Advanced Engineering Informatics 51 (2022) 101484

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Sensor placement for robust burst identification in water systems: Balancing modeling accuracy, parsimony, and uncertainties

Lu Xing [a], Tal Raviv [b], Lina Sela [a,*]

[a] Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Texas 78712, United States of America
[b] Department of Industrial Engineering, Tel Aviv University, Tel Aviv, Israel

ARTICLE INFO

Keywords:
Planning and design
Sensor placement
Transient hydraulics
Uncertainty

ABSTRACT

Urban water systems are seeing an uptake in using advanced sensing technology. Incorporating sensors for monitoring water distribution systems (WDSs) provides promising benefits to water utilities by enabling a shift from reactive to proactive pipe failure detection and from delayed water loss management to automatic sense-and-respond capabilities. Despite the opportunities that new sensing technologies create, a budget-constrained utility is challenged with identifying sensing locations in the WDS that will maximize information gain. To address this gap, this paper studies the problem of optimal placement of high-frequency pressure sensors in WDSs for pipe burst identification. This paper proposes a sensor placement strategy to address the challenges of data and modeling uncertainty by incorporating robust representation and tolerance analysis into an optimization framework with the objective of achieving the best detection and identification of burst events. Transient simulations are first used to predict system's response to burst events, demonstrating the importance of modeling accuracy over approximation methods. A robust event representation approach is then presented to summarize system response to pipe bursts using signature matrices. Subsequently, the identification problem is cast as a minimum test cover problem when the number of available sensors is unlimited, and as the maximum covering test problem when the number of available sensors is limited. The optimization problems are then formulated and solved using mixed integer linear programming. Four complementary metrics are suggested to evaluate the performance of the sensor placement designs. Multiple criteria decision analysis is then applied to select the placement design while balancing information gain and cost. The results show that incorporating more information can improve event identification, but sufficient accuracy of the extracted information is required to accrue the benefits.

1. Introduction

Water distribution systems (WDSs) are complex lifeline infrastructures, essential for providing safe and reliable drinking water to the growing urban population. As pipelines in WDSs age, the decreasing structural integrity and transmission capacity result in increasing occurrences of pipe failures, which disrupt water supply and waste a significant amount of treated water [1]. WDSs have been reported to waste approximately 20% to 30% of the treated water through pipe failures [2]. In addition to water loss, pipe failures create unintended opportunities for contamination intrusion and potentially disturb the operation of other infrastructures [3,4]. Monitoring WDSs and integrating the data collected from distributed sensing devices has been identified as one of the prominent strategies to detect and identify pipe failures, thus minimizing water losses and service interruptions. Various pipe failure detection and identification techniques have been developed, including visual inspection [5,6], acoustic-based methods [7–10], ground penetrating radar [11,12], thermography [13], as well as methods based on steady state hydraulic modeling [14–17] and transient-based methods [18–25].

Pipe failures are commonly categorized into background leaks and bursts [26]. Background leaks refer to preexisting and persistent water losses, often through pipe joints and cross connections at low flow rates, whereas pipe bursts refer to sudden pipe rupture and break events [23]. In this paper, pipe bursts are the primary events of interests (EoIs). Pipe bursts can introduce sudden and rapid disturbances to the flow conditions, which propagate through the system as pressure waves, i.e., pressure transients, with very high velocity in the range of 600–1500 m/s [27]. Transient-based methods have received increasing attention in the past two decades because a considerable amount of information about the WDS can be revealed within a very short period of time as the transient wave propagates through the network

* Corresponding author.
E-mail addresses: xinglu@utexas.edu (L. Xing), talraviv@tauex.tau.ac.il (T. Raviv), linasela@utexas.edu (L. Sela).

https://doi.org/10.1016/j.aei.2021.101484
Received 14 July 2021; Received in revised form 9 November 2021; Accepted 24 November 2021
Available online 14 December 2021

[27–29]. Thus, by monitoring for transient pressure at various locations of the WDS, these transient pressure waves can be detected and the origin of the pipe burst can be identified. Monitoring for transient pressure can be enabled by high-resolution pressure sensors that are distributed in WDSs [30,31]. Due to the fast evolving hydraulics, traditional methods such as supervisory control and data acquisition (SCADA) systems, which collect data with 5–15 min resolution, are inadequate to capture the rapidly changing transient system dynamics [23]. Flow meters can additionally provide useful information for detecting pipe burst; however, flow meters are expensive, require direct contact with the pipe, and do not react instantaneously to changes in flow, and thus, are not appropriate for high-resolution monitoring [28]. Other sensing mechanisms, such as surface and inline sensing mentioned above, are better suited for local inspection and are not suitable for continuous operation.

