**Article:**

Boatwright, S., Mounce, S. orcid.org/0000-0003-0742-0908, Romano, M. et al. (1 more author) (2023) Integrated sensor placement and leak localization using geospatial genetic algorithms. Journal of Water Resources Planning and Management, 149 (9). ISSN 0733-9496

https://doi.org/10.1061/jwrmd5.wreng-6037

INTEGRATED SENSOR PLACEMENT AND LEAK LOCALISATION USING GEOSPATIAL GENETIC ALGORITHMS

[1]Shaun Boatwright, [2]Stephen Mounce, [3]Michele Romano and [4]Joby Boxall.

[1] Affiliation: Department of Civil and Structural Engineering, University of Sheffield, Sheffield, S1 3JD, UK. Email: Shaun.Boatwright@sheffield.gov.uk

[2] Affiliation: Department of Civil and Structural Engineering, University of Sheffield, Sheffield, S1 3JD, UK. Email: s.r.mounce@sheffield.ac.uk (Corresponding Author)

[3] Affiliation: Data and Analytics Team, United Utilities Water Limited, Warrington, WA5 3LP, UK.E-mail: Michele.Romano@uuplc.co.uk

[4] Affiliation: Department of Civil and Structural Engineering, University of Sheffield, Sheffield, S1 3JD, UK. Email: J.B.Boxall@sheffield.ac.uk

## ABSTRACT

There is an urgent need to reduce water loss from drinking water distribution systems. A novel framework that integrates the placement of multiple pressure sensor and localisation using geospatial techniques is developed and validated to find leaks/bursts as they occur within District Meter Areas (DMA). A data-driven leak/burst localisation technique, featuring a novel spatially constrained inverse-distance weighted interpolation technique, was developed which quantifies the change in pressure due to a new leak/burst event using pressure sensors deployed in a DMA. The integrated framework uses the same modelling results and geospatial search techniques in both the optimal sensor placement and leak/burst localisation steps. It can be adapted for any data-driven or model-based leak/burst localisation technique and is not dependent on high hydraulic model calibration requirements such as high density smart meter deployment. Validation is presented using data from 16 engineered events (field work flushing) conducted in an operational DMA. Results show good agreement between the leak/burst localisation performance for real and modelled engineered events demonstrating that the sensor placement technique can accurately predict the expected performance of an operational DMA. This is particularly the case as the number of optimal sensors increases. Engineered events as small as 3.5% of the peak daily flow (6% of the average daily flow) were correctly localised with search areas containing as few as 14% of the pipes in the DMA (using only four pressure sensors).

**Key words:** Leak/burst localisation, optimal sensor placement, evolutionary algorithm, spatial constraint, network analysis, field testing

## INTRODUCTION

The size and capacity of a water distribution system (WDS) is dictated by the demographics of the supplied area and the distance between the source and the served population. In the Western world these are generally large, complex (and often ageing) infrastructures and water is inevitably lost between the points of treatment and consumption. Failures such as leak and burst events are a serious problem because they disturb customer supplies, lead to water loss (with subsequent impacts on cost, energy and carbon footprint), can lead to potential contamination and managing and repairing them consumes vast resources (Colombo and Karney, 2002, Fox et al. 2016). These various factors combine so that that water utilities need to develop new solutions and methods to better target the collection of data and its analysis, in order to decide where money should be invested to efficiently and effectively reduce water loss.

Much leakage management is reactive consisting of activities to detect, localise and repair leaks once they have occurred so as to minimise the impact on customers and the associated water loss. A district metered area (DMA) is a sub-section of the network which is partially hydraulically isolated from the surrounding network and flow is monitored at all entry and exit points. Detection aims to confirm the presence of leakage events and, where DMAs have been implemented, to provide an approximate location so that targeted intervention can be undertaken to resolve the problem quickly and with minimal disruption to customers. The current level of WDS monitoring makes detection (i.e. identifying the DMA(s) containing leak events) possible but localisation (or narrowing down the area within a DMA containing the event) requires a higher density of sensors than are currently deployed in most WDSs (Mounce et al. 2010, Xu et al. 2020). Most pressure sensors can be installed using a pressure tapping which means that access to only a portion of the pipe is required (and / or fire hydrants can be used) as opposed to flow sensors which require total excavation and pipe isolation.

Often, water companies are still typically reliant on their customers reporting problems with the water supply (such as low pressure / no water or a visible event) and on the use of labour intensive and expensive manual acoustic techniques. A number of modern techniques are currently used to allow the assessment, detection and control of leaks within WDS. Previous reviews by Puust et al. (2010) and Mutikanga et al. (2013) cover the range of various techniques and how they can be classified.

Engineered events are physically simulated flushing events introduced into a WDS by opening fire hydrants in the field and are used because they allow the run time, location and size of the event to be controlled. Determining the most informative size of leak or burst events to model represents a significant challenge. Large events typically impact the supplies of customers and must be dealt with swiftly to reduce the impact on targets related to customer supplies. Smaller events tend to be more difficult to detect and localise. There are a range of intermediate event sizes which are both too small to be readily apparent but also large enough to be localised using additional sensors placed throughout a DMA (referred to here as leak/ bursts).

This paper presents the details of a novel methodological framework for approximate leak/burst location. It encompasses a method for selecting the optimal number and location of sensors to be deployed in a particular DMA in order to achieve a desired level of event location performance whilst minimising the dependency on accurate hydraulic models. This is a significant advance on sensor placement techniques which only use spacing or entropy (in the sense of information theory) to determine the optimal configurations. This tight coupling between optimal sensor placement and approximate burst/leak location is of particular importance as an optimal sensor placement strategy depends on the method that is used to locate the potential leaks/bursts and likewise the efficiency of the burst/leak location depends on the sensor placement. Multiple engineered events (via field work) are used to validate the framework of methods.

## LITERATURE REVIEW

As part of the transition to smart networks, continuous, near real-time monitoring of WDS's hydraulic parameters (i.e. pressure and flow) has been the focus of significant research and development. Automatic near real-time techniques have been developed and implemented to minimise the leak/burst awareness time (Mounce et al. 2010, Romano et al. 2014) by effectively detecting, at the DMA level, leaks/bursts as soon as they occur (Li et al. 2015). A wide range of innovative data-driven leak/burst detection techniques utilising machine learning and advanced statistical tools have been developed (including with the use of deep learning, as in Wang et al. (2020)) which can analyse the large amounts of data and some have been implemented by water utilities (Wu and Liu, 2017), with wider proliferation ongoing as sensor rollout continues. After it is established that an event has occurred in a DMA, the next challenge is to determine the exact event location. Pressure data can help locate bursts when the pressure meters are properly deployed in a DMA, the temporal varying correlation of pressure drops between data from several pressure sensors in a single DMA providing a fingerprint to help with localization (Wu et al. 2018). This is challenging due to the often limited numbers of available pressure sensors within a DMA. The latest trends for leak/burst localisation pressure sensor deployment show that sensor densities of fewer than five pressure sensors per DMA are commonly deployed in the real world (Soldevila et al. 2019). Daniel et al. (2023) presented a pressure-driven algorithm for leakage identification and localization. Time series pressure data was used for leakage detection, with the sensor closest to the leak pinpointed and then this being utilized for localization via a hydraulic model (requiring a demand-calibrated model with smart meter data). Whilst the technique was able to localize leakages with an average distance of 149m in the case study presented, requirements included an almost error-free pressure calibrated model, 33 pressure sensors over 3 DMAs and demand data from 82 smart meters. Steffelbauer et al. (2022) tackled the same case study (BattLeDIM) with a method which consists of calibrating the nodal demand and pipe roughness and introducing a dual model for the calibrated primal problem to detect and locate leaks. They note that data set contains an unrealistically high number of sensors.

Methods which utilise hydraulic data to determine the approximate location of leaks and bursts within a DMA have emerged (Casillas et al. 2013a, Farley et al. 2013, Romano et al.

2013, Boatwright et al. 2016). In general these methods are not integrated with the detection step. These methods can allow a further reduction in the response time to leak/burst events by providing an indication of their likely location as a prelude to finding their precise location (i.e., pinpointing on a pipe). Key to the success of these approximate leak/burst location methods is the placement of the additional instrumentation.

As only a limited number of sensors can be installed in WDSs due to budget constraints and since improper selection of their location may seriously hamper leak/burst detection and localisation performance, the development of optimal sensor placement strategies is an important area for research and validation. A case can be made that sensor placement, in reality, is a multi-objective optimisation problem. There are many individual objectives, some which are competing, which must be satisfied simultaneously. Therefore, selecting the objectives which will be used to evaluate the quality of any given sensor configuration is a critically important step in developing a sensor placement technique which reflects the reality of the leak/burst localisation sensor placement problem. The definition of an 'optimised' sensor network depends on the context and design methodologies, relevant literature includes effective contaminant detection (e.g. Ostfeld and Salomons 2004, Shastri and Diwekar 2006) and model calibration - both hydraulic as in Kapelan et al. (2005) and water quality as in Zhang et al. (2020). Sanz et al. (2016) proposed a leak-detection and localization approach coupled with a (demand) calibration methodology that identifies geographically distributed parameters. Tested on synthetic data they were able to correctly detect and locate leaks within 200m. Qi et al. (2018a) proposed a methodology to investigate the underlying capacity of existing pressure sensors for pipe burst detection and thus provide practical guidance for effective burst management (though not per se to provide optimised sensor placement) . Five metrics were developed being: (1) undetectable nodes; (2) undetectable demands; (3)

detection dimensions (number of sensors that can simultaneously detect the bursts at each node); (4) the spatial partition, where bursts within each partitioned region can be detected by single sensor; and (5) the detectable threshold. All the metrics were calculated with the aid of a pressure-driven hydraulic simulation model (including generating pipe burst scenarios), and using Monte Carlo simulations to generate uncertain demand data. This work helps to advance the understanding of pressure sensor deployment to burst detection. In Qi et al. (2018b), similar work was presented for evaluating both hydraulic and water quality impacts. Six proposed metrics focus on identifying different aspects of impacts as follows, (i) break outflow volume, (ii) water shortage, (iii) nodes with significant pressure drops, (iv) pipes with significantly decreased pressures, (v) pipes with reversed flow directions, and (vi) pipes with significant increases in flow velocity. Computational simulation using hydraulic models was used in a similar manner to the first paper for the case studies, based on calculating metrics for each pipe break scenario. Two types of WDS case studies with varying attributes, scales, and topologies were used to demonstrate the utility of the comprehensive framework, although these did not include and field trials or real events for validation. Simulation results showed that impacts of pipe breaks not only vary with pipe diameters but are also significantly influenced by pipe locations, when the break occurs, and the specific metric considered. The understanding and findings obtained offer important guidance to developing effective pipe management / maintenance, resource planning and emergency response and break restoration strategies.

