Battle of the Leakage Detection and Isolation Methods

Stelios G. Vrachimis[1], Demetrios G. Eliades[2], Riccardo Taormina[3], Zoran Kapelan[4], Avi Ostfeld[5], Shuming Liu[6], Marios Kyriakou[7], Pavlos Pavlou[8], Mengning Qiu[9], and Marios M. Polycarpou[10]

[1]Postdoctoral Research Associate, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus; Email: vrachimis.stelios@ucy.ac.cy

[2]Research Assistant Professor, KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus; Email: eldemet@ucy.ac.cy

[3]Assistant Professor, Faculty of Civil Engineering and Geosciences, Department of Water Management, Delft University of Technology, Stevinweg 1, 2628 CN Delft, Netherlands; E-mail: r.taormina@tudelft.nl

[4]Professor, Department of Water Management, Delft University of Technology, Stevinweg 1, 2628 CN Delft, Netherlands; E-mail: z.kapelan@tudelft.nl

[5]Professor, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel; E-mail: ostfeld@tx.technion.ac.il

[6]Professor, School of Environment, Tsinghua University, Beijing 100084, China; Email: shumingliu@tsinghua.edu.cn

[7]Research Software Engineer, KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus; Email: kiriakou.marios@ucy.ac.cy

[8]Research Engineer, KIOS Research and Innovation Center of Excellence, University of Cyprus, Nicosia, Cyprus; Email: pavlou.v.pavlos@ucy.ac.cy

[9]Postdoctoral Research Fellow, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel; Email: mengning.qiu@campus.technion.ac.il

[10]Professor, KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus; Email: mpolycar@ucy.ac.cy

ABSTRACT

A key challenge in designing algorithms for leakage detection and isolation in drinking water distribution systems, is the performance evaluation and comparison between methodologies using benchmarks. For this purpose, the "Battle of the Leakage Detection and Isolation Methods" (BattLeDIM) competition was organized in 2020 with the aim to objectively compare the performance of methods for the detection and localization of leakage events, relying on SCADA measurements of flow and pressure sensors installed within a virtual water distribution system. Several teams from academia and the industry submitted their solutions, using various techniques including time-series analysis, statistical methods, machine learning, mathematical programming, meta-heuristics and engineering judgment, and were evaluated using realistic economic criteria. This paper summarizes the results of the competition and conducts an analysis of the different leakage detection and isolation methods used by the teams. The competition results highlight the need for further development of methods for leakage detection and isolation, and also the need to develop additional open benchmark problems for this purpose.

INTRODUCTION

Drinking Water Distribution Networks (DWDN) are susceptible to infrastructure failures, which may lead to water losses. The global average Non-Revenue Water (NRW) is 30%, with an estimated annual cost of $39 billion USD (Liemberger and Wyatt 2019). A significant part of NRW is due to background leakages and pipe bursts, which may occur anywhere within the distribution network. Background leakages are typically difficult to detect due to their small size, whereas pipe bursts are easier to locate as they are of larger size and may appear on the surface. The early detection and localization of any leakage event is crucial, as this reduces the time required for addressing the event and therefore reducing the risk of further infrastructure degradation, contamination events and consumer complaints.

Leakage diagnosis in water distribution systems has attracted a great deal of attention from both practitioners and researchers over the past years (Chan et al. 2018). The process of leakage diagnosis can be separated into: leakage detection, which focuses on identifying the existence of a leak in the network; and leakage localization, which aims to provide an approximate location of leakages given the available measurements. A recent review paper (Chan et al. 2018) classifies leakage detection methodologies into Passive and Active methods. Passive methods (also referred to as equipment-based, hardware or external methods) require the deployment of specialized equipment, such as acoustic sensors or ground penetrating radars, at areas that are suspect of leakage. Active methods (also referred to as internal or software methods) are methods that are based on the presence of permanently installed sensors which continuously monitor the system for leakages. The latest developments in hydraulic sensor technology and on-line data acquisition systems have enabled water companies to deploy a larger number of more accurate pressure and flow devices with less cost. These data can be used to monitor the system in real-time and develop methodologies that use the data to detect and pre-localize leaks using Active methods. Pre-localization is the process of defining an area in which the leak exists, instead of pin-pointing exactly its location. This research area has witnessed a significant interest, as indicated in recent review papers (Li et al. 2015; Chan et al. 2018; Zaman et al. 2019).

The term *model-based leakage diagnosis* is used to describe methodologies that utilize a model of the DWDN (also referred to as numerical model) and sensor measurements to estimate the steady-state hydraulic conditions in the network (Vrachimis et al. 2018b). The operating principle behind model-based leakage detection, as suggested by (Pudar and Liggett 1992), is to find discrepancies of measurements to their estimates obtained by the network model, which would indicate the existence of a leakage. Typically, model-based methods utilize a larger number of pressure sensors than flow sensors because they are cheaper and easier to install and maintain (Pérez et al. 2011). However, DWDN are large-scale systems and the number of sensors used in practice is still small compared to the system size. Moreover, to enhance leakage diagnosis, methodologies for optimal placement of pressure sensors are used (Farley et al. 2010; Casillas et al. 2013; Cuguero-Escofet et al. 2017). Finally, the consideration of measurement and model uncertainties is important when using these methods to determine if the network is operating in a normal state (Vrachimis et al. 2019) and should be taken into account before making a decision about the occurrence of a leakage in the network (Vrachimis et al. 2018a).

Leakage localization methods are typically model-based due to the limited information provided by the small number of sensors; one of the first representative examples is the work in (Wu et al. 2009) where the authors develop a model-based approach for leak localization which is applied to a large real system. Another interesting model-based approach applied on real systems is found in (Sophocleous et al. 2019), where the authors formulate an optimization problem to perform leakage diagnosis and deal with the dimentionality of the problem using search space reduction to reduce decision variables. Some approaches relate the acquired measurements with the simulated output from many simulated leakage scenarios on different locations of the network (Farley et al. 2010; Goulet et al. 2013); the geographical mapping of each model component can then be used to indicate the probability that a zone contains a leakage (Perez et al. 2014). Researchers have also used pressure residual analysis, by creating a system pressure sensitivity matrix to identify the location of leaks, based on the assumption of a single leakage occurring in the system (Pérez et al. 2011; Cuguero-Escofet et al. 2017). A more recent approach considers modeling uncertainties to create a set-bounded model of the system and then incorporates sensor measurements in an optimization-based framework to detect and pre-localize leakages using the concept of model-invalidation (Vrachimis et al. 2021).

Data-driven methods (also referred to as non-numerical model methods) do not require a model to perform detection. Leakage detection methodologies typically follow a data-driven approach; the authors in (Wu and He 2021) provide the latest review on this topic, and present a practical approach for anomaly event detection (including but not limited to leaks), classification and evaluation. Some approaches may require large amounts of reliable training data where the events are labeled by the operators or experts and they may perform poorly when data is not available (Li et al. 2015). An example of a data-driven approach is found in (Mounce et al. 2002) where the authors introduced artificial neural networks (ANNs) for burst detection and have continued to extend their work in the following years (Mounce et al. 2010). Another approach is found in (Eliades and Polycarpou 2012) where the authors proposed an algorithm which analyzes the discrete inflow signal of a District

Metered Area (DMA) by using an adaptive approximation methodology for updating the coefficients of a Fourier series and detects leakages by utilizing the Cumulative Sum (CUSUM) algorithm. The authors in (Soldevila et al. 2016) used a mixed model-based and data driven approach to improve performance. The study in (Wu and Liu 2017) provides a review on data-driven approaches for burst detection. The study concludes that these approaches are promising for use in real-life burst detection, however, reducing false alarms is still an important issue. Moreover, a comprehensive performance evaluation procedure, especially under different network configurations, might be necessary.

Leakage diagnosis methods are commonly evaluated on private commercial datasets (Chan et al. 2018), and as a result, it is not possible to objectively compare different methods in their ability to detect and isolate leaks. Moreover, data from real systems may not be readily available, while many aspects of the system operation are unknown. For example, information about the exact location, magnitude and time profile of leakages are typically unknown, but crucial when evaluating leakage diagnosis methodologies. The middle ground between evaluating algorithms on real systems and having all the available information about the system, is the development of a realistic simulation benchmark, built upon the expertise of practitioners, of which the operation resembles that of a real system. Recently, a benchmark leakage detection dataset has been developed named LeakDB (Vrachimis et al. 2018c), created using the WNTR tool (Klise et al. 2017). The dataset comprises of data generated from benchmark networks and uses pressure-driven demands and realistic leakage modelling (van Zyl et al. 2017). In this work, a realistic open benchmark for leakage detection and localization is developed and used in a "battle" (Taormina et al. 2018), to allow different teams to evaluate their methods in a unified way.

The *Battle of Leakage Detection and Isolation Methods* (BattLeDIM), was organized in 2020 initially as part of the CCWI/WDSA 2020 conference (the conference was postponed due to the COVID-19 pandemic). The competition aimed to objectively compare the performance of methods for the detection and localization of leakage events, relying on SCADA measurements of flow and pressure sensors generated using a realistic virtual city, which was based on a real water distribution network in Cyprus. The overall objective was to detect as many leakages as possible, as fast as possible and as close to the source as possible, while avoiding false alarms. Participants could use different types of tools and methods, including (but not limited to) engineering judgement, machine learning, statistical methods, signal processing, and model-based fault diagnosis approaches. In total, 18 teams from universities and industry around the world have submitted their solutions to the competition, and the results were presented on an online workshop organized on September 3, 2020.

The main contributions of this work are: 1) introduce a new benchmark network named "L-Town", developed for the purposes of the competition, along with a benchmark SCADA dataset; 2) provide an overview of the different leakage and isolation methodologies presented at the BattLeDIM competition and 3) analyze their results with respect to different objectives by proposing a comprehensive evaluation procedure.

THE L-TOWN BENCHMARK NETWORK

Below we introduce a new benchmark water distribution network, which we refer to as "L-Town". This is a city-scale model inspired by a coastal city in Cyprus, which can be used for research purposes. The network has been suitably modified and redesigned for security purposes. The L-Town is part of the *KIOS Virtual City Testbed*, an open software platform for simulating the SCADA operation of different critical infrastructures, including water, power, telecommunications and transportation systems.

Topology and structure

The L-Town model, depicted in Fig. 1, is represented using the EPANET input file format. It has 782 junctions and 905 pipe segments of approximately 50 meters length each and delivers drinking water to around 10,000 consumers and industries. It comprises of a network of steel pipes with a total length of 42.6 km and roughness coefficients (C values) between 120-140. The L-Town network has a *loop ratio* of 25%, a measure of complexity when solving the hydraulics of the network; it indicates that 25% of the pipes have to be removed in order to eliminate all loops from the network (Vrachimis et al. 2019). The node elevations range between 1.5 m and 75 m above the sea level.

The water distribution network of L-Town is receiving water from two reservoirs, and it has been designed to provide pressure head of at least 20m to all of its consumers. The normal operating pressure in the network ranges between 20-30 meters. A Pressure Reduction Valve (PRV) is installed at the lower part of the town ("Area B"), to help reduce background leakages. The network has different pressure areas, and therefore exhibits different sensitivity to leakages. PRVs are also installed downstream of the two main reservoirs, to help regulate the pressure. A pump and a water tank have been installed in the higher part of the town ("Area C"), to provide sufficient pressure to the consumers of that area. The tank has a diameter of 16 meters with a cylindrical shape. The pump has been programmed so that the tank refills during the night and empties to "Area C" during the day.

Note that the design decision to include pipes of 50 meters length is based on the following considerations: First, it is common for a real network to have consumer demand locations at a 50 meter interval, thus, in this sense, the provided benchmark can be considered a detailed version of a real network. Moreover, for the purposes of this competition, it is more efficient to allow participants to define a labeled pipe segment when localizing a leak, instead of defining a long pipe and the position of the leak on that pipe. Finally, the participating teams can apply model reduction techniques to reduce the complexity of the model and computational cost. This approach has the benefit of allowing teams to showcase the ability of their methodology to deal with complex network models. This would not have been possible if a reduced model of the benchmark network was already provided.

Water demand modelling

L-Town is assumed to be located in the Northern hemisphere, thus higher water usage is expected around July/August, and lower in December/January. No significant variations of water consumption is observed during holidays or other special days. During workdays (Monday to Friday), water consumption follows a similar pattern, whereas during the weekend (Saturday and

Sunday), there is higher consumption during late hours as the result of night life. Areas with industrial users do not follow the same pattern of consumption.

For constructing the benchmark model, open geospatial data were considered corresponding to the buildings of the actual location. A clustering algorithm was implemented in QGIS (QGIS.org 2020) to assign each building to a network node, and the node population was assigned to be proportional to the building area. This was computed using:

$$d_i^b = \sum_{j=1}^{n} (\alpha_i^j \beta_i^j) \gamma_i, \tag{1}$$

where d_i^b is the base demand of node i, n is the number of consumer types, α_i^j is the percentage of the j-th consumer type at the i-th node, β_i^j the average amount of water consumed in m^3/h for each m^2 of a building, and γ_i the total building area corresponding to node i. In this benchmark, three types of consumers ($n = 3$) were considered: *residential*, *commercial* and *industrial*.

Each node has a unique demand pattern for each consumer type, based on the statistical characteristics of real metered data from the area. Specifically, a Fourier Series model was used to approximate the demands (Vrachimis et al. 2018c), capturing seasonality (weekly, yearly) as well as the uncertainties on demand patterns (see Fig. 2). The overall water consumption is the linear combination of the base demands with the corresponding patterns.