Majority of the current transient-based methods are designed to detect background leaks by manually inducing transient events through, for example, manipulating valve operations, and investigating the impact of existing leaks on the shape of the transient response of the system [18,19,21,32]. However, these methods require meticulous design of the experiment procedure and are not suited for detecting abrupt events in real time. Motivated by the growing needs for real-time burst detection, recent studies proposed to investigate the pressure transients generated by pipe bursts themselves [22,23,25,33–35]. These studies have shown the potential of using pressure transients for timely pipe burst detection. However, previous works are primarily limited to pipe segments, small networks, or transmission mains, and were not implemented in water networks comprising multiple pipes with complex topology. Moreover, the accuracy of pipe burst detection method heavily depends on the number and locations of pressure sensors; hence, it is imperative to design a sensor network that maximizes information gain. Although several studies for optimal sensor placement for leak detection have been proposed, the majority of these studies consider the placement of low temporal resolution sensors and model the leaks using steady-state hydraulic models [33–37]. Other studies have relied on distance-based approximation methods [38–41]. Thus, in the context of sensor placement for burst detection, a gap remains in including a better representation of the WDS dynamics under pipe burst events to achieve a sensor placement that enables accurate, robust, and timely detection of pipe bursts.

To address this gap, we simulate the adverse impacts of pipe bursts using *transient hydraulics*, which represents the physical conditions in a WDS more realistically, compared to distance-based or the steady-state hydraulic models. Further, we consider *event identification* as the objective of the sensor placement problem, as opposed to the more common event detection. The detection problem maximizes the number of EoIs that can be detected, while the identification problem maximizes the number of EoIs that can be uniquely identified [42,43]. Considering the identification objective is beneficial because of the ability to pinpoint the location of pipe bursts, which can significantly reduce the response time and overall costs of recovery [40,44]. However, the ability to uniquely identify events typically comes at the expense of relying on more data, solving a more complex problem, and ultimately requiring deploying a greater number of sensors augmented with advanced techniques for data processing [45–47].

The combination of transient modeling coupled with event identification introduces new challenges in addressing modeling and data uncertainty, which can significantly impact the sensor placement design. These uncertainties generally originate from two sources: (1) data preparation and acquisition, i.e., the process of obtaining data either from computational models during the design phase or sensor measurements after deployment, and (2) information extraction, i.e., the process of extracting information from the data to inform design and decision-making. Firstly, during the design phase of the distributed sensor systems, the acquisition of data typically relies on computational models, which simulate the behavior of the physical system; however, uncertainties associated with the computational models are inevitable due to the inadequacies, such as model formulation and the incomplete information of model parameters. Over-reliance on the data generated by the computational models can potentially lead to inappropriate choices regarding optimal sensor locations. Secondly, the accuracy of the sensory data is restricted by the sensor precision (i.e., the degree of reproducibility of a measurement), sensor resolution (i.e., the smallest detectable change of the pressure signal), and calibration accuracy [48]. To exacerbate the situation, sensors are error prone and can give erroneous outputs due to degradations of sensor hardware or software components. The limitation of sensor accuracy and erroneous readings contribute to the data uncertainties, thus making it unreasonable to rely on the raw sensory data to identify the EoIs.

We propose a two-fold method to address the challenges of *model and data uncertainty* by incorporating robust signal representation and tolerance analysis. Firstly, the data uncertainties are modeled by robustifying the data representation, such as extracting characteristic features that are robust enough to account for the uncertainties, yet still contain enough valuable information to inform design and decisions. However, certain levels of data uncertainties still remain in the robust representation and additional uncertainties are also introduced during information extraction. Hence, in addition to the robust representation, a tolerance analysis is introduced to take an agnostic view to account for the fact that the extracted information and problem input are not exact [49]. We then formulate the robust sensor placement problem as mixed integer linear programming (MILP), which can explicitly address the challenges imposed by data and model uncertainties through robust representation and tolerance analysis and can achieve solutions with high performance guarantees. Multiple criteria decision analysis (MCDA) is then applied that incorporates decision-maker preferences for balancing information gain and cost and can facilitate in selecting the best sensor design [50].

The rest of the paper is organized as follows: in Section 2, we provide a brief literature review on sensor placement for pipe failure detection and identification in WDSs. In Section 3, we define the problem of event identification and give an illustrative example. In Section 4, we introduce the hydraulic transient model for simulating bursts and propose signature-based matrices to represent the EoIs. Subsequently, we discuss the difference between detection and identification problems, and then formulate the identification problem as the minimum test cover (MTC) problem when the number of available sensors is unlimited and as the maximum covering test (MCT), where the number of available sensors is limited [51]. We propose four metrics to evaluate the performance of the sensor placement design. Additionally, we introduce the Preference Ranking Organization METHod for Enrichment Evaluation (PROMETHEE) method [50] to determine the best number of sensors. In Section 5, we apply the proposed scheme to pipe burst identification in a WDS and demonstrate the benefits of enriching the signature matrix with more information and the influence of uncertainties on the identification performance. Finally, Section 6 concludes the paper and proposes several potential future extensions.

2. Literature review

Various methodologies have been previously proposed for sensor placement in WDSs for the detection and identification of pipe failures, where majority previous works cast the sensor placement problem using a simulation–optimization framework with the objective to optimize different performance measures [14,33,34,34–37,40,52–55]. In the simulation–optimization framework, hydraulic models, typically steady-state solvers such as EPANET [56], are first utilized to simulate the impacts of the EoIs. In the second stage, the problem of optimal sensor design is formulated and solved based on system responses simulated in the first stage. Ultimately, previous works differ in the choice of simulation models, optimization objectives, event representation, and optimization algorithms. We briefly review previous studies in

terms of modeling event dynamics, event representation methods, and optimization algorithms.