Almost all the sensor placement algorithms for leak/burst localisation in a DMA rely on modelling a large number of leak/burst scenarios. A number of hydraulic solver packages exist which allow leaks/bursts to be modelled relatively easily, albeit often under the (incorrect) assumption of a perfect (or at least well calibrated) model. Such a model needs to closely match the physical and hydraulic conditions which exist in the real network.

Various methodologies (e.g. Casillas et al. 2015, Steffelbauer and Fuchs-Hanusch 2016, Blesa et al. 2016) have been developed for selection of suitable sensor locations in order to maximise sensitivity to leaks/bursts whilst minimising the required number of sensors (and hence cost). Romano (2020) provides a detailed literature review of optimal sensor placement in WDSs for leak/burst detection and localisation. The almost complete lack of field tests and validation of the proposed techniques on real-life networks to assess their true value and practicality was noted along with the fact that comparing the effectiveness of the different proposed approaches remains an almost impossible task. Instead, tests and demonstrations of proposed techniques nearly always involve synthetic numerical experiments. That said, a few examples of field tests in the literature include the use of fire hydrants to simulate leaks. Farley et al. (2013) performed validation with a total of eight hydrant openings conducted in four DMAs. Six events were correctly localized to sub-DMA areas (ranging in size approximately between 50% to 33% of the full DMA). Romano et al. (2020) successfully approximately located a flushing event within a search area that was less than a quarter of the total length of mains in the DMA. In Fuchs-Hanusch and Steffelbauer (2016) a comparison of several methods including the methods proposed by Pérez et al. (2009), Casillas et al. (2013b) and Steffelbauer and Fuchs-Hanusch (2016) was carried out by opening fire hydrants to simulate different leak/burst scenarios in a real network and then assessing the leak/burst localisation capabilities of the different methods by calculating the distance between the suggested leak/burst locations and the opened fire hydrants. The results from the limited tests carried out in that study showed that using different leak/burst positions and different sensor sets, mainly those with sensors close to the leak/burst position, led to the best performance (otherwise localisation was inaccurate as regards sub-areas). Huang et al. (2020) presented a multistage bisection optimization approach to locate burst leaks within a DMA, where valve operations and water balance analysis based on smart demand meters (required for all users in the DMA, which is atypical) are iteratively performed to gradually narrow down the spatial regions associated with leaks. They were able to transfer artificial burst leak analysis to practical application for two leak regions (around 3-6% of the entire DMA) in a real WDS in China.

# FRAMEWORK OF METHODS

## Overview

A novel, integrated framework of methods to achieve the dual aims of optimal sensor placement and leak/burst localisation was developed which makes uses of a number of interconnected techniques as depicted in Figure 1. The two key steps, in the order which they are performed by the framework of methods are, to determine the optimal (or, more generally, near optimal) configuration of pressure sensors by analysing a hydraulic model of a DMA and subsequently, to analyse the pressure data from the optimal sensor configuration to determine the approximate location of a new leak/burst which has been detected in a WDS. Crucially the method utilises the same modelling results and geospatial search techniques in both the placement and localisation.

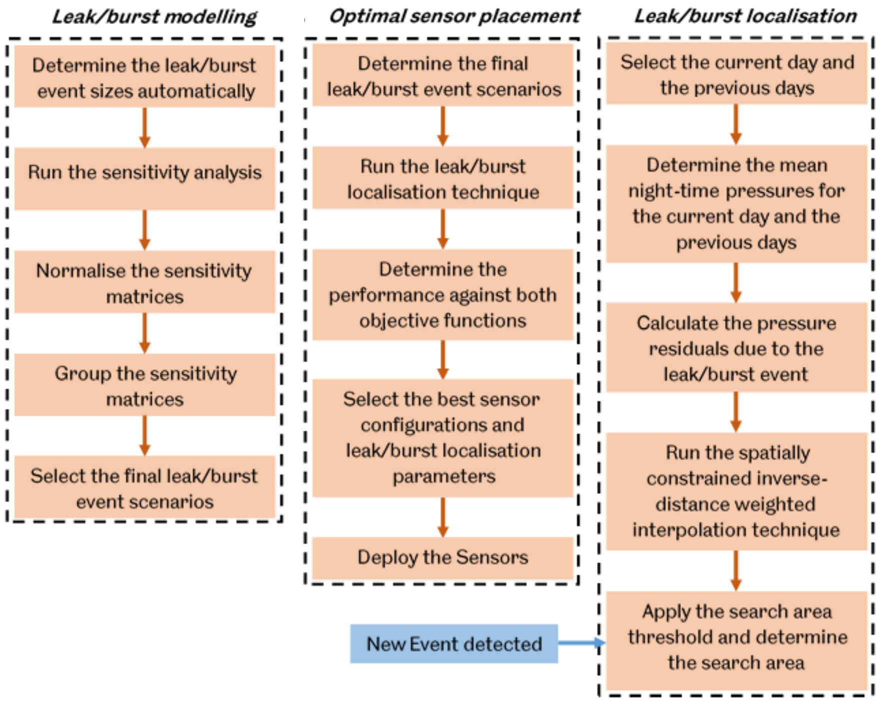


**Figure 1**: Schematic of the framework of methods

## Leak/ burst modelling

The framework of methods requires a hydraulic model. This model does not require any special additional calibration, and the standalone models used in the case study were of a typical UK water company 'top down' calibration standard.

The first step is to model a range of leak/burst event sizes at various leak/burst event locations throughout the DMA to build a sensitivity matrix. The sensitivity matrix contains the changes in pressure for all potential sensor locations for all of the leak/burst event scenarios (as in Farley et al. 2013). A mathematical representation of the sensitivity matrix is shown in Equation 1.

$$S = \begin{bmatrix} \frac{p_1^{f_1} - p_1}{f_1} & \cdots & \frac{p_1^{f_m} - p_1}{f_m} \\ \vdots & \ddots & \vdots \\ \frac{p_n^{f_1} - p_n}{f_1} & \cdots & \frac{p_n^{f_m} - p_n}{f_m} \end{bmatrix} \qquad (1)$$

Where $p_i^{f_j}$ is the pressure of sensor *i* when a leak/burst with constant flow, $f_j$, is present at node *j*, *m* is the number of node in the network (possible leak/burst locations – when leaks and burst are assumed as occurring at nodes), *n* is the number of sensors in the network and $p_i$ represents the pressure of sensor *i* without the presence of a leak/burst in the network.

For each sensor configuration, the pressures corresponding to the selected sensor locations are extracted from the sensitivity matrix and analysed by the leak/burst localisation technique. For each leak/burst event the size of the search area (percentage of total pipe length) is determined and the average size of the search area is calculated for the sensor configuration, considering all leak/burst events. For any event which is not correctly localised the whole size of the DMA is counted to assign a poor level of localisation for that leak/burst event. The average size of the search area for all leak/burst events (including the penalty for incorrectly localising leak/burst events) is used to measure the quality of each sensor configuration.

## Optimal Sensor Placement

A sensor placement technique minimises cost (numbers of sensors) whilst maximising performance when choosing where and how many sensors to place. The optimal sensor placement technique presented here is used to automatically determine the near optimal location of pressure sensors in a DMA to maximise the leak/burst localisation performance whilst simultaneously using the minimum number of pressure sensors (which is a trade-off). The sensor placement problem is formulated considering two objectives, namely the number of sensors and level of localisation performance which can be achieved. These configurations are specified with respect to multiple leak/burst event sizes which are determined automatically for each DMA. The critical point of the optimal sensor placement technique is that the leak/burst localisation technique is used by the optimal sensor placement technique to determine the level of localisation performance for a given sensor configuration.

### *Optimisation*

The GALAXY multi-objective evolutionary algorithm (MOEA) (Wang et al. 2017) is used to search for the optimal sensor configurations in an efficient way due to the potential number of possible configurations for DMAs of typical size. Once the maximum and minimum number of sensors was set for a given DMA, the sensor placement technique only needed to be run once to find all of the optimal configurations. The MOEA only requires two parameters to be specified. These are the size of the population and the number of functions evaluations (NFEs). The NFEs is the number of individual population members which are evaluated using the objective functions and is equal to the population size multiplied by the number of generations. This removes the main drawback of most GAs which need to specify and fine-tune several parameters to ensure that the performance of the MOEA is acceptable for the problem being considered. Following Wang et al. (2017), the amount of change between successive generations was used to indicate convergence, and the number of function evaluations was selected by running the sensor placement technique using several DMAs to ensure that less than 0.1% performance improvement was attained in those cases leading to a NFE of 100,000. A population size of 200 was chosen based on empirical tests over multiple DMAs. The objective function used by the MOEA is formulated to minimise the average size of the search area (percentage of DMA pipe length) and the number of sensors at the same time so that the optimal solutions with different numbers of sensors can be determined. The sensor placement technique determines the shape of the Pareto front (the curve of trade-off between multiple objectives) to allow water company personnel to select the appropriate number of sensors to deploy in a DMA.