The demand peaking factor, which is the ratio of the Maximum Daily Demand (MDD) to the Average Daily Demand (ADD) in a water system, was also considered in the design of demand patterns. The ratio, based on observations from real systems, typically ranges from 1.2 for very large water systems, to 3.0 or even higher for specific small systems. The demand peaking factor in L-Town ranges between 1.5 and 2.0, given it is an average size system.

THE BATTLEDIM CHALLENGE SCENARIO

As part of the "Battle of the Leakage Detection and Isolation Methods", all participating teams were given the following artificial scenario to establish the challenge:

"In previous years, the utility of L-Town was experiencing a large number of pipe breaks and water losses, affecting its service quality. During 2018, a number of leakage events occurred, which were detected and fixed by the water utility. However, it is believed that a number of smaller leakages occurred but not revealed. It is also assumed that some leakages occurred abruptly, whereas others developed gradually, as incipient events, from background leaks into pipe bursts.

To assist the L-Town water utility decision-making process, the utility developed an EPANET-based nominal model of the distribution network, in which base demands were assigned to nodes, following historical billing data of proximity consumers. Moreover, two nominal demand patterns were identified for residential and commercial consumer types (with some discrepancies). The utility believes that there might be some inaccuracies in the model, e.g., with respect to the pipe roughness and pipe diameters. In addition, the utility was not able to confirm the status of all the valves in the network (i.e., whether they are open or closed).

The L-Town water utility is searching for a solution to help them analyze the SCADA dataset, and detect leakage events as fast as possible. In addition, it is crucial for the utility to have an indication where approximately the leakage occurs, so that the field workers can inspect those potential leaks using their equipment.

The L-Town utility has created an open call for teams to demonstrate their ability in detecting and localizing leakage events. The teams are given a historical SCADA dataset along with information related with the leakages detected and fixed by the utility throughout 2018, to use for training purposes and for calibrating their models. It is possible that more leakage events occurred during 2018, however the utility was not able to detect and localize them.

Throughout 2019, the utility conducted periodic surveys using additional sensing equipment, pipe inspections and other methods, and was able to detect and isolate all the leakage events that occurred within that period. The most critical of these events were repaired, however it was not possible to repair some of these leakages due to financial reasons.

The overall goal of this competition, is to identify methods which are able to detect and localize the leakage events that occurred in L-Town in 2019, as fast as possible (with respect to time) and as accurately as possible (with respect to their location), in order to minimize their overall financial costs, both in water losses, as well as due to the hours spent in isolating the leakage by the utility staff. The L-Town utility will compare the different solutions and select the best one based on that objective."

SCENARIO GENERATION AND AVAILABLE DATA

To replicate the conditions of a real system, a SCADA dataset was synthetically generated using simulation, to correspond to sensor measurements from two (2) years of system "operation". For the generation of this SCADA dataset, a virtual testbed engine was designed in Python, released under the EUPL Open Source license (see Data Availability Section). This testbed uses the L-Town EPANET benchmark, and incorporates a number of assumptions with respect to the hydraulic solving, the leakage modelling, the modelling of uncertainty as well as the modelling of sensors.

Simulation and dataset generation engine

The dataset generation engine takes as input a structured file "dataset_configuration.yalm", which includes the start and end-time of the simulation, the leakages (including the start and end-time, the leak diameter, the type of the leakage and its peak time), the locations of the sensors (flow, pressure, AMRs and level sensors).

The hydraulic simulations are executed using the Water Network Tool for Resilience (WNTR), a Python package which supports pressure-driven demand simulations and leakage modelling (Klise et al. 2017). Specifically, for the pressure-driven demands, we compute a new demand for the i-th node $D_i(k)$, using the function f_{PDD}, such that $D_i(k) = f_{PDD}(p_i(k), d_i(k))$, where $p_i(k)$ is the pressure and $d_i(k)$ is the requested demand at node i: If the computed pressure is $p_i(k) < P_0$ then the demand is zero, i.e., $D_i(k) = 0$. If the pressure is $p_i(k) > P_f$, then the demand equals the requested demand, i.e., $D_i(k) = d_i(k)$. Finally, in the case where the pressure is $P_0 \leq p_i(k) \leq P_f$, then the demand is calculated as $D_i(k) = d_i(k)((p_i(k) - P_0)/(P_f - P_0))^{\delta}$. In BattLeDIM, we consider the following parameters: $P_0 = 7$, $P_f = 25$, $\delta = 0.5$. The values for P_f and δ are the default values used in WNTR, while the minimum pressure value $P_0 = 7$ was raised from 3.5 to 7 meters since this minimum value was never observed in the L-Town network during the considered scenarios.

Using the pressure dependent demand simulation, the node demand $D_i(k)$ starts to decrease compared to the requested demand $d_i(k)$ when the pressure is below P_f and goes to zero when pressure is below P_0.

Nominal and Real models

In practice, it is difficult to have an accurate model of the real system. For this reason, a "*nominal*" EPANET L-Town model was provided to the BattLeDIM participants, however a "*real*" model (which was unknown to the competitors) was used to generate the SCADA dataset. In general the nominal model approximates the real, with some uncertainties. The nominal model was generated by randomizing parameters of the real L-Town network, using the EPANET-MATLAB Toolkit (Eliades et al. 2016), as follows:

- Base demand of each consumer type at each node are randomized uniformly between ±10% compared to the 'real' value.
- Demand patterns: Nominal residential and commercial patterns are available, however industrial patterns are not available. The patterns used in the 'real' model are unique for each node and may differ significantly from the nominal patterns, while they also include a significant noise component.
- Pipe parameter uncertainty: All pipe parameters (roughness, length, and diameter) are randomized uniformly between ±10% of their 'real' value. This randomization aims to represent the uncertainty on hydraulic resistance, which is a function of all the aforementioned pipe parameters. We note that, in reality, parameter uncertainties may have different magnitudes. Typically, the most uncertain parameter is pipe roughness, while pipe length and diameter are less uncertain.
- Topological uncertainty: Two pipes ("p37" and "p251") were randomly selected to be closed in the real network, whereas in the nominal model they appeared to be open. The term "topological" uncertainty is used here to describe the variability of the topological graph of the network, due to a pipe valve with unknown status (open/closed). This can

also be considered as “operational” uncertainty since, typically, valves change status during operations, such as repairs, that have taken place in the network.

Sensors and Telemetry

We assume that there is one (1) tank water level sensor, a total of three (3) flow sensors, one at the pump and one at each of the DMA entrances, and 33 pressure sensors, all transmitting their measurements every 5 minutes to the utility’s *Supervisory Control and Data Acquisition* (SCADA) System. There are no time delays in the data transmission, and no lost packages. Pressure sensors give an average value of the last 5 minutes, which mitigates the uncertainty due to pressure transients in the system. In addition, 82 *Automated Metered Readings* (AMRs) have been installed in “Area C”, for delivering water consumption data directly to the SCADA system. Each AMR gives the aggregated consumption of many users in the AMR area.

The locations of the pressure sensors is depicted in Fig. 3, and the AMRs in Fig. 4. Sensor readings do not have errors, nor time-delays. The simulated sensor readings are rounded to 2 decimal points; in practice this reduces the amount of data sent over the telecommunications network.

Leakage modelling

We assume that the only faults affecting the system during the 2-year operation, are background leakages and pipe bursts. Any pre-existing leakages in the network are assumed to be small relative to individual node demands and have been incorporated into the pressure-dependent demands of the network. To model the leakage outflow in the i-th node, we assume the following general model (Lambert 2001; Greyvenstein and van Zyl 2007; Cassa et al. 2010):

$$l_i(k) = L(k)[p_i(k)]^{\zeta}, \tag{2}$$

where $L(k) = CA(k)\sqrt{2}\rho^{\zeta}$, for which the discharge coefficient for turbulent flow is $C = 0.75$, $A(k)$ is the area of the leak hole which may change in time, and ρ is the fluid density (for water we assume that $\rho = 1000kg/m^3$). For simplicity, we assume that the pipes in L-Town are made of steel, with roughness coefficients ranging between 120 and 140 (Hazen-Williams). Therefore, the exponent related to the characteristics of the leak, is assumed to be $\zeta = 0.5$.

A key aspect is the leakage magnitude and the time profile of the leakages. There are three (3) types of leaks in the system, categorized depending on their magnitude:

1. **Background leaks**: These are small leaks with size of 0-5% of the average inflow.
2. **Medium pipe-bursts**: Pipe breaks with flow size of 5-10% of the average inflow.
3. **Large pipe-bursts**: Pipe breaks with flow size above 10% of the average inflow.

In general, the average system inflow for the benchmark is around 180 m^3/h. The concept of background leaks is based on the categorization presented in (Lambert 1994); these are leakages that may exist in the system undetected for a long period of time. In the proposed benchmark, the smallest background leak was constrained at 2.5% of the average inflow, to enable their detection. The distinction between medium and large pipe-bursts is made assuming the latter are made visible and fixed more quickly by the water utility.

Moreover, the leak hole area $A(k)$ can be time-varying. In the case of abrupt leakage, the hole area is zero before the leakage start-time T_0, and becomes $\overline{A}$ after that time:

$$A(k) = \begin{cases} 0 & k < T_0 \\ \overline{A} & k \geq T_0 \end{cases} \tag{3}$$

In the case of incipient leak, we assume that the leak hole area $A(k)$ gradually increases after T_0, until it reaches $\overline{A}$ at time T_p:

$$A(k) = \begin{cases} 0 & k < T_0 \\ \overline{A}\left(\frac{k-T_0}{T_p-T_0}\right) & T_0 \leq k < T_p \\ \overline{A} & k \geq T_p \end{cases} \tag{4}$$

Regarding the leak time profile, the following assumptions were made: i) background leaks can exist from the beginning of the dataset and continue until the end, or they can start at any given time; ii) there are no large pipe bursts which have started before the simulation time; iii) background leaks can evolve into bursts (incipient leaks); e.g., a background leak which may have started as a small crack on a pipe may evolve into a large burst due to the stress applied on the pipe by pressure transients.

Leakage reporting

In practice, large leakages are easier to identify and fix, as they will be reported at some point by consumers or the utility staff. For the dataset leakages, we assume that large pipe-bursts are detected and fixed by the water utility, if they reach a flow magnitude larger than $\overline{l_j}$ at time T_l. The time of detection T_d is a time instance selected randomly during a maximum period of one (1) week after T_l. The repair time T_r is also defined as a time instance defined randomly, within one (1) week after T_d. After the leak is fixed, the area of the leak hole becomes zero, i.e., $A(k) = 0, t > T_r$. Specifically, large and some medium-size leakages (above 15 m^3/h) are fixed by the water utility after a reasonable time selected in random, with maximum delay of 2 months.

Leakage event simulation

All the leakage characteristics, were selected randomly, with certain constraints and assumptions:

- Based on the size of the network, statistically around 15 leakages (background and burst) events should appear each year in the network, with maximum 20 events. Eventually, we assume to have 14 events in the year 2018, and 19 events in the year 2019. Four (4) background leaks in the year 2018 continued in the year 2019. Only large pipe-bursts are detected and fixed by the water utility.
- We assume that at most 2 pipe bursts can coexist in the network during the examined periods. This is to enforce a wider spreading of the leaks during the year.
- We assume that a leakage can be detected by an L-Town staff using acoustic loggers, within 300 meters radius of its location. This is used in the evaluation of leakage isolation, and is based on actual feedback received by water utility operators from the original city considered

for the L-Town benchmark.

- We assume that in case leakages exist with overlapping detecting radius, there is a minimum 2 week different between their start time. This is to ensure separability of the alerts during the evaluation phase.

The final leakage locations for year 2018 and 2019 are found in the Fig. 5 and Fig. 6 respectively. The time profile of the leakages in 2019 is depicted in Fig. 7.

The BattLeDIM Datasets

The BattLeDIM datasets are composed of the following files, which are openly accessible via the Zenodo platform (see "Data Availability Statement" section) under the FAIR principles:

- **Configuration files**: The dataset configuration file indicates the simulation period as well as the characteristics of the 33 simulated leakages as part of BattLeDIM. It also specifies the sensors to be included in the SCADA datasets. (File format: YAML)
- **SCADA datasets**: These correspond to the SCADA measurements during the 2-year period between 2018-01-01 00:00 until 2019-12-31 23:55, at 5-minute time steps. The SCADA datasets are comprised of the water tank level, the flow sensors, the AMR measurements and the pressure sensors. (File format: CSV)
- **Leakages**: Table of times with respect to the leakage events of BattLeDIM, indicating their outflows in m^3/h. (File format: CSV)
- **Fixed Leakages reports**: This includes the repair times of pipe bursts that have been fixed in 2018 by the water utlity. (File format: TXT)
- **Network models**: Two network models are provided. i) The "*real*" model is the one used to generate the 2-year datasets, along with all the demand patterns. It contains the real network parameters and consumer demands. It does not contain any leakages. The real network should be considered as "unknown" ii) The "*nominal*" model should be used as the "known" model. This network is provided with nominal parameters for all the system elements. The nominal base demands for each node are based on average historical metered consumption.

Weekly demand profiles for three consumer types (residential, commercial and industrial) are also provided, however they do not capture the yearly seasonality. Furthermore, the EPANET model parameters may be different from the actual network parameters (e.g., diameters, roughness coefficients), and this difference is no greater than 10% of the nominal values. (File format: INP)

LIMITATIONS

The main challenge in developing effective leakage diagnosis algorithms is for them to be applicable in real systems and be able to deal with the problems arising from the scarcity and reliability of the data collected from the field. The aim of the proposed benchmark is to offer a realistic simulation scenario, built upon the expertise of practitioners, which closely resembles real conditions. It has the advantage that all the parameters and aspects of the system operation are known, and thus it can be used to compare and evaluate different methodologies. However, it has limitations and differences from real systems which are stated in this section to advice caution to researches and practitioners when using the benchmark.