Modeling event dynamics. Pressure waves induced by pipe bursts propagate rapidly in the network, thus implying that steady-state analysis of system hydraulics is not adequate to capture system response to burst events and that transient system dynamics should be considered [23]. However, due to the modeling complexity of transient hydraulics, the majority of previous works have relied on steady-state analysis to estimate system response to leak events, e.g., by modeling pressure driven discharge using emitters at leak locations and using EPANET simulations to estimate changes in nodal pressures in response to leak events [33–35]. Acknowledging the limitation of steady-state models, Deshpande et al. [38], Abbas et al. [39], Sela Perelman et al. [40] and Sela and Amin [57] proposed to simulate the network dynamics using transient wave propagation models. However, due to the lack of open-source and application programming interface for transient modeling, these studies adopted a simple distance-based model to approximate the dissipation of the pressure signal in the network. The major assumption of the distance-based models is that the disturbance caused by bursts will dissipate as it travels in the network, such that the disturbance in pressure can be sensed within a specified distance from the location of the burst, i.e., any sensors deployed within a certain distance from the origin of the burst can detect that burst, while sensors located farther away cannot detect the burst. However, this approximate model is inadequate for modeling the complex dynamics of pressure wave reflection, transmission, and propagation in the network. To illustrate the necessity of using transient simulation for modeling pipe bursts, we performed simulations using the steady-state hydraulic model as proposed in [14,15], distance-based approximation as in [39,40,57], and the newly developed open-source transient simulation package, TSNet [58]. The comparison and further discussion are presented in Section 5.1.

Event representation. The essence of sensor placement for event identification is to select sensors that are most informative for the task of distinguishing the impacts of different EoIs [59]. If different EoIs have different impacts on the system, these EoIs can be distinguished and identified. Thus, an essential step is to summarize the impact of EoIs using an efficient and robust representation. The common practice is to encode the impacts of EoIs into a boolean representation $\{0, 1\}$, based on whether or not the EoI has a visible change on the system as compared with its normal status [60]. For example, in [14], the impacts of leaks were summarized using the leak sensitivity matrix, which represented the difference between the measurements and their estimates using the steady-state hydraulic models; it follows, then, that the leak sensitivity matrix was binarized by comparing the estimates with a chosen threshold. In [39,40], the impacts of EoIs were represented using binary values indicating whether a sensor can detect a pipe burst depending on its distance from the location of the burst. In [41], categorical values were extracted to represent the different events, assuming that the outputs produced by the sensors are discrete. On one hand, these simplified representations conceal some of the uncertainties in the raw sensory data, and thus unintentionally enhance the robustness of the sensor placement design. On the other hand, valuable information is also lost through this simplification. The boolean representation only answers whether or not an EoI has a visible impact, but gives no additional information about its occurrence, e.g., magnitude of the impact. This choice of representation can be partially attributed to the previous limitations in sensing and modeling techniques. As an improvement on Pérez et al. [14], Casillas et al. [36] proposed a non-boolean leak sensitivity matrix and the projection-based leak isolation approach; however, the proposed representation approach resulted in a non-linear optimization problem, which is challenging to solve. Thus, an improved but parsimony event representation that does not impair problem complexity is imperative in order to realize full benefits of the high quality data collected by advanced sensors.

Optimization algorithms. Various optimization methods have been proposed for solving the sensor placement problem, with most prominent including the greedy approximation [37,40,52], mixed integer linear programming (MILP) [33,38,57], and evolutionary algorithms [34,36,55]. Evolutionary algorithms, such as genetic algorithms, simulated annealing, and tabu search, are widely adopted due to their flexibility of dealing with nonlinear problems. However, these heuristic approaches cannot guarantee convergence to the global optimal and are computationally expensive [61]. The advantages of MILP formulations are clear: mixed integer formulations coupled with modern solvers can guarantee solution optimality and are computationally efficient even for large scale problems. When modeling complexity or problem size hinders MILP formulation, other optimization approaches, such as greedy approximation or evolutionary algorithms, should be pursued [57,62].

3. Problem definition

In this work, we formulate the optimization problem as to find the sensor locations that achieve the best identification of pipe burst events with limited or unlimited budget, where event identification refers to the ability to distinguish between different EoIs. Specifically, we define a set of candidate sensors as $S = \{s_1, s_2, \dots, s_{N_S}\}$, where N_S is the number of candidate sensors, and s_i denotes the location of the ith sensor. We also assume that the sensors are continuously monitoring pressure (p). Similarly, a set of possible EoIs, i.e., bursts, is denoted as $E = \{e_1, e_2, \dots, e_{N_E}\}$, where N_E is the number of potential EoIs, and e_j symbolizes the attributes of the jth EoIs, such as its location and magnitude. The continuous pressure recording at sensor location s_i given that event e_j has occurred is represented as a time-series with discrete timestamps, i.e., $\boldsymbol{p}(s_i \mid e_j) = (p_{ij}^1, p_{ij}^2, \dots, p_{ij}^{n_t})$, where the subscripts represent the location where the pressure is measured and the location where the event has occurred, respectively, the superscripts stand for the time when the pressure is measured, and n_t is the total number of time stamps.