The first objective function used by the optimal sensor placement technique evaluates the ability of a sensor configuration to localise all possible leak/burst event scenarios which are contained in the sensitivity matrix resulting from an event grouping technique. This procedure was used to group together leak/burst event locations using only the changes in pressure which were caused by them. An average localisation performance is determined considering all of the event scenarios which are contained in the grouped sensitivity matrix. The total weights of all graph links, which is the same as the length of the pipes that they represent, in the search area is allocated as the size of the search area (percentage of DMA pipe length). A penalty is applied to the size of the search area for an event which is not correctly localised i.e. that the modelled leak/burst location is not one of the nodes in the search area produced by the localisation technique. This penalty steers the population generated by the MOEA towards sensor configurations which localise the greatest number of events. A second penalty is also applied for any event for which multiple search areas are produced by the localisation technique. A graph-based procedure, which uses the same graph as the localisation technique was developed. If more than one search area is produced for an event then, in the same way as for incorrect localisation, all of the graph links are counted as being in the search area.

The second objective function used by the sensor placement technique is related to the cost of each sensor configuration generated by the MOEA. The number of sensors which are being installed is assumed by the sensor placement technique to be a proxy to cost. A sensitivity analysis was conducted to investigate the best combination of leak/burst localisation parameters (search area threshold and interpolation exponent) using two simple hydraulic models.

***Hydraulic modelling***

The basis of many sensor placement techniques for leak/burst detection and localisation is modelling a range of event scenarios so that the response of a DMA to these events can be used to inform sensor locations. In this context the decision variables are the locations of the sensors. A sensitivity analysis determines, for all possible combinations of leak/burst locations and leak/burst event sizes, the changes to both the pressure and, in some cases flow, which will occur. Due to practical considerations the optimal sensor placement technique implemented considers fire hydrants as the set of potential sensor locations. For all sensor placement runs in the case study a minimum value of three sensors and a maximum value of ten sensors has been used.

Several parameters must be specified or determined prior to performing the sensitivity analysis. The first two, which are related to the leak/burst events are the event sizes and the event locations. Additionally, the allowable sensor locations must also be specified. Only pressure data from these points can be considered. Next, a single run of the hydraulic model is performed so that the normal pressures and flows are captured. These are used to determine how much the pressure and flow has changed for each leak/burst scenario. The final step is to iteratively add leak/burst events to the hydraulic model, one node at a time, and store the changes in pressure which occur as a result.

A non-binarised sensitivity matrix is used as the input to the leak/burst localisation technique used by the sensor placement technique. To normalise the sensitivity matrix, the changes in pressure for each sensor location are divided by the normalisation factor, determined using a hydraulic model specifically for each potential sensor location. This normalisation step and the method for deriving the normalisation factors is used by both the sensor placement technique and the leak/burst localisation technique to maintain consistency.

***Leak/burst event size***

An automated approach was developed to determine the minimum leak/burst event sizes to, for each leak/burst event location, incrementally increase the leak/burst event size until one potential sensor location registers a change in pressure which is greater than or equal to the sensor accuracy. This ensures that only leak/event sizes which can be measured by at least one sensor in the network being studied are considered for each leak/burst event size. It also allows different leak/burst sizes or range of leak/burst event sizes to be determined for each event location in a given network. A similar process is used to determine the maximum leak/burst event sizes. However, the maximum event sizes are not calculated with respect to the change in pressure but the increase in flow which occurs as a result of each event. To achieve this, an additional parameter is used which is called the maximum allowable flow increase. This parameter is a percentage which relates the average daily flow to the additional flow which has occurred as a result of a modelled event considered by the sensor placement technique, calculated during the analysis window.

Once the maximum and minimum leak/burst event sizes have been determined for all leak/burst locations, and the validity of the event sizes are checked, the maximum and minimum event size for the entire DMA can be determined which are used for the final event size sensitivity matrices. For the purposes of the case studies presented a multiplier of 0.1 for the maximum allowable flow increase, equivalent to 10% of the daily peak flow, was used as this was felt to provide a good compromise biased towards finding smaller leaks (at risk of running undetected) rather than larger (obvious surface water etc.) bursts. Once the leak/burst event sizes have been determined a further novel step in the framework of methods is to group together the leak/burst event scenarios according to the sensitivities. This optimisation-based approach (similar to that first introduced by Sophocleous et al. 2019) reduces the number of events which are considered by the sensor placement technique by grouping together those event scenarios which are similar in terms of the sensitivities at all possible sensor locations. The MOEA is also used for this step with the groups of event locations as the decision variables. This process is repeated for all leak/burst events sizes, in turn.

## Leak/ Burst Localisation

Pressure was utilised as the hydraulic variable to measure and analyse due to the much lower cost and ease of installation of pressure sensors relative to flow sensors. Additional pressure sensors, and the data collected from them, are central to the approach and it is crucial that the (near) optimal combination of sensor locations are selected. This framework identifies the minimum number of sensors, to minimise the associated cost of the sensors for a water company, whilst simultaneously maximising the performance of the deployed sensor configurations.

The sensor placement and leak/burst localisation techniques both rely upon the fact that the characteristics of a new leak/burst event will lead to a specific set of effects on the hydraulics of the DMA which can be measured and used to infer the approximate location of the leak/burst event (by solving the inverse problem). There will be differences between the effects depending upon whether the event is modelled or is in a real WDS because a hydraulic model cannot perfectly match the real WDS, in addition monitoring data is never completely accurate. Once the optimal configuration of pressure sensors has been determined for the technique within certain limits using leak/burst simulations in a hydraulic model, and the pressure sensors have been deployed then the novel leak/burst localisation technique can be run when a new leak/burst is detected.

The framework of methods is designed and targeted towards leaks/bursts not reported by customers, i.e. leaks that do not result in surface water or other obvious effects. These events can easily be masked by the higher flows which are typically seen during the day but this problem is improved by assessment during the night-time period. To achieve this the leak/burst localisation technique can estimate the area (by providing a set of DMA pipes) containing the new leak/burst event by comparing the night-time pressure signals after the leak/burst event has occurred with recent signals, not modelling results so reducing the dependency on model accuracy. The magnitude of the change in pressure, measured by each deployed pressure sensor, is then used as the basis for leak/burst localisation. This means that there can be, at least, 24 hours between the commencement of the leak/burst event and the approximate location being calculated. Using the night-time pressure maximises the detectability of the change in pressure effects of the leak as there are minimal other demand induced variations.

A novel spatially constrained version of the inverse distance weighted (IDW) - referred to hereafter as SC-IDW - interpolation technique (Zimmerman et al. 1999) is used here to determine the location of new leaks/bursts occurring in a DMA. In IDW interpolation, the estimated value at unmeasured locations is determined as a weighted sum of the measured values where the weight for each measured location is based upon its proximity to the location being estimated. It can be mathematically described as in Equation 2.

$$Z_{est,j} = \sum \frac{z_{\mathrm{i}}}{[d_{ij} + s]^p} \Big/ \sum \frac{z_{\mathrm{i}}}{[1 + s]^p} \qquad (2)$$

where; $Z_{est,j}$ = estimated value at location j, $z_i$ = measured sample value at location i, $d_{ij}$ = distance between i and j, $s$ = smoothing factor (set to 0), $p$ = weighting power or exponent (range = 1-6).

The leak/burst localisation technique analyses of the data collected from the additional pressure sensors in a DMA over a number of days to determine the change in pressure (expressed as residuals) at each sensor location due to the new leak/burst. The novel, SC-IDW interpolation (Boatwright et al. 2016) then estimates the changes in pressure at all locations throughout the DMA to identify the area with the highest change in pressure, which is considered as the most likely area in the DMA containing the new leak/burst event. The distances used by the novel SC-IDW interpolation technique are determined as the shortest distance travelled between two points in the DMA (rather than using Euclidean distance). The distances are determined from an undirected graph which is automatically derived from a hydraulic model (as a two-dimensional square matrix of nodal distances).

One of the advantages of using SC-IDW interpolation, as opposed to more complex spatial analysis techniques, is that only a single parameter needs to be specified. This is the exponent used to calculate the weights between nodes, which depend upon the distance between them. The effect of the exponent is to control the amount of influence that near measured locations have when compared to distant measured locations on the value at the unmeasured point which is being estimated. The exponent influences the shape of the surface which is produced and using higher values of exponent tends to have a smoothing effect on the interpolation surface. Different values for the interpolation exponent have been used depending upon the domain and application (de Mesnard 2013). A process for automatically determining the best value of the interpolation exponent was developed to overcome this. Once the SC-IDW interpolation technique has been performed, the next step is to determine which parts of the DMA should be searched. The area with the largest estimated change in pressure is most likely to be in proximity to the leak/burst event so a method of dividing the surface of estimated values of the change in pressure is required. The search area threshold, $\boldsymbol{threshold_{sa}}$, is used to perform this which is set as a proportion of the range of values on the interpolation surface determined for a leak/burst event and can take any value between 0 and 1. Any locations on the interpolation surface with an estimated value which is above the numerical threshold (see equation 3) is designated as part of the search area.

$$\boldsymbol{x_{sa}} = \boldsymbol{z_{j,min}} + \boldsymbol{threshold_{sa}}(\boldsymbol{z_{j,max}} - \boldsymbol{z_{j,min}}) \tag{3}$$

Where $\boldsymbol{x_{sa}}$ is the numerical value of the threshold, $\boldsymbol{z_j}$ are the estimated changes in pressure for all locations for a single leak/burst event.

This information is then passed on to field teams to direct the search for the leak/burst event. The value of the search area threshold affects the proportion of the DMA which is included in the search area. The amount of the DMA which is in the search area is a measure of the localisation performance for a leak/burst event. The sensor placement technique automatically determines the best value for the search area threshold whilst determining the optimal sensor configurations.

# CASE STUDY

## Overview

A case study involving a DMA which had not been used in the development and verification of the framework of methods was used for validation. The combined performance of the sensor placement technique and the leak/burst localisation technique was determined by applying them both to the same DMA using data from engineered events (by flushing from hydrants) and industry standard hydraulic models.