The realistic demands included in this benchmark were generated by analyzing demands from real networks into their components and reproducing them by randomizing the components as described in (Vrachimis et al. 2018c). Real network demands may vary compared to the proposed approximations. Moreover, pressure-driven analysis is used to make the demands more realistic; however, we note that more research may be needed in selecting appropriate values for the pressure-driven analysis parameters.

A realistic leakage modeling approach was followed in this work by modeling pressure-dependent leakages on pipes, while the leakage function is constructed such as to exhibit time-variability with respect to the orifice size. However, the function describing leakage flow may vary in practice, because data collection about the size of leaks found in the field is a challenging task. More realistic leakage functions, than the one used in equation (2), have been proposed in recent literature (van Zyl et al. 2017; Kabaasha et al. 2020) and may be considered in future versions of this benchmark.

A decision was made in the creation of this benchmark to not include sensor time-delays and errors. This was taken consciously to avoid an extra dimension of complexity to a difficult competition problem, which includes large model uncertainties and small number of sensors compared to the system size. Moreover, we wanted the participants to focus on leakage diagnosis methodologies and not on methodologies for data validation. However, data acquired from real sensors may include significant errors and a number of measurements may need to be discarded and reconstructed. The real-time processing of data may be impeded by measurements arriving at later time-steps or never arriving at all.

This benchmark does not take into account events that may happen during and after repair works. Typically, repairs require the isolation of network sections by closing valves, an action that may cause pressure increase in the network. A typically observed phenomenon is the increase of leakage flows in other parts of the network during repairs or, in the worst cases, new pipe bursts. The risk of causing new leakages during repairs was not taken into account and should be considered when using this benchmark to test leakage diagnosis methodologies designed for application on real systems.

The reward for detecting leakages is based only on the value of water lost. However, the reward could be higher if indirect costs due to water losses were taken into account. The indirect costs include the acceleration of pipe deterioration, as well as third party damages. Such effects are usually accounted for in the cost of water, however they are difficult to quantify and were not considered in the benchmark.

COMPETING LEAKAGE DETECTION AND ISOLATION METHODS

In the following paragraphs we provide a short overview of the methodologies proposed by the competing teams.

The *Cheng00* team (Cheng et al. 2020) resorted to a three-stage approach involving simulation, ensemble multivariate change point detection (EMCPD), and statistical analysis. Pressure and flow residual time series are first obtained by comparing the SCADA datasets with those of simulated normal operation, produced with the provided benchmark model. The residuals are then analyzed with EMCPD to obtain a rough estimate of the occurrence of leak events in space and time. The final localization is performed after interpolating nodal pressures around likely candidate positions and by isolating the most likely sites with a two-sample one-sided Student's t-test.

The *DandW* team (Huang et al. 2020; Huang et al. 2022) proposed a methodology that treats each area of the L-Town network in Fig. 1 separately. This methods exploits the provided benchmark model to estimate expected sensor readings during normal operations and compute the residuals with respect to the provided SCADA data. Sensitivity vectors are then computed for each pipe as the Jacobian matrix of nodal pressures to pipe flows. The *angle method*, which involves calculating the angle between the residual vectors and the sensitivity vectors, is then used to isolate leaky pipes. These are characterized by having the smallest angles.

The *Leakbusters* team (Daniel et al. 2020; Daniel et al. 2022) tackled the challenge with a high-resolution pressure-driven method for leakage identification and localization composed of two sequential modules. In the first module, linear regression models are calibrated using data with no leaks to predict pairwise sensor pressure readings. When fed with new SCADA data, the reconstruction error between predicted and observed readings is tracked to identify the start time of a potential leak and the location of its nearest sensor. The second module uses the start time and most affected sensors reported by the first module to pinpoint leaky pipes relying on an initial set of candidate pipes and the application of a simulation-based optimization framework with iterative linear and mixed-integer linear programming.

The *CIACUA* team (Saldarriaga et al. 2020) approached the BattLeDIM problem by resorting to anomaly detection analysis and a simulation-optimization framework involving EPANET and Genetic Algorithms (GA). Anomaly detection analysis was first carried out by comparing SCADA data and the output of EPANET models. If the error between any observed and predicted signals passed a certain threshold, simulation-based optimization with GA was used to find which location would best explain such discrepancy, thus identifying the leaking pipe. Emitter equations were used to simulate leaks in the EPANET model.

The *Tsinghua* team (Wang et al. 2020; Wang et al. 2022) employed a hybrid approach where statistical methods are used in combination with hydraulic modelling. Their scheme comprises three stages. In the estimation stage Empirical Model Decomposition (EMD) and Vector Auto Regressive models are used to estimate expected flow and pressure in normal conditions. The residuals between these expected values and observed SCADA data are further processed in the identification stage to place leaks in time, and infer their size. In the final localization stage, leaking pipes are isolated by a double comparison between observed and simulated (EPANET) pressure data for the week with the suspected leak and the one preceding it.

The *Under Pressure* team (Steffelbauer et al. 2020; Steffelbauer et al. 2022) also employed a hierarchical approach made of 3 stages. Similar to the *Tsinghua* team, in the first stage demand calibration for the entire network was inferred from AMR data on Zone C using EMD. The authors also performed a calibration of the roughness coefficient using weighted least squares problem with bounded constraints. The second stage of *Under Pressure*'s approach entails the creation of a *dual* hydraulic model for leak detection. In this dual model, the pressure drops due to a leak translate into additional outflows to virtual reservoirs connected to the pressure measurement nodes. These time series, and the derived residuals, have a much better signal-to-noise ratio which facilitates detection and localization. This is done in the third stage, where leaks are first identified in time with the help of change detection methods (CUSUM, likelihood-ratio) and GA. The leaking pipe is then isolated based on the computation of Pearson correlation between residuals of virtual leak flows and pipe sensitivities, similar to what done by the *DandW* team.

Fuzzy methods are at the core of the *Zhiyun Shuiwu* team (Zhang et al. 2020). In the first stage, Deep Fuzzy Mapping is used to calibrate model demands from observations. Secondly, leaks are identified in time based on anomalies between observed and modeled pressure values and an analysis of the most affected nodes. Localization is finally performed based on fuzzy similarity between real bursts characteristics and pipe network characteristics.

The *IRI* team (Romero et al. 2020; Romero-Ben et al. 2022) devised a data-driven approach for Area A of L-Town due to the high density of pressure sensors. On the other hand, a model-based approach is used for both Area B and Area C to respectively overcome the lack of pressure sensors and exploit the availability of AMRs. In the data-driven approach, graph-based interpolation is first performed to estimate the state of the entire network from available data of leaky and non-leaky scenarios. The selection of candidate leak location is then performed by nodal pressure comparison between these estimated states. In the model-based approach, EPANET simulations are carried out after inferring the demands for Area B and Area C. The results of the simulations with leaks added at different locations are compared against the SCADA data to find the most likely placement for the leak.

The *KU Hydrosystems* team (Min et al. 2020) proposed a two-stage method where leak identification in time and space is tackled separately using a data-driven and a model-based approach. After pre-processing the data and performing feature selection, the detection of the leak in time is performed jointly by resorting to k-means clustering. Leak locations are then identified via a comparison between real data and the output of multiple simulations using a calibrated EPANET model accounting for leaks (with emitter coefficients). The initial calibration is performed with the Harmony Search algorithm in order to find optimal values of roughness coefficients and nodal demands.

InfraSense Labs (Blocher et al. 2020) devised a method involving three main steps. Firstly, the daily demand profiles are partitioned into clusters using the k-means algorithm. The clusters correspond to days with similar flow patterns so that variations in the derived clusters can be used to identify changes in demand that may be attributed to leaks. Leaks are then detected by comparing the difference between expected demands (derived from flow profiles of five preceding days based on cluster membership) and observed flows. If the residuals indicate the presence of a leak, hotspots are localized by solving a regularized inverse problem that includes a pressure-driven model for the leak flow.

DHI China (Liu et al. 2020) proposed a method that relies on genetic algorithms and Machine Learning (ML) techniques. GA is used to calibrate the provided nominal model, whose demand patterns are defined based on the analysis of the provided AMR data. Leak detection in time is done with the use of both Deep Learning methods (an LSTM neural network) and gradient boosted trees (LightGBM). GA-based simulation-optimization (with EPANET) is employed to localize the leaky pipe, similar to what done by other teams.

The Multiple Leaks Detection and Isolation Framework (MLDIF) proposed by the *Tongji* team (Li and Xin 2020) consists of three stages as "calibration-identification-localization". First, a model calibration stage is performed to get a calibrated hydraulic model using a time-period where little or no leakages are assumed to exist. Any pre-existing leakages in the selected time-period are incorporated into the calibrated model, which is then used to estimate the overall yearly leakage flows and to predict nodal pressures under a 'leak-free' scenario. Then, the pressure residuals between observed and predicted pressure are processed by integrating STL decomposition method and the K-means clustering method to identify different leak scenarios during the analysis period. Finally, by adding no-repaired but identified leaks to the calibrated hydraulic model in the localization stage, a new and simple leakage scenario is reconstructed to facilitate leakage localization. Therefore, the pipe with the highest probability of leakage can be isolated by a step-wise method based on matching degrees between the actual leakage feature and the simulated leakage features.

The *Wu BSY* team (Wu and He 2020) presented an integrated data analysis with hydraulics-based modeling approach consisting of three main steps: i) *data pre-processing* to prepare for analysis, where flow and pressure time-series are decomposed to get rid of trend and seasonality using the Seasonal-Trend decomposition procedure ; ii) *data analysis* for leakage event detection, where the decomposed time-series are analyzed using Statistical Process Control methods; and iii) *model analysis*, where simulation-based optimization in *Bentley WaterGEMS*, a hydraulic model calibration tool, is used to localize the leaky pipes using a pressure-driven approach where the emitter coefficients and locations are the parameters to be optimized.

The *CUBALYTICS* team (Bhowmick and Seifert 2020) also devised an approach combining data-driven methods with hydraulic simulations. This method is based on the computation of an *anomaly matrix* (AM) for leak detection and localization. This matrix is created by first applying statistical methods to identify anomalies in the Master Data Set, i.e., the overall table having timestamps as indexes and sensor readings as columns. The AM is a binary matrix (1 = anomaly detected), obtained from the previous operation after keeping only the rows for which there is at least an anomaly. Leaks are identified in time by analyzing contiguous rows in the AM having multiple anomalies. The list of nodes, i.e., the headers of all columns with non-zero entries, is checked to find valid node combinations identifying potential leaky pipes. The isolated pipe for each leak is selected after comparison with pressure-driven simulations.

Decision trees are at the core of the methodology of the *Artesia* team (Adanza Dopazo 2020). The approach consists of three main steps. In the first step, data normalization and feature engineering is performed to extract minimum and maximum daily peaks, as well as averages for different parts of the day for all pressure, water level and flow sensors. Decision trees are then trained on this refined dataset to predict the mean night pressure values expected for each pressure sensor. The mean pressure during the night is chosen as the target to predict since pressure during this time of the day is more steady and less affected by randomness. In the last stage, the differences between predicted and observed mean night pressure values in the test dataset are used to identify leaks in time, while comparison of results across neighboring pressure sensors is used to improve localization.

The *DHI Singapore* team (Tan et al. 2020) employed WNTR, a Python wrapper of EPANET, to generate extra data for training a deep neural network (DNN) using Tensorflow. Before generating the leak events, the team calibrated the provided nominal model to find optimal values of pipe diameters, roughness coefficient, as well as determining optimal seasonality of residential and commercial demands. Calibration was performed using GA and the 2018 pressure readings. The DNN development dataset is generated from 400 simulations with random leaks at different locations, with different start time and duration. A five hidden layer DNN is trained on this data to isolate the leaky location having as inputs the readings from the 33 pressure sensors. After its validation, the DNN is tested on the competition dataset.

The *UNIFE* team (Marzola et al. 2020; Marzola et al. 2022) adopted a pragmatic approach to detect and localize leakage events, based on the analysis of the SCADA data and the use of the provided hydraulic model of the network. After inferring demand patterns for the entire network based on the provided AMR data, the hydraulic model is calibrated (roughness and diameters) to realistically represent the hydraulic behaviour of the network. The observed inflows and water demands are then analysed to identify leakage number, entity and time of occurrence with engineering judgment. Each identified leakage is then spatially localised through an enumerative procedure. This is done by i) performing simulation after assigning the leakage to each pipe of the network in turn, ii) assessing the error in terms of differences between observed and simulated pressures, and iii) selecting the pipe characterized by the lowest error.

The *FluIng* team (Barros et al. 2020) resorted to a mixed approach using signal processing for leak identification, and simulation-based optimization for leak localization. The first phase of leak identification entails the use of blind source separation to decompose each measured flow time series into a main signal, primarily related to water consumption, and a "noisy" signal in which leak events are more visible. Change detection is then performed on this noisy component to detect leaks in time. Localization of leaky pipes is then carried out with a two-steps approach based on Particle Swarm Optimization where i) the provided nominal model is first calibrated in an *offline* fashion, and ii) leak locations are inferred via iterative *online* fine tuning of nodal demands.

Analysis of methodologies

Table 1 summarizes the key elements of each method, highlighting similarities and differences between them. The general features which are listed in Table 1 and their use as part of the different methodological approaches is described in Table 2.

In general, the solutions proposed may be comprised of one or more of the following parts: the detection procedure, the localization approach, and the calibration method. Each methodology utilized various tools in order to solve each problem. For example, some model-based approaches relied on the use of nominal water network models provided (such as the EPANET L-Town model). To accommodate the differences between the measurements and the nominal model, calibration methods were used to design a more accurate representation, by updating the demands and certain pipe parameters. The calibrated model can be used to create datasets describing the operation of the system under normal and faulty operation conditions, e.g., using the EPANET libraries. This can allow the comparison of the computed pressure residuals with the observed pressure sensor measurements.