It is assumed that EoIs induce changes in the pressure; thus, the occurrence of EoIs can be recognized by monitoring the changes in pressure, making burst detection feasible. Notably, precision and accuracy limitations introduce uncertainties in the transient pressure. The main idea behind event identification is to compare the changes recorded by different sensors, such that combinations of changes recorded by several sensors can uniquely pinpoint the occurred EoIs, while accounting for the uncertainties.

3.1. Illustrative example

To illustrate the sensor placement problem for event identification and challenges imposed by the uncertainties, consider a simple example with two candidate sensors $S = \{s_1, s_2\}$ and three potential events $E = \{e_1, e_2, e_3\}$. The goal in this example problem is to choose one sensor such that maximum number of events can be identified, i.e., distinguished from each other, by analyzing the pressure measured by the selected sensor. The normalized pressure measured by s_1 in the event of e_1, e_2, and e_3 are denoted as $\boldsymbol{p}(s_1 \mid e_1)$, $\boldsymbol{p}(s_1 \mid e_2)$, and $\boldsymbol{p}(s_1 \mid e_3)$, respectively, and are depicted in Fig. 1(a). Fig. 1(b) illustrates the normalized pressure observed by s_2 under the occurrence of the three events, i.e., $\boldsymbol{p}(s_2 \mid e_j),\ j = 1, 2, 3$.

It can be observed that all the six normalized pressure time series, i.e., $\boldsymbol{p}(s_i \mid e_j),\ i = 1, 2;\ j = 1, 2, 3$, are different from each other; thus, ideally, both sensors can distinguish the three events. However, further inspection of Fig. 1(a) reveals that $\boldsymbol{p}(s_1 \mid e_1)$ and $\boldsymbol{p}(s_1 \mid e_2)$ are fairly similar to each other. Intuitively, this similarity indicates that if uncertainties in sensor measurements, information extraction, and modeling, are taken into account, it is likely that the changes induced by events e_1 and e_2 will be indistinguishable by sensor s_1. In other words, it is possible that the difference between $\boldsymbol{p}(s_1 \mid e_1)$ and $\boldsymbol{p}(s_1 \mid e_2)$

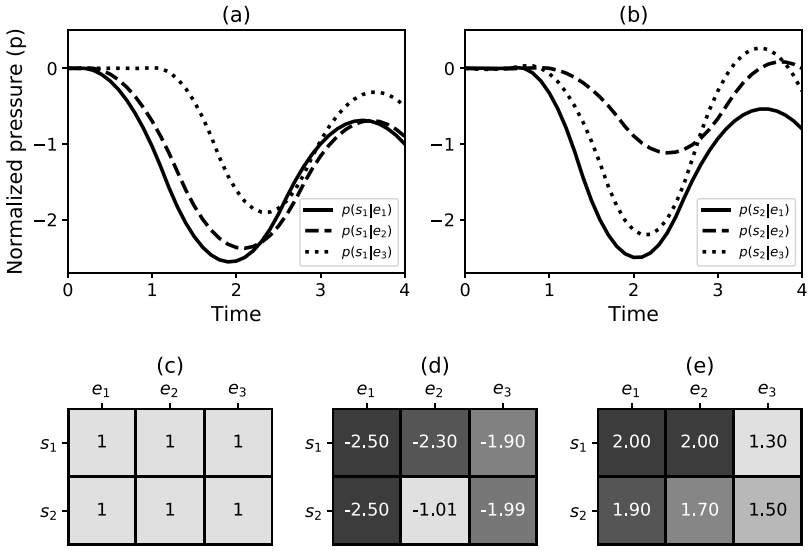

Fig. 1. An illustrative example for robust representation and sensor placement: (a) normalized pressure at s_1 under the occurrence of e_1, e_2, and e_3; (b) normalized pressure at s_2 under the occurrence of e_1, e_2, and e_3; (c) boolean signature matrix; (d) signature matrix with amplitude of the first change as the characteristic feature; and (e) signature matrix with duration of the first change as the characteristic feature.

is not significant enough to overcome the uncertainties, thus suggesting that s_1 alone cannot distinguish between events e_1 an e_2. The rather apparent difference between $\boldsymbol{p}(s_1 \mid e_1)$ and $\boldsymbol{p}(s_1 \mid e_3)$ implies that e_1 and e_3 should be distinguishable by analyzing the pressure observed by s_1. On the other hand, the time series recorded by s_2, $\boldsymbol{p}(s_2 \mid e_j)$ for the three events $j = 1, 2, 3$, appears to be different from each other as shown in (b), which suggests that s_2 can distinguish between all three events. Thus, this qualitative analysis indicates that s_2 should be chosen as the best one-sensor design.

This simple example qualitatively illustrates that the existence of uncertainties discourages the practice of determining whether two pressure time series are distinguishable by a naive point-wise comparison. However, the challenge remains on quantifying and incorporating these uncertainties. In the next section, we propose a parsimony representation of the time series that is robust against uncertainties and, at the same time, preserves the key features of the full signal.

