M. Mahardani, S. Sharma, F. Arregui, M. Kennedy, "Impact of float-valves on water meter performance under intermittent and continuous supply conditions," *Resources, Conservation and Recycling*, vol. 163, p. 105091, 2020, doi: <https://doi.org/10.1016/j.resconrec.2020.105091>
- [21] UNESCO and UNESCO i-WSSM, *Water security and the Sustainable Development Goals (Series I). Global Water Security Issues (GWSI) Series*. UNESCO Publishing, 2019. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000367904>
- [22] S. G. Vrachimis, M. S. Kyriakou, others, "LeakDB: a benchmark dataset for leakage diagnosis in water distribution networks:(146)," in *WDSA/CCWI joint conference proceedings*, 2018, doi: <https://doi.org/10.5281/zenodo.1313116>
- [23] L. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, T. Haxton, "EPANET 2.2 User Manual," 2020. [Online]. Available: <https://www.epa.gov/water-research/epanet>
- [24] K. A. Klise, M. Bynum, D. Moriarty, R. Murray, "A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study," *Environmental modelling & software*, vol. 95, pp. 420–431, 2017, doi: <https://doi.org/10.1016/j.envsoft.2017.06.022>
- [25] E. Protection Agency, "API documentation – WNTR 1.4.0 documentation,"

2026. [Online]. Available: <https://usepa.github.io/WNTR/wntr-api.html>
- [26] E. Protection Agency, "USEPA/WNTR," 2026. [Online]. Available: <https://github.com/USEPA/WNTR>
- [27] L. Rossman, "06 EPANET Net 3," 2016. [Online]. Available: https://uknowledge.uky.edu/wdst_manuals/2
- [28] P.N.M.S.M. SETH, *Hydraulics and Fluid Mechanics Including Hydraulics Machines*. Hydraulics and Fluid Mechanics Including Hydraulics Machines, Amit Publisher and Distributors, 2019.
- [29] M.T. Tsutiya, *Abastecimento de água*. Departamento de Engenharia Hidráulica e Sanitária da Escola Politécnica da Universidade de São Paulo, 2004.
- [30] K. K. Kuok, P. C. Chiu, D. C. M. Ting, "Evaluation of "c" values to head loss and water pressure due to pipe aging: Case study of uni-central Sarawak," *Journal of Water Resource and Protection*, vol. 12, no. 12, pp. 1077–1088, 2020, doi: <https://doi.org/10.4236/jwarp.2020.1212064>



Matheus Pilotto Figueiredo

He is an electrical engineer and holds a M.Sc. in Electronics and Computer Engineering at the Catholic University of Pelotas. His current research interests are focused on telemetry, sustainable management, and anomaly diagnostics in Water Distribution Systems using artificial intelligence.



Lizandro de Souza Oliveira

Received his Ph.D. and M.Sc. in Computer Science from the Federal University of Pelotas (UFPel), and the B.Sc. in Electrical Engineering. He is currently a professor and researcher at Catholic University of Pelotas (UCPel). His research interests include fuzzy consensus measures for hybrid memory management, embedded systems and telecommunications, particularly microstrip antennas and

numerical methods in electromagnetics.



Giancarlo Lucca

Received his Ph.D. from Public University of Navarra (UPNa), Spain. He is currently a professor and researcher at Federal University of Rio Grande (FURG), with research interests in fuzzy logic, aggregation functions, and fuzzy rule-based classification systems (among other areas). His research contributions include advancements in the Choquet integral, pre-aggregation functions, and their

applications in machine learning and decision-making. He actively engages in international collaborations and scientific events, contributing to the development of intelligent systems and computational intelligence techniques.



Adenauer Corrêa Yamin

Graduated in Electrical Engineering from the Catholic University of Pelotas, with a master's and a doctorate in Computer Science from the Federal University of Rio Grande do Sul. Currently serves at the Federal University of Pelotas (UFPel) as an advisor for the Graduate Program in Computing at the Center for Technological Development (PPGC/CDTec/UFPel). Actively involved in organizing

events and participating in Program Committees, the researcher contributes to projects and scientific production in the field of distributed systems, with a focus on high performance and ubiquity.



Tiago da Rosa Lopes

Holds a degree in Business Civil Engineering from the Federal University of Rio Grande (2009). Currently works as a Civil Engineer I at the Autonomous Water and Sewage Service of Pelotas (SANEP). Has experience in Civil Engineering, with a focus on Hydraulics.