Another approach is to consider the mathematical model of the system, to create a *pressure sensitivity matrix*, through a linearization of the hydraulic equations. Using the above, residuals can be computed, using model-based approaches which compare simulation-based estimations and SCADA measurements, as well as by using model-free approaches. The residuals, as well as other relevant time-series, can be analyzed using change detection techniques (CUSUM, angle method etc.), time series analysis and signal processing, empirical method decomposition, regression analysis, hypothesis testing and other statistical approaches. More advanced statistical approaches, such as machine learning, and computational intelligence methods based on fuzzy systems, have also been proposed.

A subset of methodologies considers optimization formulations, which may rely on simulations to evaluate the objective functions, or on explicit mathematical formulations which can be solved using Integer/Dynamic/Mixed Integer Programming. Where this is not possible due to the complexity of the optimization formulation, meta-heuristics (such as genetic algorithms or particle swarm optimization) can be used. Finally, it's important to note that some approaches analyzed the AMR-area in a different way, by creating a model of the water demands in the area, to exploit the additional information provided due to the significant penetration of the smart meters.

EVALUATION PROCEDURE

Participants were required to submit their results in the format specified in a template file, which includes the location and start time of each detected leakage event. The start time of a leakage is specified in the ISO 8601 time format YYYY-MM-DD hh:mm. The location of the leakage is specified by the link ID, as defined in the EPANET model of the network "L-Town.inp". Participants were allowed to specify any number of leaks.

Competition evaluation criteria

Evaluation of participant results follows a pure economic approach. The water utility of L-Town calculates the profit from water saved in a single year from successful detections. The utility also considers the cost of the repair crew every time it is sent to search for a leakage.

A correct detection is one that points at a link ID which is inside a predefined pipe length radius around the leak location, and the given leakage start time is during the lifetime of the same leakage. The predefined pipe radius is defined by the capability of the close range equipment used by the repair crew (e.g. acoustic sensors) to exactly pinpoint the location of the leakage in a single workday.

The scoring methodology is described here in detail. Given a user defined set of detections $\mathcal{D}$ and the set of leakages $\mathcal{L}$ (2019 BattLeDIM dataset), the total score S is calculated using the following rules:

1. **True detection (True Positive):** A given detection $i \in \mathcal{D}$ is considered a True Detection of a leakage $j \in \mathcal{L}$ if the detection time t_d^i and the distance $x_{ij} \geq 0$ from the center of the isolated link to the leak location, satisfy the following conditions:

$$t_{st}^j \leq t_d^i \leq t_{end}^j, \tag{5a}$$

$$x_{ij} \leq x_{max}, \tag{5b}$$

   where t_{st}^j and t_{end}^j are the start and end time of leakage j respectively, and x_{max} is the predefined pipe length radius around the leak location.

2. **False detection (False Positive):** False detections are the detections which do not satisfy the True detection condition above.
3. **Missed detection (False Negative):** Missed detections are the set of leakages in $\mathcal{L}$ which have not been detected by any detection in $\mathcal{D}$ (includes 4 leakages starting in 2018 and 19 leakages starting in 2019).
4. **Order of evaluation:** Detections in $\mathcal{D}$ are evaluated in chronological order, i.e., from the earliest detection to the latest detection, against all leakages in $\mathcal{L}$. Note that detections given by participants which are outside the year 2019 are ignored.
5. **Repeated detections:** Once a leak is detected, it is added to the list $\mathcal{L}_{\lceil}$. Successful

detections of leaks in $\mathcal{L}_{\lceil}$ are given a score of zero (0), i.e., repeated detections of the same leakage are ignored.

6. **Multiple detections:** A single detection may detect only one leakage, even if more than one leakage is in the detection area. Note that detection of multiple leakages is limited due to the leakage placement algorithm used to create the dataset. In the case of the existence of multiple leakages in the detection radius of detection i, e.g., leakage $j \in \{1, \ldots, m\}$, only the leakage closest to the detected link is considered to be discovered. The discovered leakage $l \in \mathcal{L}$ in the case of multiple true detections is given by:

$$l = \left\{ j : x_{ij} = \min \left(x_{ij}, j \in \{1, \ldots, m\} \right) \right\} \tag{6}$$

7. **Profit from water saved:** The profit p_w^i (euro) from water saved by detection i, for a detected leakage j, is calculated as follows:

$$p_w^i = \left(\sum_{k=t_d^i}^{t_{end}^j} q^j(k) \Delta t \right) c_w, \tag{7}$$

where by detection i, $q^j(k)$ is the flow rate of leakage j at each discrete time step k, Δt is the duration of the discrete time step and c_w is the cost (euro) of water per cubic meter.

8. **Repair crew cost:** All detections in $\mathcal{D}$ are associated with a utility repair crew cost. The repair crew checks for leakages only within a predefined radius of x_{max} from the given location. The repair crew cost for a given detection i is assumed to be proportional to the distance x_{ij} from the leakage j and is calculated as follows:

$$c_r^i = \begin{cases} -\left(\frac{x_{ij}}{x_{max}} \right) c_r, & x_{ij} < x_{max} \\ -c_r, & x_{ij} \geq x_{max} \end{cases} \tag{8}$$

where c_r^i is the repair crew cost for detection i, and c_r is the maximum repair crew cost for a given leakage search assignment.

9. **Total score:** The total score S for a given set of detections $\mathcal{D}$ is given by:

$$S = \sum_{i \in \mathcal{D}} s^i = \sum_{i \in \mathcal{D}} \left(p_w^i + c_r^i \right), \tag{9}$$

where s^i is the score per given detection i.

The parameters of maximum detection radius x_{max}, cost of water per cubic meter in euro c_w and the maximum repair crew cost c_r are given in Table 3. The cost of water is selected assuming a water utility which operates in Cyprus. The maximum repair crew cost is calculated assuming a three-person repair crew searching for the leakage location for a whole 8-hour workday, with an hourly rate of approximately 20 euro per hour. The maximum detection radius is selected assuming the repair crew is able to search using acoustic sensors a maximum pipe length of 1 km in a single workday. In order for this distance to be translated into a radius, an average of three pipe branches emerging around any given location is assumed. The maximum score in this problem, given the parameters of Table 3 and the leakages existing in the dataset, is achieved when all leakages are detected at their exact start time and location, while no false detections are given. The "perfect" score of the competition was calculated using equation (9) to be €523,124.

For illustration purposes, an example of the evaluation function is shown in Fig. 8, where all possible values of the detection score are plotted for detecting a leakage with constant flow of $q(k) = 100\ m^3/h$. The evaluation parameters were arbitrarily chosen as follows: cost of water $c_w = 1$ euro$/m^3$, max crew cost $c_r = 500$ euro/detection and max detection distance $x_{max} = 50m$.

Alternative evaluation criteria

The evaluation methodology used in this competition has some disadvantages which arise from using a score which is proportional to the amount of water saved from each successfully detected leakage. Specifically, the current methodology favors the detection of large and abrupt leakages as well as leakages which start early in the dataset.

To avoid this issue, an alternative evaluation approach is demonstrated which takes into account the total volume of water lost from each leakage, given in Fig. 9. The volumes are derived by calculating the area under the leakage flow curves of Fig. 7. It can be observed from Fig. 9 that each leakage will be rewarded differently since the reward for each detection directly relates to the water volume loss of each leakage.

This alternative evaluation approach alters the reward function of (7) which calculates the profit from each detected leakage, by normalizing the profit by the volume of the corresponding leakage. Specifically, given detection i which successfully detects leakage j, the profit from water saved (euro) is calculated as follows:

$$p_w^i = \frac{v_s^j}{v^j} v_m c_w \tag{10}$$

where v_s^j is the volume of water saved given detection i, v^j is the total volume of water loss from leakage j, and v_m is the mean volume of water loss of all leakages in the dataset. The mean leakage volume v_m is calculated for this dataset to be $v_m = 28432\ m^3$.

Notice that using the normalized reward function, the maximum reward for each detected leakage is $v_m c_w$. The most obvious drawback of this alternative evaluation approach is that the Economic score loses its literal meaning.

COMPETITION RESULTS AND DISCUSSION

Team rankings are defined by calculating the *Economic* score of the results submitted by each team. The Economic score of each team is given in Fig. 10(a), where the names of the teams have been substituted by generic labels, specifically the letters A–R. It is interesting to note that the Economic score does not necessarily reflect the ranking when the True Positive Rate (TPR) and False Positives (FP) of each submitted result is considered. The TPR and FP of each team are illustrated in Fig. 10(b) and Fig. 10(c) respectively.

The winning teams of the BattLeDIM competition, were the 6 teams with the highest economic score and with the highest true-positive rate. The name of these teams are provided in Table 4, along with their Pareto ranking. For instance, "Tongji-Team" and "Under Pressure" are non-dominated solutions and are ranked to the first Pareto front with an economic score of €264,873 and €260,562, and a True Positive Rate of 56.52% and 65.22%, respectively. The "Perfect" score of the competition was €523,124 (no time delay in detection, no false positives, exact position), which implies that the best solutions in BattLeDIM achieved a score around 50%.

Evaluation parameter sensitivity analysis and alternative criteria results

The sensitivity of the total score to the cost of water per cubic meter in euro c_w is evaluated here in order to analyze the effect that different assumptions on cost may have on the ranking of solutions provided. The cost of water affects the Economic score the most since this is proportional to the amount of water lost from leakages, while it does not affect the number of True Positives or False Positives achieved by each team. Five different water prices were used to re-evaluate the competition results ranging from 0.40 €$/m^3$ to 1.20 €$/m^3$.

The sensitivity analysis results are illustrated in Fig. 12. The results indicate that the increasing water price favors teams which had a larger number of False Positives and of which the Economic score was affected due to the cost of sending out repair crews. This result draws the conclusion that, given a difficult challenge such as the BattLeDIM problem, the cost of water should be taken into account when deciding how conservative a leakage diagnosis methodology should be. Another interesting observation is that the first five teams do not change rank with the increasing water price since they outperform the rest of the methodologies in the TPR metric.

Moreover, the results using the alternative evaluation criteria described in Section 7 are shown in Fig. 13. It can be observed that the normalized score rankings follow more closely the rankings of the True Positive Rates, except in the cases where the corresponding teams have a high number of False Positive detections.

Discussion

The BattLeDIM competition provides valuable insights on the state-of-the-art in leakage detection and isolation methods, their limitations as well the different ways that the results should be evaluated. For instance, by analyzing the methodological approaches followed by the top teams, as shown in Table 1, it is apparent that different approaches have been used by the teams, and the robustness of each approach to different evaluation functions may vary. Some of the observations are discussed below:

- Most top-scoring teams make use of a nominal model, of which the parameters are calibrated in some form using sensor data, to construct a water distribution model which describes the normal operation of the system (such as Tsinghua, Under Pressure, IRI and UNIFE), by incorporating existing leakages into the calibrated node demands. This allows the computation of the expected flows and pressures at different locations in the network. Moreover, they also consider the AMR measurements separately from the rest of the network, and use them to estimate/calibrate demands.
- For the detection of events, model-based residuals along with some form of a change detection algorithm (e.g., Leak-Busters, UNIFE, Under Pressure) or time-series/signal processing (e.g., Tongji) analysis was preferred by most of the top-scoring teams. Some of these residuals were also utilized for localization purposes (e.g., IRI, Tsinghua).
- For the leak isolation, top-scoring teams used some form of optimization framework, to identify the most likely leakage point (e.g., Leak-Busters, Tsinghua, IRI and Tongji).
- Some solutions, had a high True Positive Rate, but with a significantly higher number of False Positives (210) with respect to the other participants (such as team 'E' in Fig. 10). Based on the BattLeDIM assumptions for the cost of water and staff cost, this solution received a low score. However, sensitivity analysis of the result indicates that, for higher cost of water, this solution could have received a higher rank. This indicates that it may be beneficial to accept higher number of false positives, if the cost of water lost is significantly higher than the staff cost.

CONCLUSIONS AND OPEN CHALLENGES

In this paper we presented the results from the "Battle of the Leakage Detection and Isolation Methods" (BattLeDIM), an open competition which aimed to objectively compare different methodologies in their ability of detecting and isolating leakage events within a virtual water distribution system. For the purposes of this work, a new benchmark network was introduced, "L-Town", based on a realistic water distribution system. Moreover, a synthetic 2-year SCADA benchmark dataset was generated with leakages of various types and magnitudes, which can be used by the research community to develop leakage diagnosis methodologies, keeping in mind the limitations of this benchmark mentioned in Section 5. An economic objective metric was defined to evaluate the different solutions, considering realistic operational costs. In total, 18 teams from the academia and the industry participated in the BattLeDIM competition. The teams used various methodologies, including model-based and model-free approaches, simulation and optimization tools, machine learning and others; these techniques are summarized in Table 2. We presented the evaluation methodology and discussed its limitations.

Overall, the competition demonstrated that multiple technologies could be used for solving the problem and that there is potential for significant improvement, since the top solutions achieved 50% of the maximum possible score. However, it is important to make a distinction between the 'maximum possible score' and the 'maximum feasible score' in this problem: the former is the score achieved when all leakages are detected perfectly without false positives, while the latter is the maximum score that can be achieved by any methodology given the limited information provided about the problem. The methodology to calculate the maximum feasible score for the BattLeDIM benchmark is an open research question. Since the goal of this benchmark is to recreate, as realistically as possible, a real-world problem, the development of such methodology will be useful in determining the conditions that should exist in real systems to make it at least theoretically feasible to achieve a certain performance in leakage diagnosis. Many factors are in play that affect the maximum feasible score, such as the selected water network, the size of leakages and the magnitude of the considered uncertainty. Moreover, it is safe to say that the maximum feasible score will change by varying some parameters of the BattLeDIM problem to make it even more realistic; for example, including sensor noise and missing measurements in the dataset.