4. Methodology

In this section, we propose a novel methodology for sensor placement problem to achieve optimal event identification performance under model and data uncertainties. The proposed approach consists of four main steps. First, pipe burst events are simulated using transient hydraulics. Second, characteristic features that represent system's response to the EoIs are extracted from the transient pressure signals and are represented using signature-based matrices. Third, a tolerance level for which the system is assumed to produce equal response is specified, and two optimization problems are formulated and solved: specifically, finding the minimum number of sensors such that all events can be identified, and selecting a subset of sensors with limited budget such that maximum number of events can be identified. Then, the sensor placement design is evaluated using four complementary performance metrics. Finally, multiple criteria are systematically accounted for to determine the optimal sensor placement.

4.1. Modeling transient hydraulics

WDS response to burst events can be modeled using hydraulic transient flow, which, for a single pipeline, can be described using a system of partial differential equations characterizing mass and momentum conservation [63], as follows:

$$\frac{\partial H}{\partial t} + \frac{a^2}{g}\frac{\partial V}{\partial x} - V \sin\alpha = 0 \tag{1}$$

$$\frac{\partial V}{\partial t} + g\frac{\partial H}{\partial x} + h_f(f, V) = 0 \tag{2}$$

where H is the hydraulic head, V is the flow velocity in the pipe, t is time, a is the wave speed, g is the gravity acceleration, f is the pipe friction coefficient, α is the pipe slope, and h_f represents the head loss model, which is a function of flow velocity and friction coefficient.

For a networked system, Eqs. (1) and (2) are discretized in time and space and can be solved using the method of characteristics (MOC) [63]. The essence of MOC is to transform the set of partial differential equations to an equivalent set of ordinary differential equations (ODEs) applicable along specific numerical grid lines, i.e., characteristics lines, which represent the directions in which the disturbance in a pipe propagates. Once the MOC characteristic grid and numerical scheme are established, the explicit time marching MOC involves computing the head and flow velocity, H_t^i, V_t^i, at new point in time and space given that the conditions at the previous time step are known. To model various boundary conditions, such as pipe connections, valves, pumps, reservoirs, as well as leaks and bursts, auxiliary equations, such as continuity, work-energy principles, are supplemented to characterize the flow and head behavior at the boundary [63].

In this work, a burst is modeled using the conservation of mass and momentum equations coupled with the orifice equation [63], $Q_b(t) = k_b(t)\sqrt{H_{p_b}(t)}$, where $H_{p_b}(t)$ is the pressure head at the location of the burst, $Q_b(t)$ is the pressure-dependent burst discharge, and $k_b(t)$ is the lumped burst coefficient, which aggregates the size of the burst, units, and burst coefficients at time step t. Before burst occurs, $k_b(t)$ equals zero and it increases with time as the burst develops. It should be noted that a background leak can be modeled using the same orifice equation. The difference is that k_b is constant when modeling preexisting leaks, while for bursts $k_b(t)$ is a time-dependent function. In other words, the burst is the process of first occurrence of the pipe failure, while the background leak is a preexisting condition [23]. We utilize TSNet, the recently developed open-source python package for transient simulations in water networks [58], to simulate changes in pressures in response to pipe bursts at different location in a WDS. For details on the numerical scheme of burst simulation in WDSs, the reader is referred to Xing and Sela [58].

4.2. Event representation

As discussed in the previous section, transient simulations can be performed to predict pressure at a sensing location, s_i, given that a

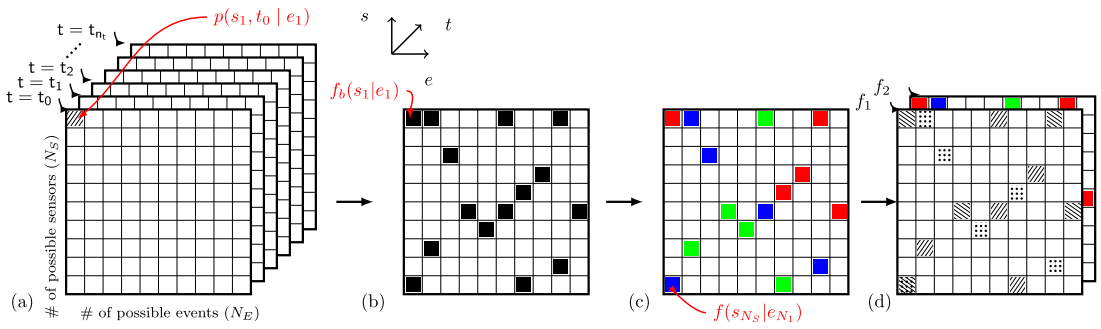

Fig. 2. Event representation: (a) full signal tensor (FST), where element (i, j, k) represents the signal observed at sensor s_j at time t_k under the occurrence of event e_i; (b) boolean signature matrix (SM^B), where element (i, j) represents the boolean signature at sensor s_j under the occurrence of the event e_i (black pixel $f^b(s_j \mid e_i) = 1$, white pixel $f^b(s_j \mid e_i) = 0$) (c) single feature signature matrix (SM), where element (i, j) represents the signature at sensor s_j under the occurrence of the event e_i (different colors represent different signature values), and (d) combined signature matrix with two features, f_1 and f_2 (different colors and fill patterns represent different signature values). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

burst occurs at junction j, i.e., $\boldsymbol{p}(s_i \mid e_j)$. The simulation results contain information about systems response to the EoIs, as recorded by the sensors. The question follows is how to represent this information in an adequate, efficient, and robust manner. In this section, we first elaborate the necessity of transforming the full time series representation to lower-dimension robust representation. Subsequently, we propose a robust event representation method based on extracting characteristic features.