The validation DMA, situated in a small urban town, is connected to a single service reservoir which has a standard diurnal profile. 12 additional sensors were deployed (non-optimally and for a variety of purposes) in the DMA as illustrated in Figure 2 (in which the locations of the engineered events are shown as red circles). The DMA is split into two discrete pressure areas (DPAs). The pressure in one of these is controlled using a pressure reducing valve (PRV) to regulate the pressure at a sufficient level as to maintain the minimum required pressure whilst also ensuring that there is not excessive pressure. All of the pressure sensors collected a pressure measurement every minute for a period of four months. The sensors were manually spread evenly throughout the validation DMA by water company staff to ensure that the entire DMA was covered in spatial terms (using the hydraulic model and GIS). These 12 hydrant locations (rather than all fire hydrants) effectively formed a constrained version of the sensor placement technique that was used to validate the framework.

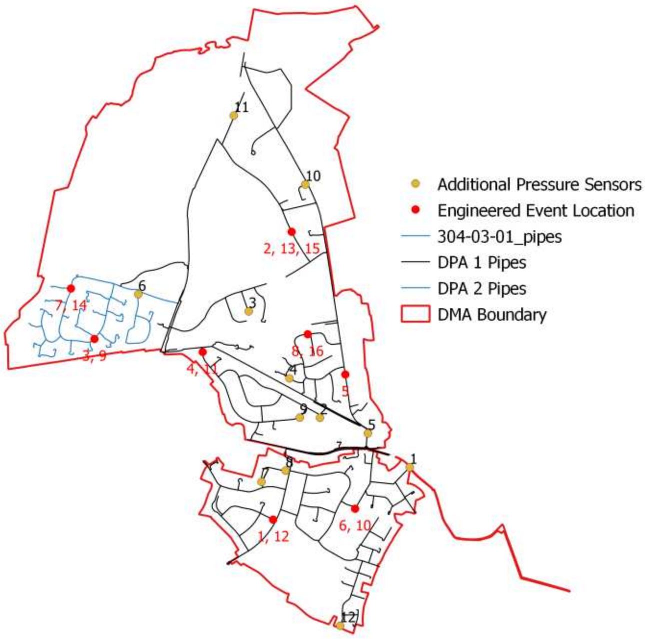


**Figure 2**: The 12 deployed sensor locations and DPAs for the validation DMA

## Sensor placement

Optimal sensor configurations with between 3 and 10 sensors (inclusive) were determined. In addition to identifying the optimal sensor configurations, some baseline sensor configurations were selected at random from the available sensor locations in the validation DMA, these were based on expert human judgement (academic modeller). Baseline configurations with between 3 and 10 sensors (inclusive) were selected so that a comparison of the performance of the baseline and optimal sensor configurations could be completed for all numbers of sensors considered. The baseline level of performance was used to represent the typical level of leak/burst localisation performance which would be achieved if sensors were deployed at random without using the sensor placement technique.

The optimal sensor configurations, which were constrained by the deployed sensor locations, were determined. Figure 3 shows these for the optimal case. Once the optimal sensor configurations were determined for the DMA, and combinations of parameters were found, the leak/burst localisation performance was determined for each engineered event that is; the size of the search area (percentage of DMA pipe length), whether the event was correctly localised and whether multiple search areas were determined. The average size of the search area, across all engineered events, was used as the measure of performance for each of the optimal sensor configurations. The same penalties for producing multiple search areas and incorrectly localising leak/burst events used by the sensor placement objective function were applied to the engineered events. This allowed a direct comparison to be made between the leak/burst localisation performance achieved using modelled data and real data collected during engineered events.

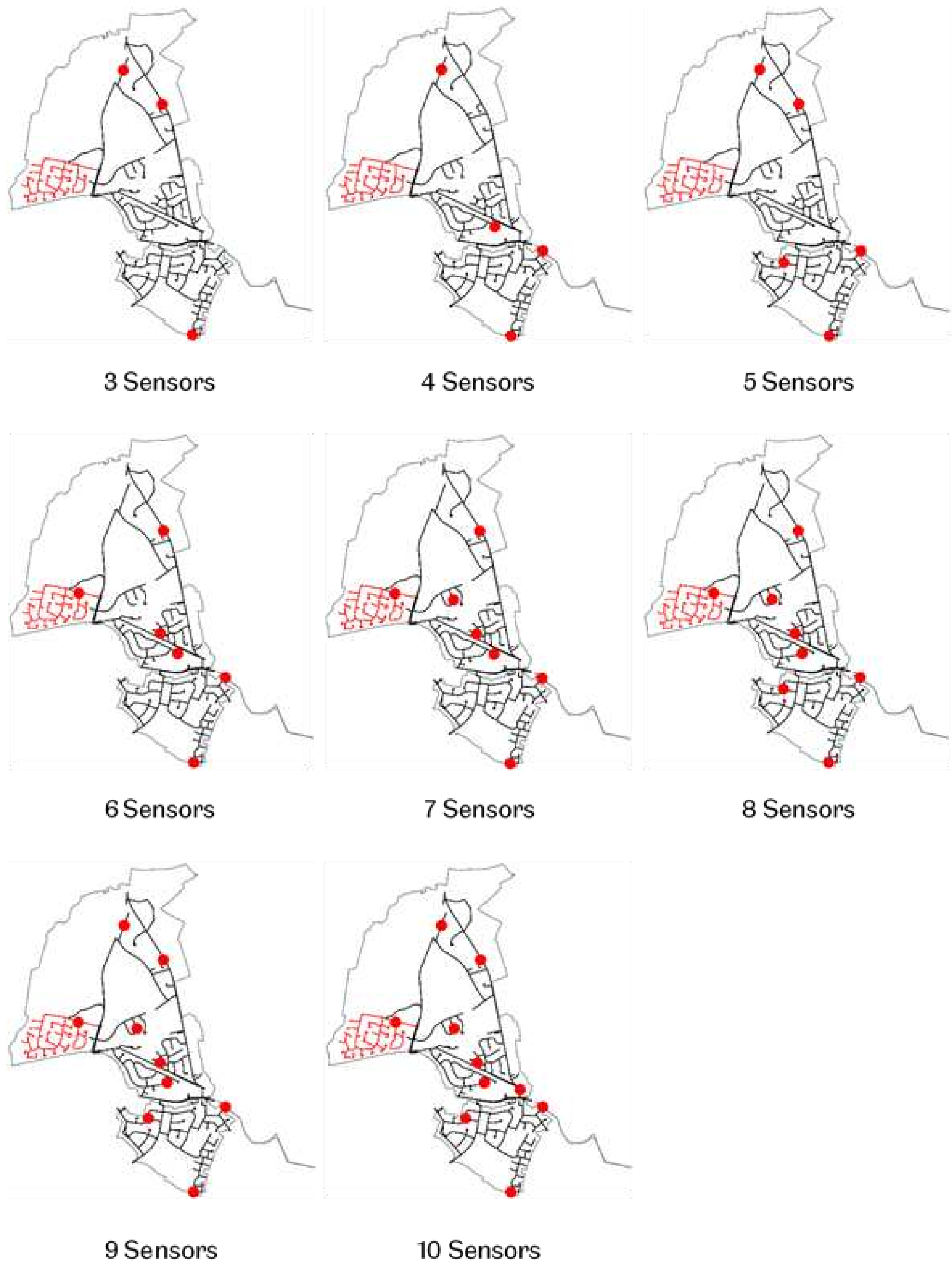


**Figure 3**: Optimal sensor configurations (red circles) with varying numbers of sensors for the validation DMA

To determine the maximum and minimum leak/burst event sizes an event size increment of $0.1 l/s/m^{0.5}$ was specified. The maximum allowable increase in flow was specified as 0.1 which equated to a 10% increase in the flow relative to the maximum daily flow. The automatically determined leak/burst event sizes and the corresponding results from the leak/burst event grouping technique are provided in Table 1. For each of the considered leak/burst event sizes, in the first column, the number of valid leak/burst event groups produced are given in the second column.

**Table 1**:

## Engineered Events

In a similar fashion to the deployed sensor locations, the engineered event locations were manually chosen to ensure that at least one engineered event was conducted in each section of the DMA. The sizes of the engineered events were selected to be just sufficiently large to be distinguishable from anomalous but legitimate customer demand which could still occur, even during the night.

The locations of the engineered events for the validation are shown as red circles in Figure 2. The indices for each event location are given by the red numbers next to the locations. Some event locations were used multiple times, providing replication, on different days such that 8 unique hydrants were opened across the 16 engineered events.

The engineered events were conducted during the period of minimum flow which is used by the leak/burst localisation technique. Each event was started before 03:00am and ended after 04:00am. The opening of each hydrant was controlled to achieve an approximate flow rate of 0.6 l/s at the start of the event and the same hydrant opening was maintained for the duration of the engineered events.

# COMPUTATIONAL RESULTS

**Pareto front for optimal versus baseline performance**

The Pareto front for the constrained optimal sensor configurations and parameters and the performance achieved for the baseline sensor configurations and parameters are shown in Figure 4. For all numbers of sensors, the constrained optimal sensor configurations outperformed the baseline sensor configurations, as expected. The difference in performance between the optimal and baseline configurations and parameters ranged from 18.1%, for the cases with 3 sensors, and 1.9% for the case with 10 sensors.