In closing, the BattLeDIM competition demonstrated the need for open benchmarks, which can assist the research community towards reproducibility and open science.

DATA AVAILABILITY STATEMENT

All data, models, or code generated or used during the study are available in a repository online in accordance with the FAIR data retention policies, under the European Union Public License

(EUPL) v1.2:

- Dataset generation and scoring algorithm: https://github.com/KIOS-Research/BattLeDIM
- SCADA Dataset: https://zenodo.org/record/4017659
- Reproducible code: https://codeocean.com/capsule/2366240/tree/v1

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the ERC Synergy Grant 'Water-Futures' (Grant agreement No. 951424); by the European Union Horizon 2020 programme Grant Agreement No. 739551 (KIOS CoE), and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy; by the Interreg V-A Greece-Cyprus 2014-2020 program, co-financed by the European Union (ERDF) and National Funds of Greece and Cyprus under project "SmartWater2020"; by the WaterAnalytics Project ENTERPRISES/0916/23 which is co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation; and by the Deutsche Forschungsgemeinschaft (DFG).

REFERENCES

Adanza Dopazo, D. (2020). "A leakage detection system extracting the most meaningful features with decision trees.." *Zenodo*, <10.5281/zenodo.3906668> (June).

Barros, D., Resende, P., Brentan, B., Meirelles, G., Montalvo, I., Luvizotto, E., and Izquierdo, J. (2020). "Signal processing and optimization process for leakage detection and localization." *Zenodo*, <10.5281/zenodo.4011894> (September).

Bhowmick, S. and Seifert, K. (2020). "Water Leakage Detection and Localization: Anomaly Matrix - A Deterministic Approach." *Zenodo*, <10.5281/zenodo.3906850> (June).

Blocher, C., Pecci, F., Jara Arriagada, C., and Stoianov, I. (2020). "Detecting and localizing leakage hotspots in water distribution networks via regularization of an inverse problem: An application to the battle of leakage detection and isolation methods 2020 competition." *Zenodo*, <10.5281/zenodo.3921800> (June).

Casillas, M., Puig, V., Garza-castañón, L., and Rosich, A. (2013). “Optimal sensor placement for leak location in water distribution networks using genetic algorithms.” *Sensors*, 13(11), 14984–15005.

Cassa, A. M., van Zyl, J. E., and Laubscher, R. F. (2010). “A numerical investigation into the effect of pressure on holes and cracks in water supply pipes.” *Urban Water Journal*, 7(2), 109–120.

Chan, T. K., Chin, C. S., and Zhong, X. (2018). “Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection.” *IEEE Access*, 6, 78846–78867.

Cheng, T., Li, Y., Harrou, F., Sun, Y., Gao, J., and Leiknes, T. (2020). “A hybrid leakage detection and isolation approach based on ensemble multivariate changepoint detection methods.” *Zenodo*, <10.5281/zenodo.3964167> (June).

Cuguero-Escofet, M. A., Puig, V., and Quevedo, J. (2017). “Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: Application to the Barcelona water network.” *Control Engineering Practice*, 63, 1–12.

Daniel, I., Pesantez, J., Letzgus, S., Khaksar Fasaee, M. A., Alghamdi, F., Berglund, E., Mahinthakumar, G., and Cominola, A. (2022). “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*, 148(6), 04022025.

Daniel, I., Pesantez, J., Letzgus, S., Khaksar Fasaee, M. A., Alghamdi, F., Mahinthakumar, K., Berglund, E., and Cominola, A. (2020). “A high-resolution pressure-driven method for leakage identification and localization in water distribution networks.” *Zenodo*, <10.5281/zenodo.3924632> (June).

Eliades, D. G., Kyriakou, M., Vrachimis, S., and Polycarpou, M. M. (2016). “Epanet-matlab toolkit: An open-source software for interfacing epanet with matlab.” *Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI)*, The Netherlands, 8 (Nov).

Eliades, D. G. and Polycarpou, M. M. (2012). “Leakage fault detection in district metered areas of water distribution systems." *Journal of Hydroinformatics*, 14(4), 992–1005.

Farley, B., Mounce, S. R., and Boxall, J. B. (2010). "Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network." *Urban Water Journal*, 7(6), 345–356.

Goulet, J.-a., Coutu, S., and Smith, I. F. (2013). "Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks." *Advanced Engineering Informatics*, 27(2), 261–269.

Greyvenstein, B. and van Zyl, J. E. (2007). "An experimental investigation into the pressure - leakage relationship of some failed water pipes." *Journal of Water Supply: Research and Technology-Aqua*, 56(2), 117–124.

Huang, L., Du, K., Guan, M., Huang, W., Song, Z., and Wang, Q. (2022). "Combined Usage of Hydraulic Model Calibration Residuals and Improved Vector Angle Method for Burst Detection and Localization in Water Distribution Systems." *Journal of Water Resources Planning and Management*, 148(7), 04022034.

Huang, L., Du, K., Guan, M., and Wang, Q. (2020). "The combined usage of the hydraulic model calibration residual and an improved vectorial angle method for solving the BattLeDIM problem." *Zenodo*, <10.5281/zenodo.3925507> (June).

Kabaasha, A. M., van Zyl, J. E., and Mahinthakumar, G. K. (2020). "Correcting Power Leakage Equation for Improved Leakage Modeling and Detection." *Journal of Water Resources Planning and Management*, 146(3), 06020001.

Klise, K. A., Bynum, M., Moriarty, D., and Murray, R. (2017). "A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study." *Environmental Modelling & Software*, 95, 420–431.

Lambert, A. (1994). "Accounting for Losses: The Bursts and Background Concept." *Water and Environment Journal*, 8(2), 205–214.

Lambert, A. (2001). "What do we know about pressure: Leakage relationships in distribution systems?." *Proc. IWA Syst. Approach to Leakage Control Water Distrib. Syst. Manag.*, 8.

Li, R., Huang, H., Xin, K., and Tao, T. (2015). “A review of methods for burst/leakage detection and location in water distribution systems.” *Water Science and Technology: Water Supply*, 15(3), 429–441.

Li, Z. and Xin, K. (2020). “Fast localization of multiple leaks in water distribution network jointly driven by simulation and machine learning.” *Zenodo*, <10.5281/zenodo.3911045> (June).

Liemberger, R. and Wyatt, A. (2019). “Quantifying the global non-revenue water problem.” *Water Supply*, 19(3), 831–837.

Liu, R., Zhang, Z., and Zhang, D. (2020). “Leakage detection and isolation in water distribution network based on data mining and genetic optimized hydraulic simulation.” *Zenodo*, <10.5281/zenodo.3911523> (June).

Marzola, I., Mazzoni, F., Alvisi, S., and Franchini, M. (2020). “A pragmatic approach for leakage detection based on the analysis of observed data and hydraulic simulations.” *Zenodo*, <10.5281/zenodo.3900989> (June).

Marzola, I., Mazzoni, F., Alvisi, S., and Franchini, M. (2022). “Leakage Detection and Localization in a Water Distribution Network through Comparison of Observed and Simulated Pressure Data.” *Journal of Water Resources Planning and Management*, 148(1), 04021096.

Min, K. W., Kim, T., Choi, Y. H., Jung, D., and Kim, J. H. (2020). “A two-phase model to detect and localize water distribution system leakages.” *Zenodo*, <10.5281/zenodo.3922019> (June).

Mounce, S., Day, A., Wood, A., Khan, A., Widdop, P., and Machell, J. (2002). “A neural network approach to burst detection.” *Water Science and Technology*, 45(4-5), 237–246.

Mounce, S. R., Boxall, J. B., and Machell, J. (2010). “Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows.” *Journal of Water Resources Planning and Management*, 136(3), 309–318.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., and Peralta, A. (2011). “Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks.” *Control Engineering Practice*, 19(10), 1157–1167.

Perez, R., Sanz, G., Puig, V., Quevedo, J., Cuguero Escofet, M. A., Nejjari, F., Meseguer, J., Cembrano, G., Mirats Tur, J. M., and Sarrate, R. (2014). “Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in barcelona.” *IEEE Control Systems*, 34(4), 24–36.

Pudar, R. S. and Liggett, J. A. (1992). “Leaks in pipe networks.” *Journal of Hydraulic Engineering*, 118(7), 1031–1046.

QGIS.org (2020). “QGIS Geographic Information System, <http://www.qgis.org>.

Romero, L., Blesa, J., Alves, D., Cembrano, G., Puig, V., and Duviella, E. (2020). “Leak localization in water distribution networks using data-driven and model-based approaches.” *Zenodo*, <10.5281/zenodo.3923501> (June).

Romero-Ben, L., Alves, D., Blesa, J., Cembrano, G., Puig, V., and 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.

Saldarriaga, J., Solarte, L., Salcedo, C., Montes, C., Martínez, L., González, M., Cuello, M., Ariza, A., Galindo, C., Ortiz, N., Gómez, C., and Vanegas, S. (2020). “Battle of the leakage detection and isolation methods: An energy method analysis using genetic algorithms.” *Zenodo*, <10.5281/zenodo.3923227> (June).

Soldevila, A., Blesa, J., Tornil-sin, S., Duviella, E., Fernandez-canti, R. M., and Puig, V. (2016). “Leak localization in water distribution networks using a mixed model-based/data-driven approach.” *Control Engineering Practice*, 55, 162–173.

Sophocleous, S., Savić, D., and Kapelan, Z. (2019). “Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction.” *Journal of Water Resources Planning and Management*, 145(7), 04019024.

Steffelbauer, D. B., Deuerlein, J., Gilbert, D., Abraham, E., and Piller, O. (2022). “Pressure-Leak Duality for Leak Detection and Localization in Water Distribution Systems.” *Journal of Water Resources Planning and Management*, 148(3), 04021106.

Steffelbauer, D. B., Deuerlein, J., Gilbert, D., Piller, O., and Abraham, E. (2020). “A dual model for leak detection and localization.” *Zenodo*, <10.5281/zenodo.3923907> (June).

Tan, C. A., Phipatanasuphorn, V., and Lai, C. H. A. (2020). "Deep learning of complex pipe leakages events in drinking water distribution networks for effective spatiotemporal pre-detections and isolations of leak conditions." *Zenodo*, <10.5281/zenodo.3902945> (June).

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., Ostfeld, A., Eliades, D. G., Aghashahi, M., Sundararajan, R., Pourahmadi, M., Banks, M. K., Brentan, B. M., Campbell, E., Lima, G., Manzi, D., Ayala-cabrera, D., Herrera, M., Montalvo, I., Izquierdo, J., Luvizotto, E., Chandy, S. E., Rasekh, A., Barker, Z. A., Campbell, B., Shafiee, M. E., Giacomoni, M., Gatsis, N., Taha, A., Abokifa, A. A., Haddad, K., Lo, C. S., Biswas, P., Pasha, M. F. K., Kc, B., Somasundaram, S. L., Housh, M., and Ohar, Z. (2018). "Battle of the Attack Detection Algorithms: Disclosing Cyber Attacks on Water Distribution Networks." *Journal of Water Resources Planning and Management*, 144(8), 04018048.

van Zyl, J. E., Lambert, A. O., Collins, R., Zyl, J. E. V., Ph, D., Asce, M., Lambert, A. O., Collins, R., and Ph, D. (2017). "Realistic Modeling of Leakage and Intrusion Flows through Leak Openings in Pipes." *Journal of Hydraulic Engineering*, 143(9), 04017030.

Vrachimis, S. G., Eliades, D. G., and Polycarpou, M. M. (2018a). "Leak detection in water distribution systems using hydraulic interval state estimation." *2018 IEEE Conference on Control Technology and Applications (CCTA)*, IEEE, 565–570 (aug).

Vrachimis, S. G., Eliades, D. G., and Polycarpou, M. M. (2018b). "Real-time hydraulic interval state estimation for water transport networks: a case study." *Drinking Water Engineering and Science*, 11(1), 19–24.

Vrachimis, S. G., Kyriakou, M. S., Eliades, D. G., and Polycarpou, M. M. (2018c). "LeakDB: A benchmark dataset for leakage diagnosis in water distribution networks Description of Benchmark." *WDSA / CCWI Joint Conference Proceedings*, Vol. 1 (7).

Vrachimis, S. G., Timotheou, S., Eliades, D. G., and Polycarpou, M. M. (2019). "Iterative hydraulic interval state estimation for water distribution networks." *Journal of Water Resources Planning and Management*, 145(1), 04018087.

Vrachimis, S. G., Timotheou, S., Eliades, D. G., and Polycarpou, M. M. (2021). "Leakage detection and localization in water distribution systems: A model invalidation approach." *Control Engineering Practice*, 110, 104755.

Wang, X., Li, J., Liu, S., Yu, X., and Ma, Z. (2022). "Multiple Leakage Detection and Isolation in District Metering Areas Using a Multistage Approach." *Journal of Water Resources Planning and Management*, 148(6), 04022021.

Wang, X., Li, J., Yu, X., Ma, Z., and Huang, Y. (2020). "A multistage approach to detect and isolate multiple leakages in district metering areas in water distribution systems." *Zenodo*, <10.5281/zenodo.3924109> (June).

Wu, Y. and Liu, S. (2017). "A review of data-driven approaches for burst detection in water distribution systems." *Urban Water Journal*, 14(9), 972–983.