4.2.1. Full signal representation

To begin with, the most intuitive way to represent the impact of all possible bursts on the WDSs is to store the pressure signals at all candidate sensor locations under the occurrence of each possible burst. Thus, by enumerating over all the potential events ($e_j \in E$) and sensor locations ($s_i \in S$) during a given period of time (n_t simulation time steps), a full signal tensor (FST) of $N_E \times N_S \times n_t$ dimension can be assembled where $FST(i, j, k) = \boldsymbol{p}(s_j, t_k \mid e_i)$, $k = 0, 1, 2, \ldots, n_t$. The FST encompasses the expected observations of all candidate sensors during the occurrence of all potential events. The structure of FST can be illustrated as a 3-dimensional array, as shown in Fig. 2(a), where each element is defined by three indices: sensor, event, and time index. For example, the shaded element can be accessed as the $FST(1, 1, 0)$, representing the pressure recorded by sensor s_1 at time t_0 given that event e_1 occurred.

Although the FST contains all the information about response of the entire system to the potential EoIs, it is not the ideal input to the sensor placement problem for several reasons. Firstly, to capture the rapidly changing hydraulic transients, high temporal resolution in the magnitude of 0.001–0.1 s is required in transient simulations, which results in a large number of simulation time steps, n_t. Additionally, the number of potential EoIs, N_E, and sensor candidates, N_S, is typically large in real-life systems. Thus, the dimension of $FST^{N_E \times N_S \times n_t}$ can easily get impractical. For example, consider a WDS with 1000 possible burst locations and 1000 sensor candidates with a sampling frequency of 64 Hz. To represent a ten-minute window of data, the $FST^{1000 \times 1000 \times 38{,}400}$ requires around 307 GB of RAM when stored as a full array. Secondly, uncertainties in model parameters and inadequacies of the numerical models can lead to discrepancies between model estimations and actual observations, which indicates that two sensors, exhibiting small differences in model estimations, may not have distinguishably different readings in practice. Thirdly, due to the complicated nature of the sensing, telecommunication, and power in a sensing unit, erroneous data is inevitable [64], and thus, we cannot rely on the sensors to record perfect data at all times. Hence, with the presence of model and measurement uncertainties, robust signal representation should be used instead of the full time-series. The intuition behind robust signal representation is that the pressure signals should be represented in a way such that the impacts of EoIs can be characterized and preserved without overemphasizing the details, which can be unreliable due to model and data uncertainties. This is illustrated in the above example, where the differences between $\boldsymbol{p}(s_1 \mid e_1)$ and $\boldsymbol{p}(s_1 \mid e_2)$ may not be significant enough to overcome the model and measurement uncertainties (Fig. 1(a)).

4.2.2. Reduced signal representation

Instead of the full representation, we propose to extract the representative features of the transients, such as the amplitude and duration, which characterize the transients and can be extracted using various detection algorithms [65]. If η features are to be extracted, i.e., $f_1, f_2, \ldots, f_\eta$, the observed quantity within a certain time window can then be reduced from a time-series to a discrete representation using the extracted features, $[f_1, f_2, \ldots, f_\eta]$, where $\eta \ll n_t$.

The simplest feature is a boolean representation, where $f^b(s_j \mid e_i) \in \{0, 1\}$ is equal to zero if no changes are detected at sensor s_j when event e_i occurs, and equal to 1 otherwise. This can be achieved by examining whether the detected changes are significant enough as compared with some expected or empirical values [14]. In the example shown in Fig. 1(a), changes can be detected, for example using the cumulative sum (CUSUM) algorithm [66] (as will be detailed in Section 5) in all three signals, indicating that s_1 can detect all the three events; thus, the impact of the each event at sensing station s_1 can be reduced from $\boldsymbol{p}(s_1 \mid e_j), j \in [1, 2, 3]$ to the boolean representation. Subsequently, the *boolean signature* of event e_i is defined as the boolean vector of the outputs of sensors in the set S under the occurrence of e_i, and denoted as $\boldsymbol{f^b}(S \mid e_i) = [f^b(s_1 \mid e_i), f^b(s_2 \mid e_i), \ldots, f^b(s_{N_S} \mid e_i)]$. Consequently, for a sensor set S and the set of potential events E, we can instantiate a *boolean signature matrix* of dimension $N_S \times N_E$, denoted by SM^B that summarizes the impacts of all EoIs on the WDS. The structure of SM^B is illustrated in Fig. 2(b). The ith row of SM^B comprises of sensor s_i responses to all potential events, and the jth column represent the boolean signature of the event e_j, i.e., $\boldsymbol{f^b}(S \mid e_j)$. Moreover, $SM^B(i, j) = 1$, shown as black pixels in Fig. 2(b), indicates that the sensor s_i detected the event e_j, while $SM^B(i, j) = 0$, shown as white pixels, suggests otherwise. However, valuable information contained in the original signal is lost under the boolean representation. For example, the SM^B for the illustrative example is a 2×3 matrix with all elements equaling one (as shown in Fig. 1(c)), which is uninformative for the task of distinguishing between the three EoIs.