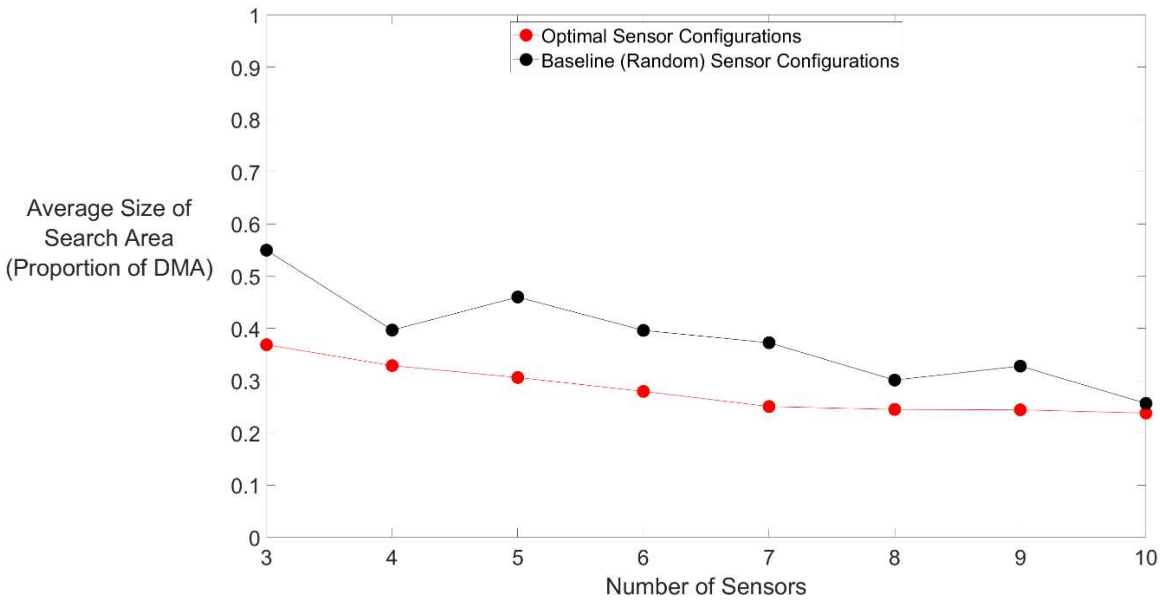


**Figure 4**: Comparison of the performance for the constrained optimal and baseline sensor configurations for validation DMA

**Overall results**

The performance of the leak/burst localisation technique, when applied to the each of the engineered events in the DMA, is given in Table 2. For each engineered event, the size of the search area (percentage of DMA pipe length) and whether the event was correctly localised within this are shown for each optimal sensor configuration. Correctly localised events are denoted by bold text and an asterisk denotes events for which multiple search areas were produced. At the bottom of the table the equivalent value for the objective function, accounting for both objective function penalties, is given so that results obtained using the modelled data and the real data can be compared.

**Table 2**

**Engineered versus modelled results**

A comparison between the localisation performance was conducted for engineered event 8. In Figure 5, the size of the search area produced using the modelled data has been plotted on the x-axis and the size of the search area determined using the engineered event data was plotted on the y-axis. A dashed black line, representing the line of perfect agreement between the modelled and real engineered event performance is also shown. The proximity of each of the optimal and baseline sensor configurations to the black line denotes the level of agreement between the two. In the ideal case, the size and location of the search area would be identical for both the modelled and the real engineered event when the same number of sensors are considered. Of all the sensor configurations only the optimal sensor configuration with 3 sensors showed perfect agreement between the search areas determined using the hydraulic model and the engineered event. The agreement for this sensor configuration was perfect even though there was a difference between the relative changes in pressure determined using the hydraulic model and for the engineered event. The level of agreement, denoted by the proximity of each configuration to the line of perfect agreement, for the optimal sensor configurations was higher than for the baseline sensor configurations. Other optimal configurations demonstrate how the leak/burst localisation technique can be used with some level of reasonable accuracy even when a very well calibrated hydraulic model is not available for use by the sensor placement technique.

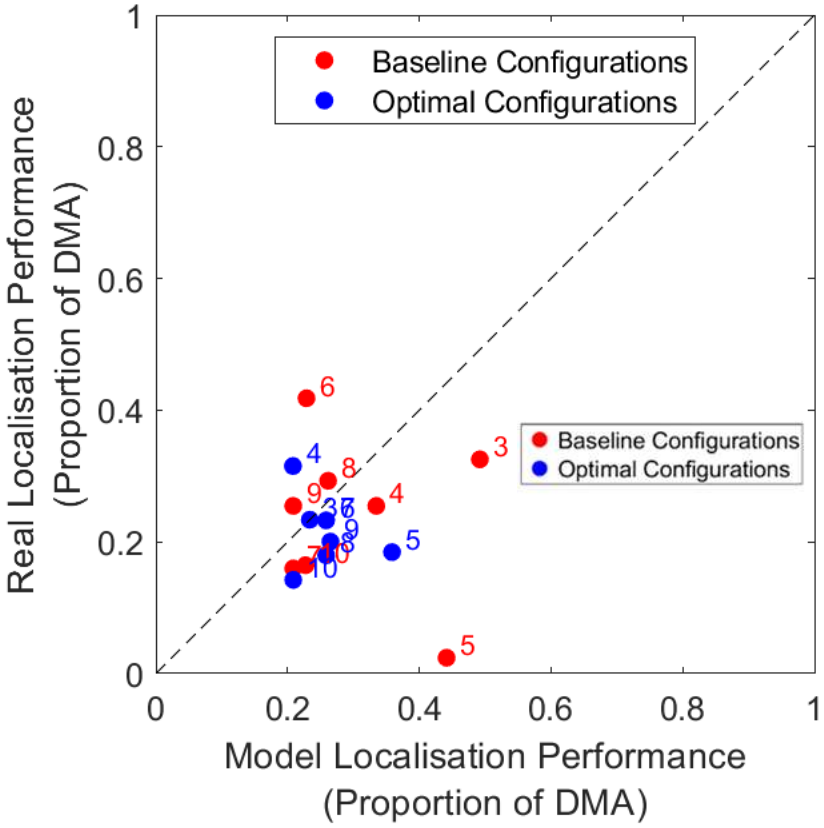


**Figure 5**: Agreement between the modelled and real search area sizes (percentage of DMA pipe length) for engineered event 8 using the optimal and baseline sensor

### Baseline versus constrained results

One key factor which influenced the leak/burst localisation results in Table 2 for the optimal (and baseline) sensor configurations was that the optimal and baseline sensor configurations were determined using a range of leak/burst event scenarios. It is informative in assessing the performance of the technique to determine what the expected level of leak/burst localisation performance was for the size (flow rate) of the engineered events, in the same way as for the sensor placement technique, so that they could be more directly compared. The search areas (percentage of DMA pipe length) determined by the leak/burst localisation technique using the modelled data for each of the engineered events, were corrected to account for the two penalties applied within the sensor placement objective function. The same correction was also applied to the search areas determined using the real engineered event data. The Pareto front obtained by running the constrained sensor placement technique was plotted against the corrected average search areas for the constrained optimal and baseline sensor configurations as shown in Figure 6. The constrained optimal configuration was selected from only 12 locations whereas the results plotted in black were determined using the sensor placement technique selecting from 145 hydrants. The Pareto fronts, determined by the sensor placement technique considering all of the leak/burst event scenarios determined by the event grouping procedure, are plotted in black.

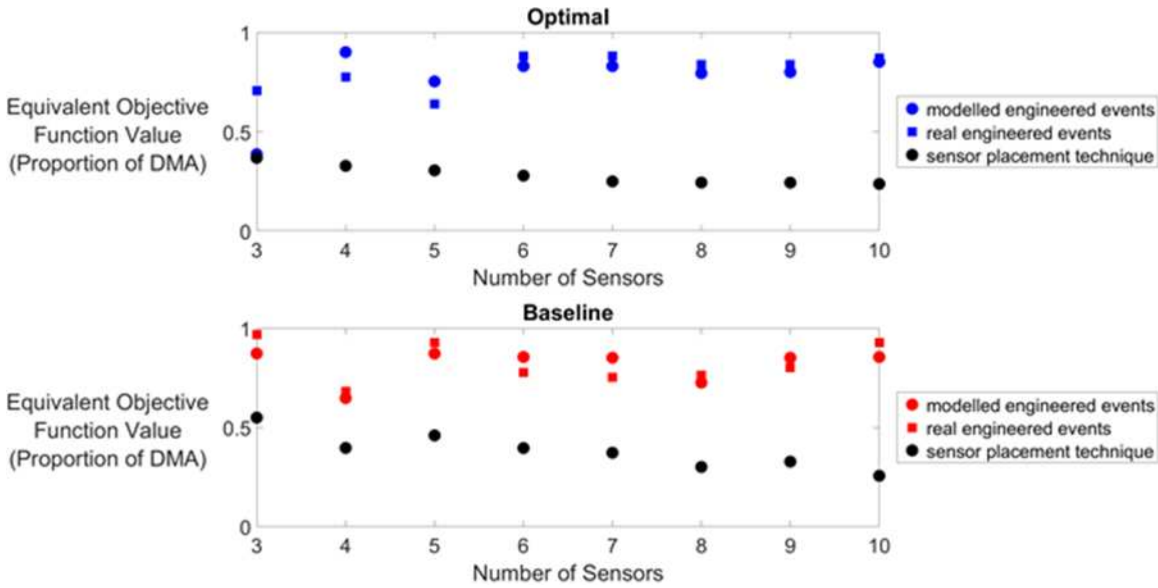


**Figure 6**: Comparison of the equivalent objective function values for the baseline and constrained optimal sensor configurations

A clear trend is that the leak/burst localisation performance achieved for the constrained optimal sensor configurations agrees very well when the modelled engineered events are compared to the real engineered events. Aside from the configuration with 3 sensors, the remaining constrained optimal configurations produced average search areas within 10% of each other. As the number of sensors increases the agreement between the results also improves, demonstrating that comparable performance was achieved even when a water industry standard hydraulic model was used to model the engineered events. By comparing the sensor placement results, plotted in black, with the modelled engineered events it was clear that the selected engineered event locations and sizes were much more difficult to localise. The only possible reason for this, aside from the different leak/burst event sizes and locations, was that the results plotted in blue were determined using the constrained optimal configuration.

Limiting the sensor placement technique to consider only the 12 deployed sensor locations resulted in there being a high degree of similarity between the constrained optimal and baseline sensor configurations. The fact that the performance achieved using the modelled engineered events for the baseline sensor configurations agreed more closely to the sensor placement technique than for the constrained optimal configurations indicates that the baseline configurations were more favourable for localising the set of engineered events. Constraint on the available sensor locations combined with extremely small engineered events made localisation difficult and essentially meant that the random and baseline performed very similarly. The agreement between the engineered events and their modelled equivalents was used as justification that the sensor placement objective function could accurately predict leak/burst localisation performance (the agreement was good for both baseline and optimal).