Wu, Z. Y. and He, Y. (2020). "Decomposition-based data analysis with hydraulic model calibration for leakage detection and isolation." *Zenodo*, <10.5281/zenodo.3908525> (June).

Wu, Z. Y. and He, Y. (2021). "Time Series Data Decomposition-Based Anomaly Detection and Evaluation Framework for Operational Management of Smart Water Grid." *Journal of Water Resources Planning and Management*, 147(9), 04021059.

Wu, Z. Y., Sage, P., and Turtle, D. (2009). "Pressure-Dependent Leak Detection Model and Its Application to a District Water System." *Journal of Water Resources Planning and Management*, 136(1), 116–128.

Zaman, D., Tiwari, M. K., Gupta, A. K., and Sen, D. (2019). "A review of leakage detection strategies for pressurised pipeline in steady-state." *Engineering Failure Analysis*, 104264.

Zhang, w., Liu, j., Han, l., Li, y., Li, x., Shi, z., Yu, j., and Wang, j. (2020). "A real time method to detect the burst location of urban water supply network." *Zenodo*, <10.5281/zenodo.3923929> (June).

List of Tables

1 Summary of the different approaches used by the competing teams, at each stage of their proposed leakage diagnosis methodologies (D: used during detection, L: used during localization, C: used for calibration, Y: used in general). 40

2 Explanation of features included in the methodologies of the competing teams. . . 41

3 Parameters used in the evaluation procedure. 42

4 BattLeDIM competition results and ranking of top 6 participating teams. 43

FEATURE	Cheng00	DandW	Leak-busters	CIA-CUA	Tsing-hua	Un-der Pres-sure	Zhiyun Shuiwu	IRI	KU Hydro-systems	Infra-Sense Labs	DHI China	Tongji	Wu BSY	Cuba-lytics	Arte-sia	DHI Singa-pore	UNIFE	Flu-Ing
Use nominal model	D	D		D	D	DL	DL	L	L			L		L		D	CDL	CL
Model calibration					Y	Y		Y	Y			Y				Y	Y	Y
AMR based demands					C	C		C			Y	Y					Y	
Normal operation dataset and/or dual model	D	D		D	D	DL					Y	Y					D	
Areas treated differently	Y	Y						Y				Y					Y	
Pressure Sensitivity Matrix		DL				L												
Pressure reconstruction/comparison	L							DL							L			
Residuals - Model-based	D	DL		D	DL		DL		L	L		DL					D	
Residuals - Model-free			D					DL		D					D			
Change Detection	D	L	D	D	DL	D	D	L					D				D	D
Time Series Analysis/Signal Processing/EMD					DL	DL						D	D					D
Statistical Methods	L		D											D				D
Machine Learning and Soft Computing					D						D				D	DL		
Simulation-based optimization			L	L	L			L			L		L			C	CL	CL
Simulating leaks			L	L	L		L	L	L			L	L	L		C		
Mathematical Programming			L							L		C						
Meta-heuristics				L		D	DL		C		CL					C		CL
Ad-hoc/Engineering judgement														DL	D		DL	

TABLE 1. Summary of the different approaches used by the competing teams, at each stage of their proposed leakage diagnosis methodologies (D: used during detection, L: used during localization, C: used for calibration, Y: used in general).

FEATURE	DESCRIPTION
Use nominal model	Making use of the provided EPANET model for L-Town
Model calibration	Nominal model calibration of demands and/or pipe parameters
AMR based demands	Use of AMR data to model demand patterns
Normal operation dataset and/or dual model	Use of a (calibrated) EPANET model to create dataset under normal operations (no leak) and/or a normal operation model
Areas treated differently	Whether the algorithms treat different areas of the network separately
Pressure Sensitivity Matrix	Linearization of hydraulic equations
Pressure reconstruction/comparison	Reconstruction/comparison of pressure of neighboring nodes
Residuals - Model-based	Residuals computed between simulated readings from available nominal model simulations and observed SCADA
Residuals - Model-free	Residuals computed between predicted readings from model-free approach and observed SCADA
Change Detection	Technique to identify abrupt change in residuals/observations in time (CUSUM, angle method)
Time Series Analysis/Signal Processing/EMD	Methods pertaining to TSA/SP such as Empirical Model Decomposition, spectral methods used at different stages of the algorithm
Statistical Methods	Methods based on comparison with statistical distribution of the observed data, hypothesis testing, linear regression, etc.
Machine Learning and Soft Computing	Includes supervised/unsupervised machine learning (also feature engineering), fuzzy methods
Simulation-based optimization	Use of an optimization method with objective function based on simulation via hydraulic model
Simulating leaks	Use of an EPANET model to simulate leaks
Mathematical Programming	Methods including Integer Programming, Dynamic Programming, Mixed Integer Programming
Meta-heuristics	Global optimization methods such as Genetic Algorithms, Harmony Search and Particle Swarm Optimization
Ad-hoc/Engineering judgement	Techniques that cannot be framed in the methods above or methods based on engineering common sense

TABLE 2. Explanation of features included in the methodologies of the competing teams.

Parameter	Value	Description
x_{max}	300 (Meters)	Maximum detection radius
c_w	0.80 (Euro)	Cost of water per m^3
c_r	500 (Euro)	Maximum repair crew cost

TABLE 3. Parameters used in the evaluation procedure.

Team Name (Label)	Pareto Rank	True Positive Rate	False Positives Count	Economic Score (Euro)
Tongji-Team (L)	1	56.52%	3	€264,873
Under Pressure (O)	1	65.22%	4	€260,562
IRI (H)	2	43.47%	1	€210,772
Leakbusters (K)	2	47.83%	7	€195,490
Tsinghua (M)	3	47.83%	5	€167,981
UNIFE (N)	4	43.47%	4	€127,626
PERFECT	-	100%	0	€523,154

TABLE 4. BattLeDIM competition results and ranking of top 6 participating teams.

List of Figures

1 The L-Town Benchmark Network. 45

2 Demand signal decomposition using Fourier Series. 46

3 Location of pressure sensors in the L-Town network. 47

4 Location of AMRs (nodes with red colour) in “Area C” of the L-Town network. . . 48

5 Location of leakages in 2018 dataset. 49

6 Location of leakages in 2019 dataset. 50

7 Evolution of leakages in 2019 dataset. 51

8 Example of the scoring function for a true detection: $q(k) = 100\ m^3/h$ (leakage flow), $c_w = 1$ euro/m^3 (water cost), $c_r = 500$ euro/detection (max crew cost) 52

9 Total volume of water lost from each leakage in the BattLeDIM problem, sorted chronologically and identified by the corresponding link ID. 53

10 (a) Final scores of the BattLeDIM competition: Team rankings are based only on the *Economic score*. The ‘Perfect’ score is the theoretical upper bound; (b) Team scores with respect to the *True Positive Rate* metric; (c) Team scores with respect to the number of *False Positives*. 54

11 Multi-parameter score (Economic score and True Positive Rate) of the submitted results. The best scores are in the upper-right corner of the graph. 55

12 Sensitivity analysis of the Economic score with respect to the price of water: (a) 0.40, (b) 0.60, (c) 0.80, (d) 1.00, (e) 1.20 Euro. Note that the True Positive Rates (TPR) and number of False Positives (FP) remain the same in these scenarios. . . . 56

13 (a) Alternative *Economic Score* and ranking of teams in the BattLeDIM competition using the alternative evaluation criteria in which the leakage volume is normalized; (b) *True Positive Rate* score; (c) Number of False positives. 57

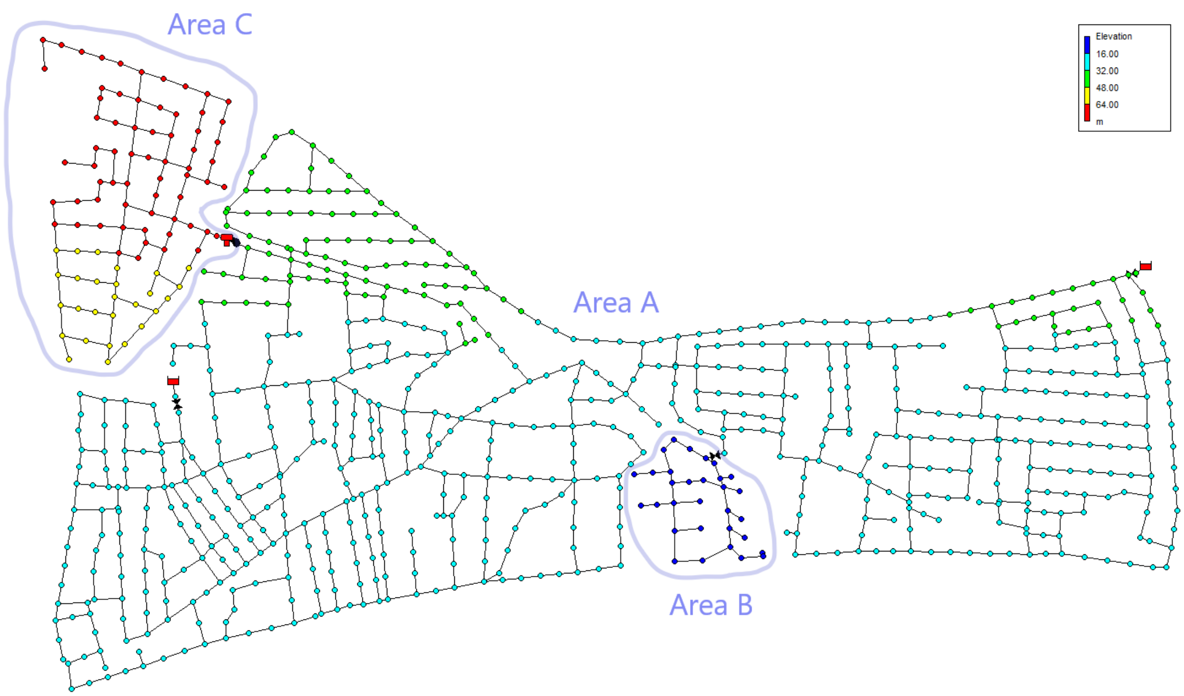

Fig. 1. The L-Town Benchmark Network.

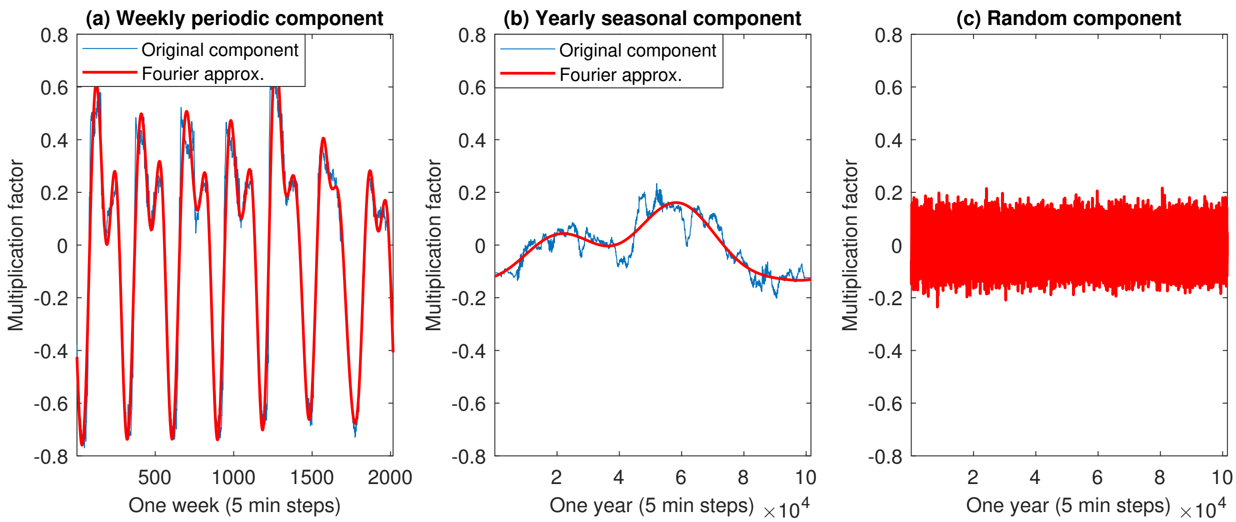

Fig. 2. Demand signal decomposition using Fourier Series.

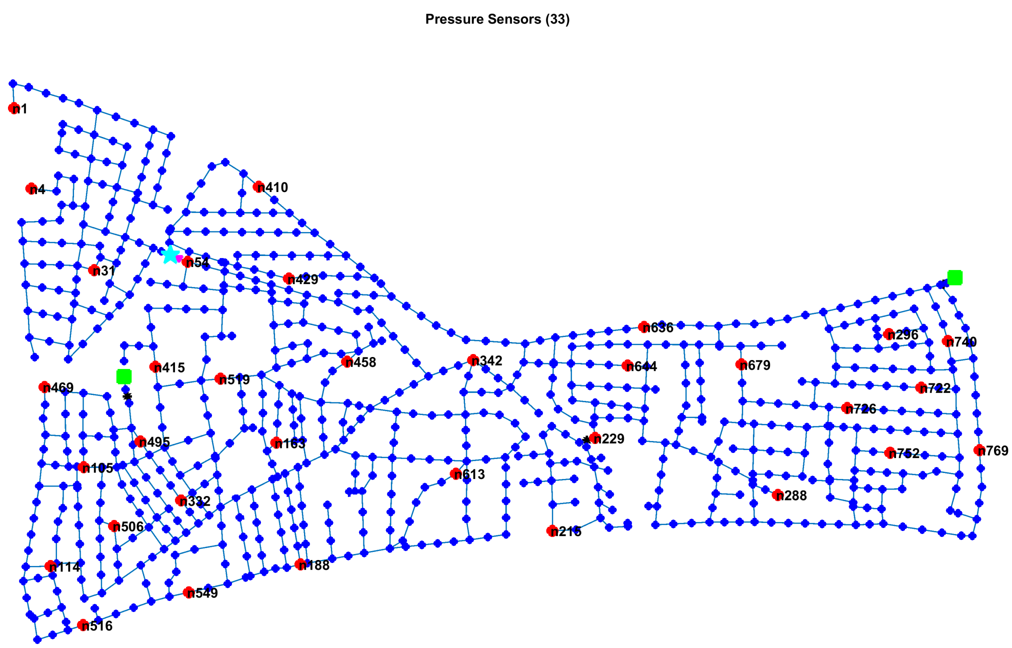

Fig. 3. Location of pressure sensors in the L-Town network.