To preserve more information, additional characteristics features, $\boldsymbol{f}(s_j \mid e_i) \in \mathbb{R}^\eta$, of the changes in the time series can be extracted and exploited. For example, two features ($\eta = 2$), the amplitude (dp) and duration (dt) of the first change in the signal, can be extracted from $\boldsymbol{p}(s_1 \mid e_j), j \in [1, 2, 3]$, such that the EoIs can be characterized by the enriched two-element representation, i.e., $[dp(s_1, j), dt(s_1, j)], j \in [1, 2, 3]$, where the two indices denote the sensor and event index, respectively. Now, for the pressure signal recorded by a given sensor

under a certain event, the original time series can be reduced to the two-dimensional vector, transforming the representation from the high-dimensional temporal space to the low-dimensional feature space. The number of features that can be reliably extracted from the signal depends on the expected data quality that will be collected by sensors, where the higher the data quality, the more features can be extracted. The extracted features can then enrich the signature matrix, which is referred to as the *signature matrix* (SM). The structure of signature matrix is shown in Fig. 2(c), where different colors and fill patterns represent different values of f. If no changes are detected by sensor s_j at the occurrence of event e_i, we set $f(s_j \mid e_i) = 0$, as denoted by the white pixels. For example, in the illustrative example in Fig. 1, the amplitude of the change is extracted to formulate the signature matrix ($SM^{2\times3}$) as shown in Fig. 1(d), where the element (i, j) represents the amplitude of the change as recorded by s_i under the occurrence of e_j. Unlike the SM^B in Fig. 1(c), the elements of the SM shown in Fig. 1(d) are all different each other, indicating that this SM is more informative for the task of distinguishing between the EoIs. Moreover, multiple features can be combined to enrich the representation, as shown in Fig. 2(d), where each element is represented by two features, $[f_1, f_2]$, respectively.

4.3. Robust event identification

Following the event representation methods, we define and formulate the robust event identification problem. To begin with, it is recognized that even the continuous signature matrix may still not be robust to modeling and measurement uncertainty. For example, relatively small difference exits between $SM(1,1)$ and $SM(1,2)$ in Fig. 1(d). Hence, we would like to introduce tolerance analysis to further robustify the event identification problem. Intuitively, the idea for the tolerance analysis relies on examining whether the difference between the values of pair-wise continuous features is significant enough in terms of producing distinguishable system responses to the EoI. In other words, two signals cannot be distinguished if their characteristic features are within a specified tolerance [49]. Specifically, if the signatures of two events e_j, e_k at a given sensor s_i are within the tolerance interval as specified by the *identification tolerance* δ, i.e., $\mid SM(i,j) - SM(i,k) \mid \leq \delta$, then s_i cannot distinguish between e_j and e_k. The identification tolerance is preferred over categorization to avoid arbitrary groupings that eliminate coherent information [14,47,67,68]. In the context of event identification, the proposed representation makes sense, because the main concern is related to how the response features compare with each other rather than focusing on the exact values of the characteristic features. The identification tolerance indicates the overall uncertainty level, which is conditional on the confidence in the precision of sensory data, the adequacy of the simulation model, and the accuracy of the detection method. As the uncertainty level decreases and confidence in sensory data and modeling increases, the identification tolerance can be reduced accordingly, such that more signatures can be distinguished. For example, if the precision (or typical noise level) of the pressure sensor is ±3 m then the comparison between two signals is only relevant if the difference is greater than ±3 m, i.e., the tolerance is $\delta_{dp} = 3$ m; however, if the precision is greater, e.g., ±1 m, then the tolerance can be decreased, e.g., $\delta_{dp} = 1$ m, since we are more confident in our measurements.

Recall the illustrative example, where the value of change amplitude observed by s_2 under the occurrence of e_1, e_2 and e_3 was $\boldsymbol{f}(s_2 \mid e_1, e_2, e_3) = [-2.5, -1.01, -1.99]$. If the identification tolerance is specified at $\delta = 0.6$, the tolerance analysis will indicate that e_1 and e_3 are indistinguishable; while if the identification tolerance is decreased to $\delta = 0.5$, all three EoIs can be distinguished. A greater value of δ will cause more EoIs to be deemed as indistinguishable, thus increasing the robustness but sacrificing identification performance. Thus, with the trade-off between robustness and performance in mind, δ should be chosen according to the expected accuracy of the model and observations. We note that with boolean representation distinguishable events imply that $\delta = 1$.