**Pareto fronts for constrained sensor placement technique**

The constraint included in the sensor placement technique, which limited the available sensor locations to only hydrants, has a limiting effect on the sensor configurations which can be made and on the resulting sensor placement performance. This is also true for the constraint due to the deployed sensor locations for validation. Therefore, a comparison between the Pareto fronts for several different versions of the sensor placement technique, with varying levels of constraint on the sensor locations, was conducted to determine the effect on performance of the sensor placement technique. In Figure 7 the sensor placement performance, measured using the average size of the search areas produced (over the engineered events), is plotted against the number of sensors for the three different levels of constraint.

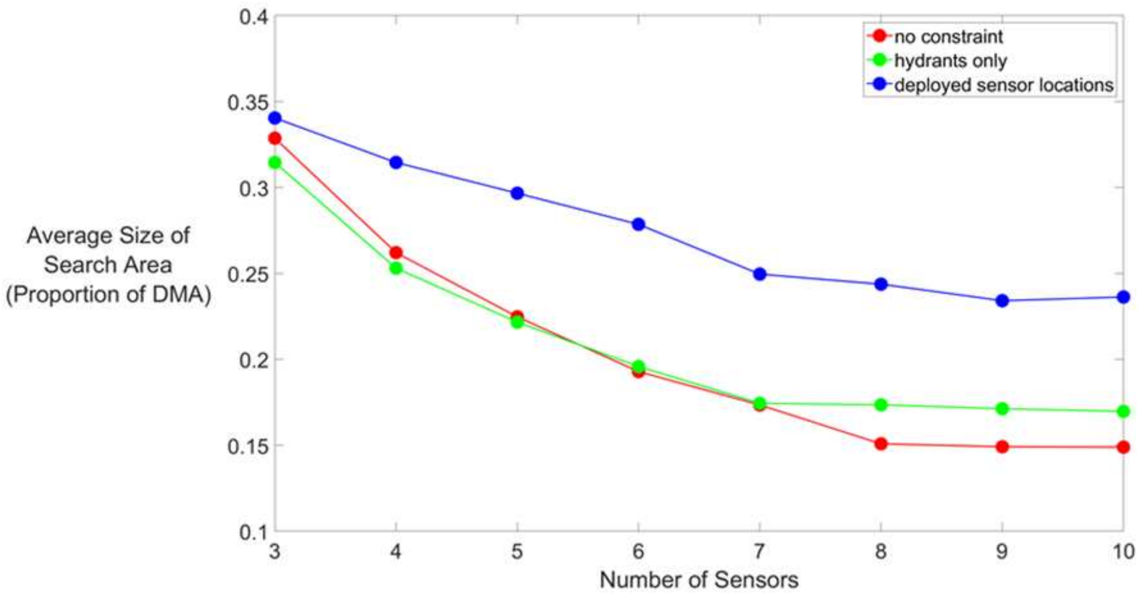


**Figure 7**: Comparison of the sensor placement performance for three levels of constraint of the sensor placement technique

The most constrained version, plotted in blue, is limited to just 12 deployed sensor locations (see Figure 2) and the least constrained version, plotted in red, could choose from 731 junctions in the hydraulic model. Between these two, plotted in green, the sensor placement technique could choose from 145 hydrants in the DMA. The lower the level of constraint, were there were more potential sensor locations to choose from, the better the sensor placement technique performed. The most constrained version of the sensor placement technique which only considered the 12 deployed sensor locations (blue) performed worse. The version with no constraint (red) performed very similarly to the version considering hydrants (green) for the cases with fewer than 8 sensors but when 8 or more sensors were considered the version with no constraint performed better by between 2-3% of the DMA. Hence, constraining to hydrants was only a little different to allowing any node as potential sensor location thus minimising the reduction in the leak/burst localisation performance.

The average size of the search areas produced, including the penalties for incorrect localisation and multiple search areas, for all 16 engineered events was determined for each number of sensors. The results for the three level of constraint are plotted in Figure 8 as regards leak/ burst localisation performance.

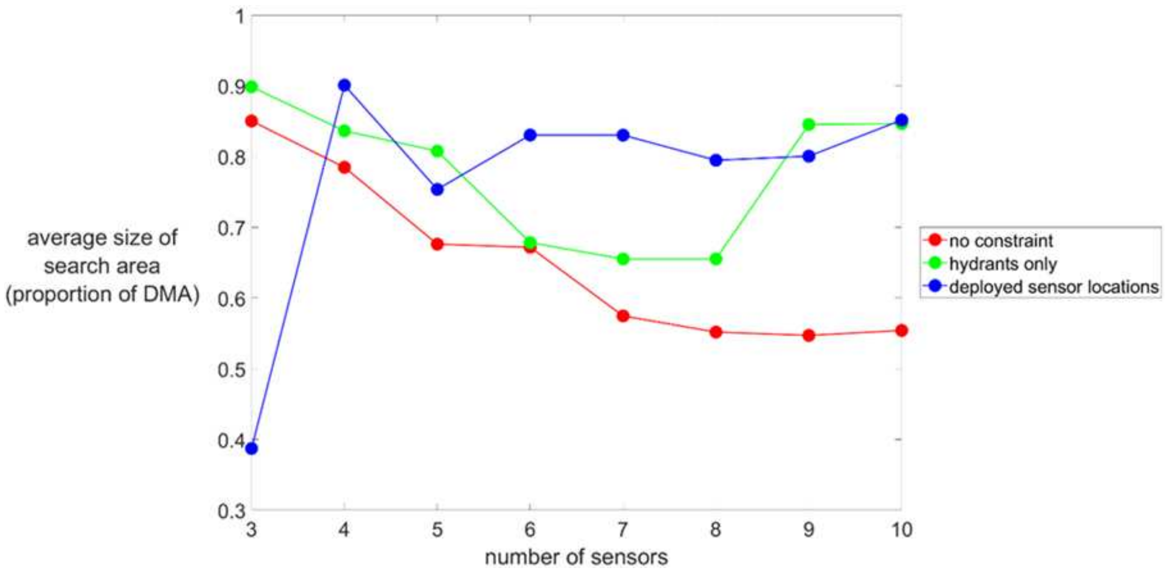


**Figure 8**: Comparison of the leak/burst localisation performance for the 16 modelled engineered events for three different levels of constraint of the sensor placement technique

The least constrained version of the sensor placement technique, where all hydraulic model junctions were considered as valid sensor locations, performed best overall. The sensor configurations used in this figure were determined with respect to all events used by the sensor placement technique and not the 16 modelled events. Aside from the case with three sensors, smaller average search areas were produced for all sensor configurations than for either of the two other constraint levels. As the constraint was increased, by considering all hydrants as valid sensor locations, a marked decrease in the leak/burst localisation performance was seen. The most constrained version of the sensor placement technique, which considers only the 12 deployed sensor locations, was, on average the worst performing. By comparing the results in Figure 7 with those in Figure 8 the effect of only considering a small number of leak/burst events was clear. The localisation results in Figure 8 did not produce smooth lines since the smaller number of event locations considered was much more sensitive to individual sensor locations. By comparing the green and blue lines in Figure 8 the effect of constraining the sensor placement technique to only selecting from the deployed sensor locations illustrated that the effect was to reduce the leak/burst localisation performance by approximately 1%. For the sensor configurations with more than 3 sensors the average reduction in localisation performance was approximately 6% of the size of the DMA. Concerning the case with 3 sensors for the deployed sensor locations (blue line) unexpectedly good performance was seen compared to other sensor configurations and levels of constraint. This was because the sensor configurations were determined relative to all possible leak/burst locations but the results shown pertain to only the engineered event locations. In this case, using 3 sensors enabled the leak/burst localisation technique to correctly localise more events than configurations with more sensors, reducing the applied penalty and thereby improving the average performance.

**Effect of small engineered event sizes**

Figure 9 demonstrates that the small size of the engineered events negatively impacted the leak/burst localisation performance when compared to the range of event sizes generated by the sensor placement technique. As the leak/burst event size increased the average size of the search area decreased across the 16 modelled engineered events. The emitter coefficient of 0.5 produced the smallest average search areas and further increases in event size above this did not lead to further reductions in the average search area size. The actual size of the engineered events led to the significantly worse performance than for any of the event sizes determined by the sensor placement technique. Comparing the actual event sizes (black line with crosses) with the worst performing of the determined leak/burst event sizes (C = 0.1 -

red line with crosses) the average reduction in performance was on average approximately 6% of the DMA. For the best performing of the determined leak/burst event size (C = 0.7 – blue line with circles) the reduction in performance was 47% of the DMA on average. Therefore the small size of the leak/burst events was more impactful than using the constrained version of the sensor placement technique.