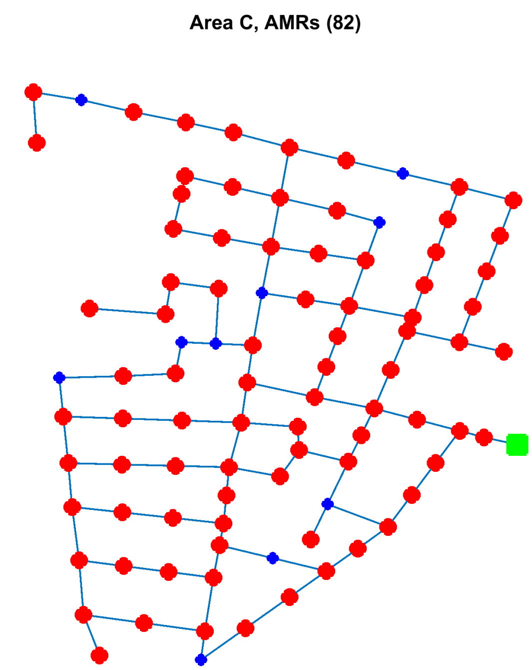

Fig. 4. Location of AMRs (nodes with red colour) in “Area C” of the L-Town network.

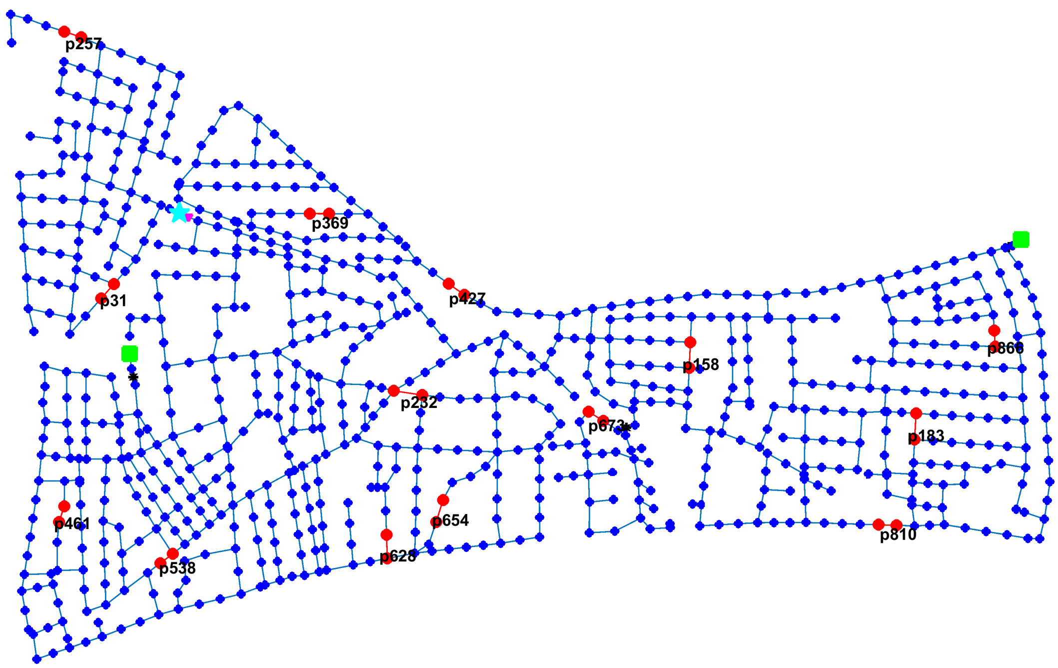

Fig. 5. Location of leakages in 2018 dataset.

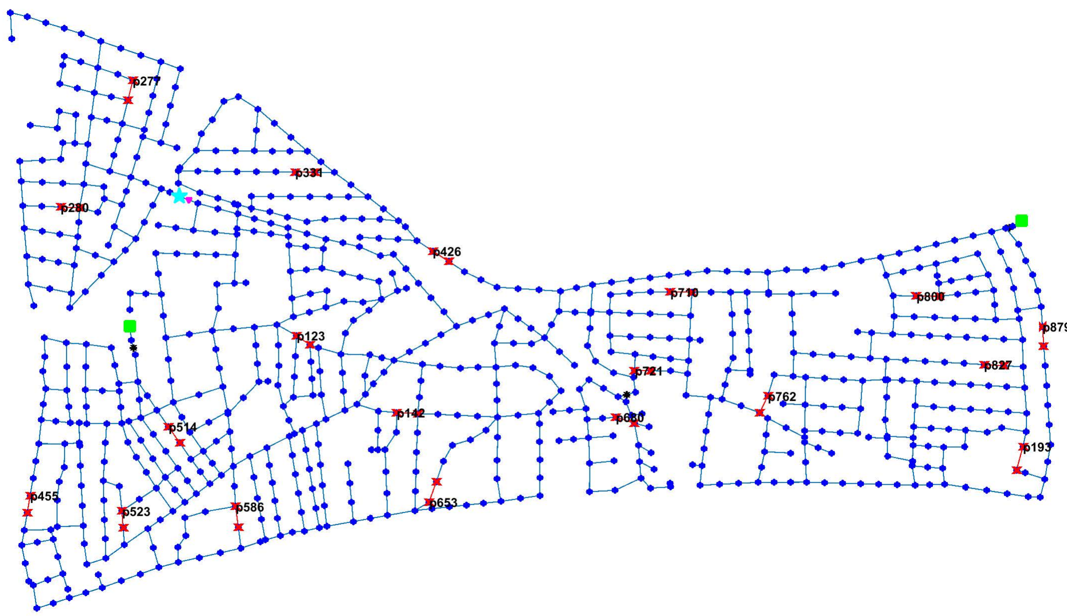

Fig. 6. Location of leakages in 2019 dataset.

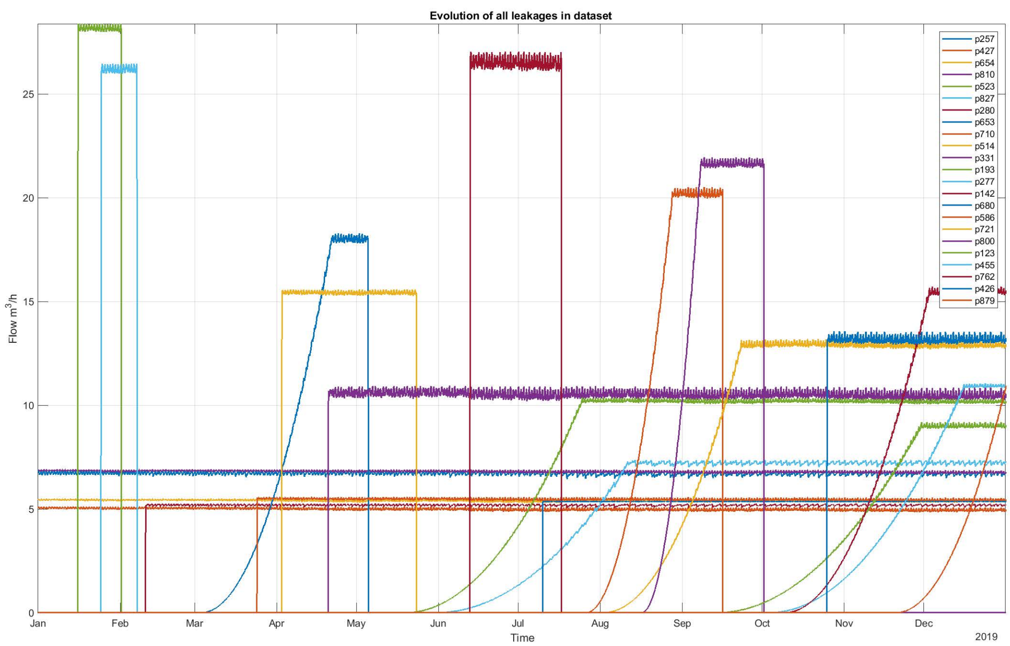

Fig. 7. Evolution of leakages in 2019 dataset.

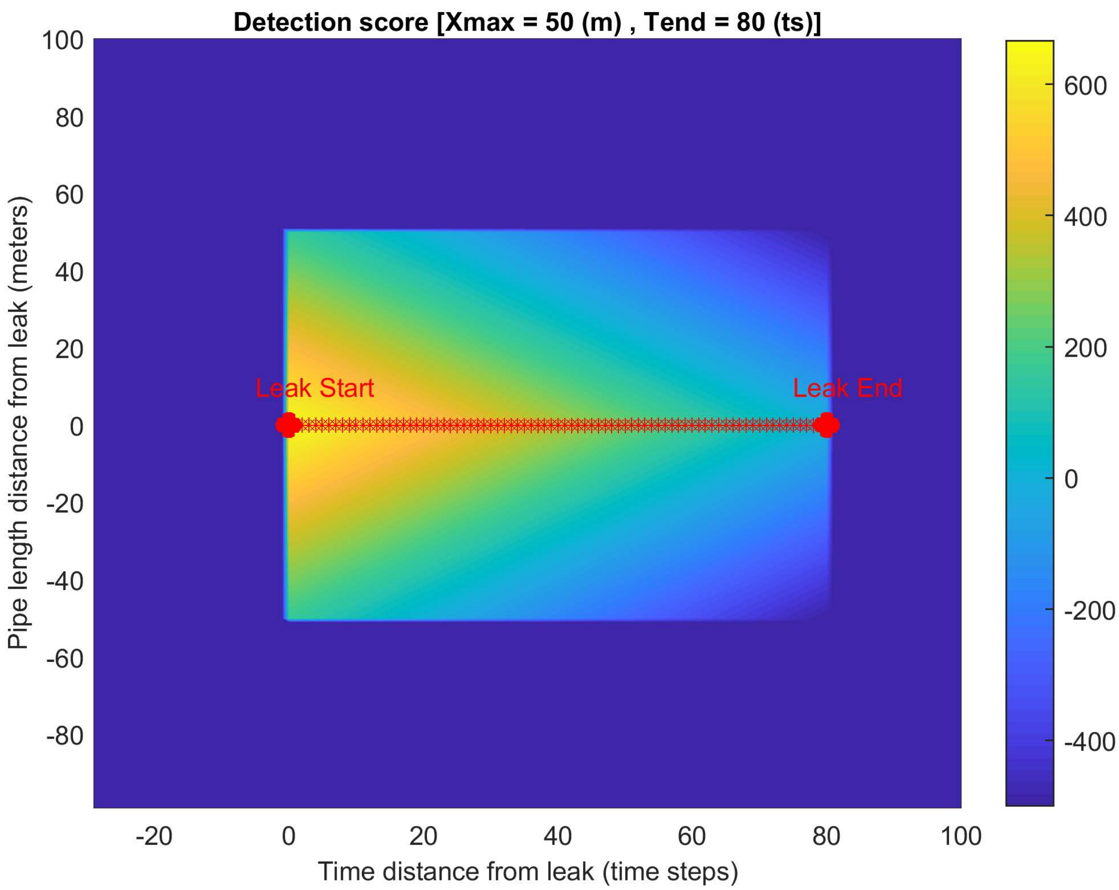

Fig. 8. Example of the scoring function for a true detection: $q(k) = 100\ m^3/h$ (leakage flow), $c_w = 1$ euro/m^3 (water cost), $c_r = 500$ euro/detection (max crew cost)

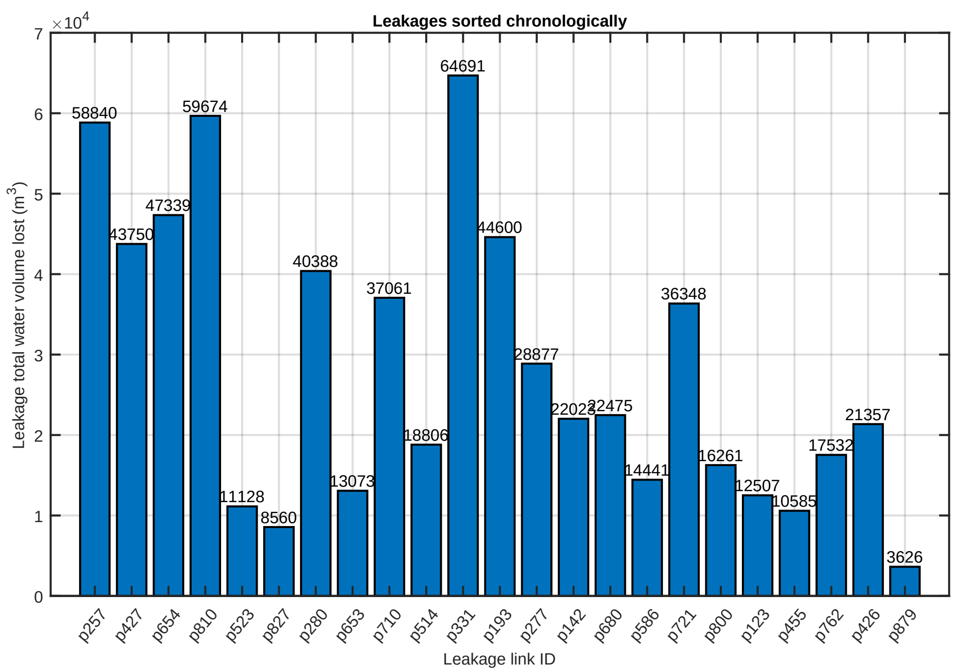

Fig. 9. Total volume of water lost from each leakage in the BattLeDIM problem, sorted chronologically and identified by the corresponding link ID.

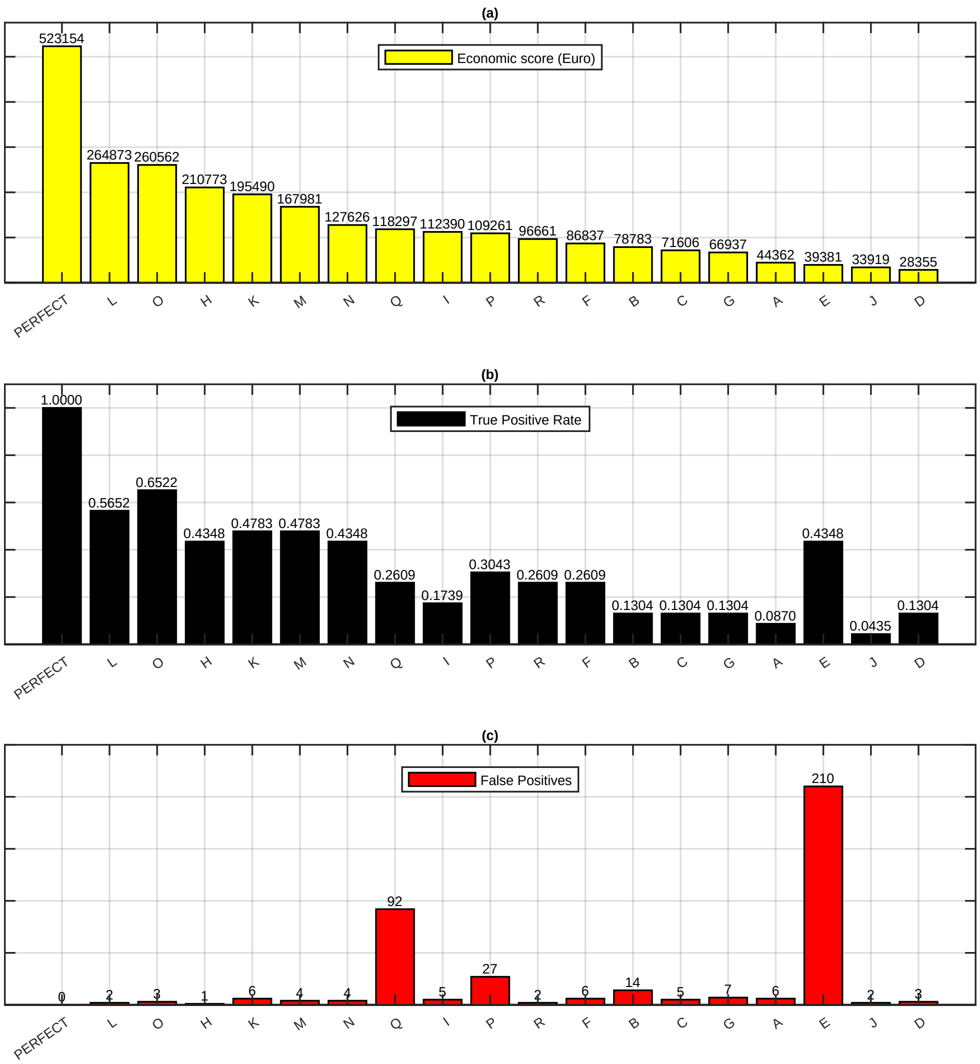

Fig. 10. (a) Final scores of the BattLeDIM competition: Team rankings are based only on the *Economic score*. The 'Perfect' score is the theoretical upper bound; (b) Team scores with respect to the *True Positive Rate* metric; (c) Team scores with respect to the number of *False Positives*.

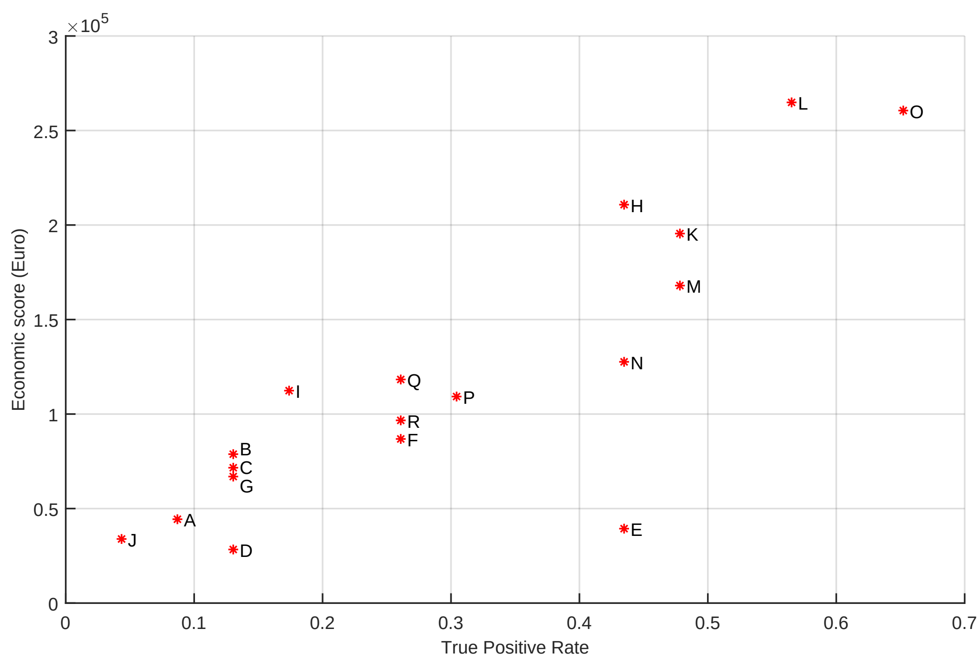

Fig. 11. Multi-parameter score (Economic score and True Positive Rate) of the submitted results. The best scores are in the upper-right corner of the graph.

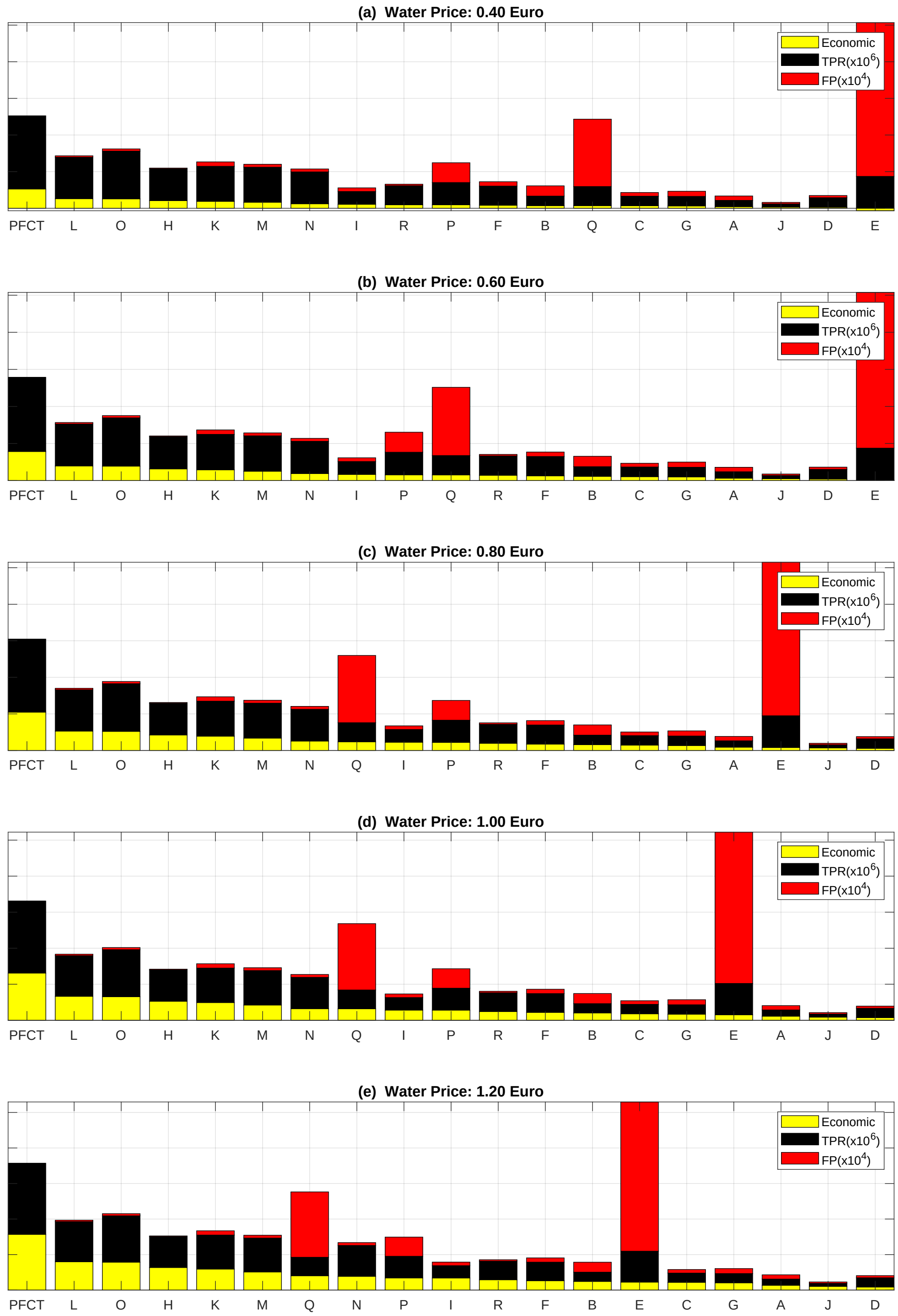

Fig. 12. Sensitivity analysis of the Economic score with respect to the price of water: (a) 0.40, (b) 0.60, (c) 0.80, (d) 1.00, (e) 1.20 Euro. Note that the True Positive Rates (TPR) and number of False Positives (FP) remain the same in these scenarios.

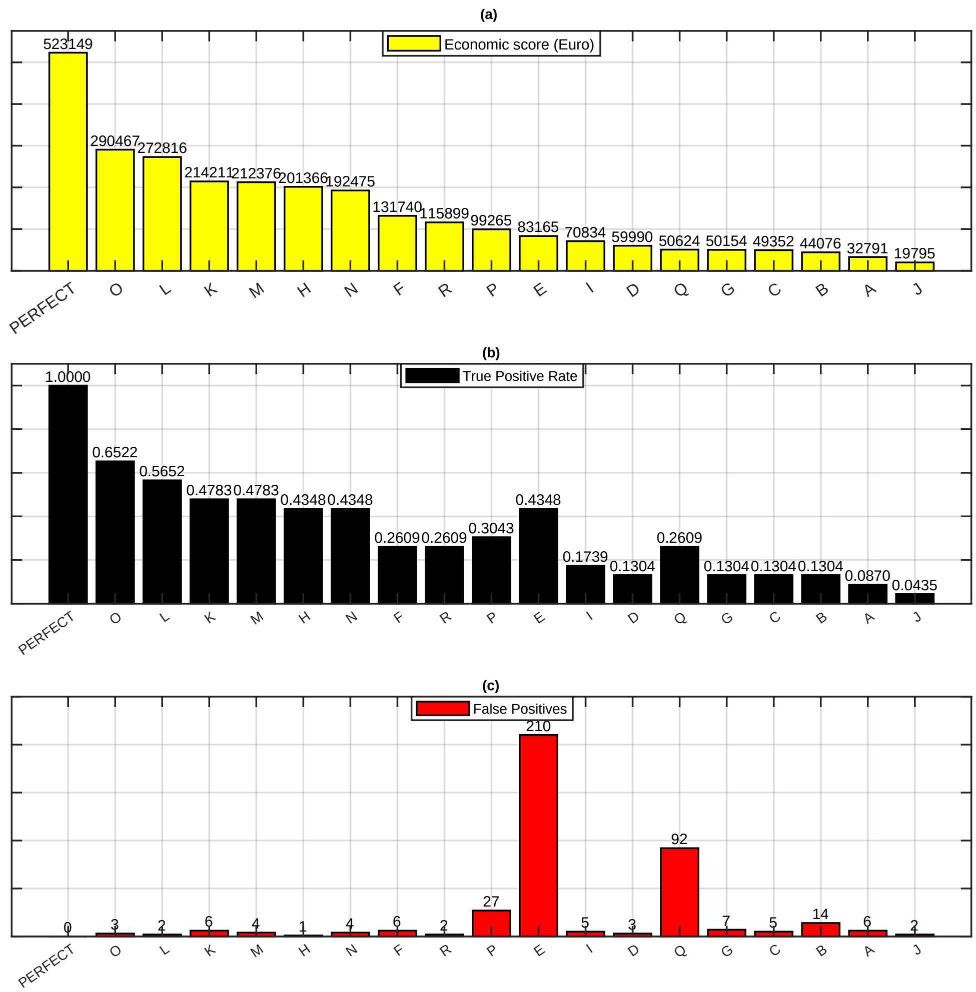

Fig. 13. (a) Alternative *Economic Score* and ranking of teams in the BattLeDIM competition using the alternative evaluation criteria in which the leakage volume is normalized; (b) *True Positive Rate* score; (c) Number of False positives.