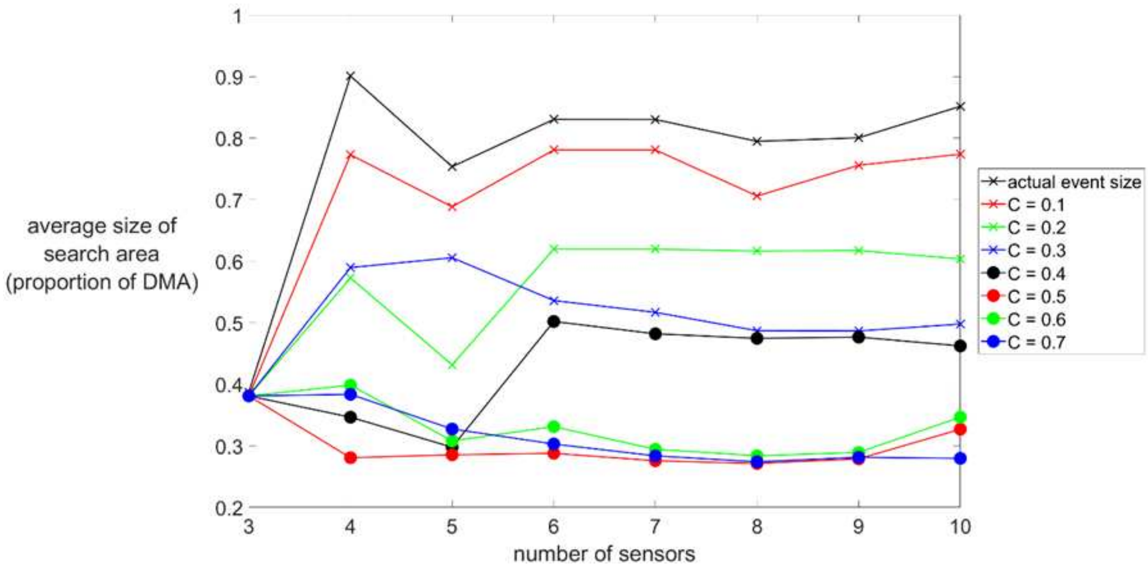


**Figure 9**: Comparison of the leak/burst localisation performance for 16 modelled events considering different sizes of event using the constrained optimal sensor configurations

## DISCUSSION

The framework developed is integrated so that the leak/burst localisation technique is used by the sensor placement technique to determine the optimal configuration(s) of pressure sensors. The only difference is that real data is used by the leak/burst localisation technique whereas hydraulic model simulated data is used by the sensor placement technique. This ensures that the sensor configurations consider not only the hydraulic changes which are caused by leak/burst events but also the technique which is being used to localise them. A range of optimal sensor configurations are determined, although this is not guaranteed for every DMA, so that they can be presented to decision makers in the form of a Pareto front to make an informed decision considering the trade-off between the number of sensors and the leak/burst localisation accuracy. Field teams can use this information to focus the search for the leak/burst event so that it can be precisely located and repaired. This framework presented can be easily adapted for any data-driven or model-based leak/burst localisation technique including the traditional approaches which are currently mainly used by many water utilities such that the localisation accuracy, including the number of leak/burst events correctly localised, is maximised.

The leak/burst localisation technique is data-driven and does not require a calibrated hydraulic model. Many leak/burst localisation techniques are developed for use with hydraulic model data only, in other words they only ever use simulated data and this poses serious transferability problems to real WDSs (e.g. Marzola et al. 2022). Model-based leak/burst localisation techniques are highly sensitive to the quality of model being used and will only perform well when a highly calibrated model is available. The sensor placement technique does not rely on a very well-calibrated hydraulic model, instead a UK industry standard calibrated model, because the leak/burst localisation technique uses the differential changes in pressure throughout the DMA to infer the approximate location of a leak/burst event.

Evaluating performance using engineered events gives greater confidence in the achieved results than solely using model simulations when related to the achievable performance on real leak/burst events in WDSs. The past pressures are determined for each pressure sensor in the DMA, calculated over 14 nights, and then compared to the pressure in the presence of a new leak/burst event. Using the night time means that the uncertainty in the customer demands is minimised and this has less influence over the leak/burst localisation performance. This is also aided by the use of the SC-IDW technique because, for every location in a DMA, multiple pressure sensors are used to estimate the change in pressure. This reduces the reliance on a single sensor which, in turn, reduces the influence of sensor uncertainty on the determined search area. The SC-IDW technique is an improved version of IDW interpolation which accounts for the layout of pipes in a DMA. By using the novel distance function (calculated using the shortest path along pipe lengths) as opposed to Euclidean distance, more realistic and often smaller search areas were produced than when compared to Romano et al. (2013). The novel distance function is able to distinguish between two pipes which are close together but which are not connected directly to each other which represents a significant improvement over results from Romano et al. (2013) where pipes which are distant, in terms of the length travelled along pipes, were still included in the area with the highest probability of a leak/burst event.

Leak/burst events as small as 3.5% of the peak daily flow were correctly localised using as few as 3 optimal pressure sensors. This compares favourably with the technique developed by Soldevila et al. (2019) which localised (to within 200m) a leak/burst event at night of approximately 8% of the peak inflow using 5 sensors. This performance was achieved even though the size of the engineered events was smaller than the minimum theoretical leak/burst event sizes determined prior to running the sensor placement technique.

This is the first automatic procedure for determining a range of leak/burst event sizes for sensor placement in a DMA for the purpose of leak/burst localisation. A previous attempt at reducing the size of the search space by Sophocleous et al. (2019) grouped together the leak/burst events by proximity. However, even leak/burst events which are close together can have very different effects on a DMA. Using the effect on pressure directly to group the leak/burst events overcomes this problem. A further novel procedure, developed as part of the sensor placement technique, groups together the leak/burst event scenarios to reduce the number of leak/burst events which must be considered by sensor placement.

The number of additional pressure sensors deployed throughout a DMA is a critical factor which dictates the achievable leak/burst localisation performance. The improvement in both localisation and sensor placement performance diminished as higher numbers of sensors were considered meaning that water utilities would likely prefer to deploy fewer sensors although this would depend upon their strategic aims.

The potential sensor locations were limited to hydrants only. Despite this, leak/burst localisation performance reduction was found to be only around 2-3% when compared to considering all junctions as candidate sensor locations. As illustrated in Figures 7 and 8, the constraint of only selecting from the hydrant locations did not lead to a significant reduction in sensor placement performance, indicating the validity for such a constraint particularly in light of the ease of installation and the reduced cost of deploying sensors at hydrants.

## Further Work

A number of avenues are apparent for development of this system in future work. Firstly, a study could be conducted to determine the relationship between the level of hydraulic model calibration and the sensor configurations which are determined as a result. Further to this the leak/burst localisation performance, determined using engineered events, can be ascertained for the optimal sensor configurations with varying levels of hydraulic model calibration.

A detailed cost-benefit analysis could be performed for deploying additional pressure sensors for leak/burst localisation and built into the sensor placement technique. There is a cost which can be associated with each type of individual leak/burst event which can occur in a DMA which is related to regulatory targets (and associated penalties) and the cost of repairing it. For each DMA, an associated cost of leak/burst events can be determined and compared to the cost of deploying pressure sensors.

## CONCLUSIONS

A framework of methods has been developed to localise new leak/burst events in WDS using a data-driven leak/burst localisation technique integrated with a sensor placement algorithm and utilising the same modelling results and geospatial search techniques in both. The leak/burst localisation technique is used by the sensor placement technique ensuring that the determined sensor configurations are optimal with respect to the leak/burst localisation technique. The localisation technique does not require a hydraulic model to be compared with the measured pressures from a real DMA which reduces its reliance on a highly calibrated hydraulic model when compared to model-based leak/burst localisation techniques. Multiple leak/burst event sizes and locations were considered in real world engineered events ensuring that the range of leak/burst event scenarios of interest, namely those events which are typically unreported and cause changes in pressure throughout a DMA, are used to determine the optimal sensor configurations. Key findings include:

- The sensor placement technique was used to determine a Pareto front of optimal sensor configurations (and parameter combinations) with varying numbers of sensors. A MOEA was used to determine the sensor configurations and combinations of parameters which simultaneously minimises the required number of sensors and maximises the localisation performance.

- The distance used by the SC-IDW interpolation technique is determined as the shortest path between two points in a DMA travelling along the pipes ensuring the topology of the pipes in the DMA is respected unlike using Euclidean distance as per traditional IDW interpolation.
- By incorporating the novel sensor placement technique into the integrated framework (and using the same SC-IDW interpolation technique) the leak/burst localisation accuracy can be improved. This was demonstrated by comparing the localisation performance of the integrated framework against a baseline sensor placement technique.
- A novel procedure for determining the smallest and largest leak/burst event size for all considered leak/burst event locations was developed. This automatic procedure requires the specification of three parameters so that all leak/burst event sizes can be determined for an entire DMA.
- A novel leak/burst event grouping procedure was used to group together leak/burst event locations using only the changes in pressure which were caused by them. This significantly reduced computational effort and ensured that leak/burst events which caused very similar changes in pressure for all candidate sensor locations were not considered separately by the sensor placement technique.
- A total of 16 engineered leak/burst events, by opening fire hydrants, were carried out in a real DMA to conduct framework validation. The constrained optimal sensor configurations were determined from the deployed sensor locations by using a heavily constrained version of the sensor placement technique. Constraining to hydrants was almost the same as allowing any node as potential sensor location, and which results in a no-dig solutions in practice. Baseline sensor configurations were also selected at

random from the deployed sensor locations and a comparison between the optimal and baseline configurations was conducted.

- Validation demonstrated that the novel sensor placement technique can tolerate inconsistencies/inaccuracies which typically exist in the water utility hydraulic models which have been used to determine the optimal sensor configurations.
- Results demonstrated that the system can typically locate leaks/bursts to a small fraction of the DMA (best localisation performance achieved was approximately 14%). The framework was proven to be able to successfully locate leaks/bursts as small as 3.5% of the peak daily flow with as few as 3 additional pressure sensors installed.

## DATA AVAILABILITY STATEMENT

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions. Direct requests for these materials may be made to the provider (United Utilities). The thesis underpinning this paper is available for download (Boatwright, 2020).

## ACKNOWLEDGEMENTS

The authors wish to thank the innovation team at United Utilities and the engineering and physical sciences research council (EPSRC) for funding this work under a STREAM IDC research project (grant number EP/L015412/1).

## REFERENCES


Blesa, J., Nejjari, F. and Sarrate, R. (2016). Robust sensor placement for leak location: analysis and design. Journal of Hydroinformatics, 18, 1, 136–148.

Boatwright, S., Romano, M., Mounce, S. R., Woodward, K. and Boxall, J. B. (2016). Approximate location of leaks and bursts in a district metered area using statistical process control and geostatistical techniques, Proc. 14th International Computing and Control for the Water Industry Conference, Amsterdam, Netherlands.

Boatwright, Shaun (2020) Integrated optimal pressure sensor placement and localisation of leak/burst events using interpolation and a genetic algorithm. EngD thesis, University of Sheffield. https://etheses.whiterose.ac.uk/29227/

Casillas, M. V.; Garza-Castanon, L. E. and Puig, V. (2013a). Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. Journal of Hydroinformatics, 16, 3, 649–670.

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

Casillas, M. V., Garza-Castanon, L. E. and Puig, V. (2015). Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms". Water, 7, 11, 6496–6515.

Colombo, A. F., Karney, B. W. (2002) Energy and costs of leaky pipes: toward comprehensive picture. J Water Resour Plan Manag 128(6):441–450.

Daniel, I., Pesantez, J., Letzgus, S., Fasaee, M. A. K., Alghamdi, F., Berglund, E., Mahinthakumar, G. and Cominola, A. (2023) A Sequential Pressure-Based Algorithm for Data-Driven Leakage Identification and Model-Based Localization in Water Distribution Networks. J. Water Resour. Plann. Manage., 2022, 148(6): 04022025 de Mesnard, L. (2013). Pollution models and inverse distance weighting: Some critical remarks. Computers and Geosciences. 52:459-469.

Farley, B., Mounce, S. R. and Boxall, J. B. (2013). Development and Field Validation of a Burst Localization Methodology. J Water Res. Plan. Man. 139(6):604-13.

Fox, S., Shepherd, W., Collins, R. and Boxall, J. B. (2016). Experimental quantification of contaminant ingress into a buried leaking pipe during transient events. Journal of Hydraulic Engineering, 142 (1), 04015036.

Fuchs-Hanusch, D. and Steffelbauer, D. (2016). Real-world comparison of sensor placement algorithms for leakage localisation. Proceedings of the 18th Water Distribution Systems Analysis Conference, Cartagena de Indias, Colombia.

Huang, Y., Zheng, F., Kapelan, Z., Savic, D., Duan, H.-F., & Zhang, Q. (2020). Efficient leak localization in water distribution systems using multistage optimal valve operations and smart demand metering. Water Resources Research, 56, e2020WR028285. https://doi.org/10.1029/2020WR028285.

Kapelan, Z., Savic, D. A. and Walters, G. A. (2005). “Optimal sampling design methodologies for water distribution model calibration”. Journal of Hydraulic Engineering, 131, 3, 190–200.

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 & Technology: Water Supply, 15, 3, 429–441.

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. J. Water Resour. Plann. Manage., 2022, 148(1): 04021096. DOI: 10.1061/(ASCE)WR.1943-5452.0001503

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. J. Water Res. Plan. Man. 136(3):309-18.

Mutikanga H. E., Sharma, S. K., Vairavamoorthy, K. (2013). Methods and tools for managing losses in water distribution systems. J Water Resour Plan Manag 139(2):166–174.

Ostfeld, A. and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security. Journal of Water Resources Planning and Management, 130, 5, 377–385.

Pérez, R., Puig, V., Pascual, J., Peralta, A., Landeros, E. and Jordanas, L. (2009). "Pressure sensor distribution for leak detection in Barcelona water distribution network". Water Science & Technology Water Supply, 9, 6, 715–721.

Puust, R., Kapelan, Z., Savic, D. A. and Koppel, T. (2010) A review of methods for leakage management in pipe networks," Urban Water J., vol. 7 (1), pp. 25-45, 2010.

Qi, Z., Zheng, F., Guo, D., Maier, H. R., Zhang, T., Yu, T. and Shao, Y. (2018a). Better understanding of the capacity of pressure sensor systems to detect pipe burst within water distribution networks. J. Water Resour. Plann. Manage., 2018, 144(7): 04018035.

Qi, Z., Zheng, F., Guo, D., Zhang, T., Shao, Y., Yu, T., Zhang, K., & Maier, H. R. (2018b). A comprehensive framework to evaluate hydraulic and water quality impacts of pipe breaks on water distribution systems. Water Resources Research, 54, 8174–8195. https://doi.org/10.1029/2018WR022736

Romano, M., Kapelan, Z. and Savic, D. A. (2013). Geostatistical Techniques for Approximate Location of Pipe Burst Events in Water Distribution Systems, J. Hydroinform. 15(3):634–651.

Romano, M., Kapelan, Z. and Savic, D. A. (2014). Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems, J. Water Res. Plan. Man. 140(4):457-67.

Romano, M. (2020). Review of Techniques for Optimal Placement of Pressure and Flow Sensors for Leak/Burst Detection and Localisation in Water Distribution Systems, in Dawei H., Mounce, S., Scozzari, A., Soldovieri, F., and Solomatine, D. (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem.

Romano, M., Boatwright, S., Mounce, S. R., Nikoloudi, E. and Kapelan, Z. (2020). AI- based event management at United Utilities. Hydrolink, IAHR. No. 4/2020, pp.104-108.

Sanz, G., Perez, R., Kapelan, Z. and Savic, D. (2016). Leak detection and localization through demand components calibration. J. Water Resour. Plann. Manage., 2016, 142(2): 04015057.

Shastri, Y. and Diwekar, U. (2006). Sensor placement in water networks: a stochastic programming approach. Journal of Water Resources Planning and Management, 132, 3, 192–203.

Soldevila, A., Blesa, J., Fernandez-Canti, R. M., Tornil-Sin, S. and Puig, V. (2019). Data-Driven Approach for Leak Localization in Water Distribution Networks Using Pressure Sensors and Spatial Interpolation. Water, 11(7):1500.

Sophocleous, S., Savić, D., & Kapelan, Z. (2019). Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction. J. Water Res. Plan. Man. 145(7):04019024.

Steffelbauer, D. and Fuchs-Hanusch, D. (2016). "Efficient sensor placement for leak localization considering uncertainties". Water Resources Management, 30, 14, 5517–5533.

Steffelbauer, D., Deuerlein, J., Gilbert, D., Abraham, E. and Piller, O. (2022). Pressure-Leak duality for leak detection and localization in water distribution systems. J. Water Resour. Plann. Manage., 2022, 148(3): 04021106.Wang, Q., Savic, D. A. and Kapelan, Z. (2017). GALAXY: A new hybrid MOEA for the optimal design of Water Distribution systems, Water Resour. Res. 53:1997-2015.

Wang, X., Guo, G., Liu, S., Wu, Y., Xu, X. and Smith, K. (2020). Burst detection in district metering areas using deep learning method. J. Water Resour. Plann. Manage., 2020, 146(6): 04020031.

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, DOI: 10.1080/1573062X.2017.1279191

Wu, Y., Liu, S. and Wang, X. (2018). Distance-Based burst detection using multiple pressure sensors in district metering areas. J. Water Resour. Plann. Manage., 2018, 144(11): 06018009.

Xu, W., Zhou, X., Xin, K., Boxall, J., Yan, H., and Tao, T. (2020). Disturbance extraction for burst detection in water distribution networks using pressure measurements. Water Resources Research, 56, e2019WR025526. https://doi.org/10.1029/2019WR025526.

Zhang, Q., Zheng, F., Kapelan, Z., Savic, D., He, G. and Ma, Y. (2020). Assessing the global resilience of water quality sensor placement strategies within water distribution systems. Water Research, Volume 172, 2020, 115527, ISSN 0043-1354, https://doi.org/10.1016/j.watres.2020.115527.

Zimmerman, D., Pelvic, C., Ruggles, A. and Armstrong, M. P. (1999). An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting, Math. Geol. 31(4),375–390.

**Table 1**: Leak/burst event sizes and leak/burst event grouping results for the validation DMA

| **Leak/burst event size (emitter coefficient ($l/s/m^{0.5}$)** | **Number of valid leak/burst event groups** |
|---|---|
| 0.1 | 3 |
| 0.2 | 437 |
| 0.3 | 469 |
| 0.4 | 18 |
| 0.5 | 4 |
| 0.6 | 2 |
| 0.7 | 1 |
| **Total** | **934** |

**Table 2**: Results for the optimal sensor configurations for the 16 engineered events in validation DMA (% of DMA search area). Events in bold were correctly localised

| **Engineered Event Number** | **Number of Optimal Sensors** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **1** | **38.7** | 33.2 | **36.2** | **61.2** | **61.2** | **61.1** | **61.3** | **61.3** |
| **2** | 30.7 | **29.5** | 30.7 | **9.5*** | **9.5*** | **15.4*** | **13.5*** | 0.8 |
| **3** | 23.4 | 20.8 | 23.4 | 2.4 | 2.4 | 9.9* | 10.1* | 14.4 |
| **4** | 39.4 | **22.1** | **35.5** | 6.2 | 6.1 | 14.3 | 15.6 | 18.4 |
| **5** | 30.7 | **65.9** | 25.1 | **20.9** | **20.9** | **17.6** | **18.7** | **11.9** |
| **6** | 28.6 | 28.4 | 25.1 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| **7** | **28.6** | 28.9 | **25.1** | 8.6* | 8.6* | 5.6 | 5.8 | 5.8 |
| **8** | **23.4** | 31.5 | **18.5** | **23.3*** | **23.3*** | **18.0*** | **20.1*** | **14.2*** |
| **9** | **28.3** | 29.5 | **25.1** | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| **10** | 28.3 | 28.4 | 25.1 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| **11** | **28.6** | **27.1** | 25.1 | 8.6* | 8.6* | 5.6 | 5.8 | 5.8 |
| **12** | 28.3 | 21.4* | **35.5** | 20.9 | 20.9 | **39.7** | **39.9** | **31.1*** |
| **13** | 30.7 | **29.5** | 25.0 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| **14** | **30.7** | 21.7 | **65.0** | 8.5* | 6.1 | 8.1* | 8.2* | 8.2* |
| **15** | 30.7 | **65.9** | **37.9** | 29.4* | 29.4* | 23.3* | 26.9* | 20.1* |
| **16** | **54.3** | 29.5 | **43.6** | **28.9** | **28.6** | **23.4** | **23.5** | **20.4** |
| **Average Search Area** | 33.2 | 40.0 | 35.8 | 28.8 | 28.7 | 29.3 | 29.5 | 27.8 |
| **Correct Events** | 7 | 6 | 9 | 5 | 5 | 6 | 6 | 5 |
| **Events with multiple search areas** | 0 | 1 | 1 | 6 | 5 | 5 | 6 | 4 |
| **Equivalent objective function value** | 70.8 | 77.5 | 63.9 | 88.2 | 88.2 | 83.9 | 84.0 | 87.1 |