In this study, the robust tolerance analysis is applied on the signatures of the EoIs to examine whether the pair-wise events can be distinguished. Specifically, for the set of events E and the set of sensors S, the set of all pair-wise events that can be distinguished by s_i is denoted as C_i^I and defined as:

$$C_i^I = \{(e_j, e_k) \in E : \ |SM(i,j) - SM(i,k)| \geq \delta\} \tag{3}$$

where SM represents either boolean or continuous signature matrix, and δ is identification tolerance. If C^I is a collection of all such C_i^I's, i.e., $C^I = \{C_i^I \mid \forall i\}$, then for a given subset of sensors $S \subseteq S$, we define $C_S^I \subseteq C^I$ as a set of subsets of distinguishable pair-wise EoIs, where a subset corresponds to a sensor in S that identifies the pair-wise EoIs in that subset, i.e., $C_S^I = \{C_i^I \mid s_i \in S\}$. Then, pair-wise events are a *distinguishable pair* if $(e_j, e_k) \in C_S^I$. It follows that event e_i is an *identifiable event* if it can be distinguished from all other events, i.e., all pair-wise events $e_i, e_j, \forall j \neq i$ are identifiable.

The set of all identifiable events is then denoted as *identifiable set* E_I. Subsequently, we define the *identification function*, which gives the number of distinguishable event pairs using the subset of sensors S, $\mathcal{F}_I$ as:

$$\mathcal{F}_I(S) = \left| \bigcup_{C_i^I \in C_S^I} C_i^I \right| \tag{4}$$

Event identification is different from the well-studied event detection problem, in which the goal is simply to recognize that some EoIs have occurred. Only the boolean signature matrix is required for the task of event detection. We denote the events that can be detected by sensor $s_i \in S$ as C_i^D, where $C_i^D = \{e_j \in E \mid SM^B(j,i) = 1\}$. A collection of detection sets for all sensors given a sensor subsets S is then defined as $C_S^D = \{C_i^D \mid s_i \in S\}$. The event (e_i) is an *detectable event*, if there exists at least one sensor in S that can detect e_i, i.e, $\exists s_i \in S \ e_i \in C_i^D$. The set of all detectable events is then defined as *detection set* and denoted as $E_D \subseteq E$. The *detection function*, $\mathcal{F}_D(S)$, yields the number of EoIs that can be detected by the sensors in S. It should be noted that any identifiable event is instinctively a detectable event, i.e., the identification set is subset of detection set, $E_I \subseteq E_D$ [42].

Let us revisit the example shown in Fig. 1 and illustrate its detection and identification set. As presented in Fig. 1(c), both sensors can detect all three events; thus, $C_1^D = C_2^D = \{e_1, e_2, e_3\}$. Assuming we use the signature matrix with change amplitude shown in Fig. 1(d) and set the identification tolerance $\delta = 0.5$, the pair-wise events that can be identified by each sensor can be presented as $C_1^I = \{(e_1, e_3)\}, C_2^I = \{(e_1, e_2), (e_1, e_3), (e_2, e_3)\}$. If only sensor s_1 is chosen, all events can be detected, but only event e_1 and e_3 can be identified, i.e., $\mathcal{F}_D(s_1) = 3$ while $\mathcal{F}_I(s_1) = 1$. The detection function output is bounded such that $0 \leq |\mathcal{F}_D(S)| \leq N_E$, where the lower bound corresponds to case where no events can be detected, while the upper bound represents that all the events in E can be detected by the subset of sensor (S). On the other hand, the identification function has a zero lower bound, indicating that no pair-wise events can be identified, and an upper bound $\binom{|E_D|}{2}$, in which any two detectable events can be distinguished from each other.

4.4. Sensor placement for event identification

The underlying idea behind the sensor placement problem for event identification (SP-EI) is to find a collection of sensors that results in a collective output that is unique for a maximum number of EoIs. The essence of SP-EI is to minimize the information loss in the transformation $\mathcal{F}_I(S) \rightarrow \mathcal{F}_I(S)$, where $S \subseteq S$. We consider two different formulations for the SP-EI problem. In the first formulation, the objective is to select the minimum number of sensors, such that every distinguishable pairs can be uniquely distinguished. This formulation is referred to as SP-EI1, and can be formulated as the *minimum test cover problem* [40], which is defined as follows [51]:

Definition 1. Minimum test cover (MTC): Consider a finite set E, of which all item pairs comprise the set $E^P = \{(e_h, e_j) \mid \forall e_i, e_j \in E, h \neq j\}$, and a set of *tests* $T = \{T_i \mid T_i \subseteq E^P, \forall i\}$. A test T_i *covers*, or differentiates, the item pair (e_h, e_j) if $(e_h, e_j) \in T_i$. The minimum test cover is to find $\mathcal{T} \subseteq T$ with the minimum cardinality such that each item pair can be covered by at least one test in $\mathcal{T}$, i.e., $\bigcap_{T_i \in \mathcal{T}} T_i = E^P$, where $|E^P| = \binom{|E|}{2}$.

In the context of event identification, the tests refer to the sensors, and the item pairs covered by the tests refer to the pair-wise events distinguished by the sensor, i.e., $T \rightarrow C^I$. The objective of SP-EI1 is to find a subset $C^I_S \subseteq C^I$ of minimum cardinality, or equivalently the minimum number of sensors $\mathcal{S} \subseteq S$, such that all distinguishable pairs can be distinguished by at least one sensor in $\mathcal{S}$, i.e., $\arg\min_{\mathcal{S}} \left(|\mathcal{S}| \mid \mathcal{F}_I(\mathcal{S}) = \mathcal{F}_I(S)\right)$.

However, in pragmatic sensor placement design, due to budget constraints, the number of available sensors is typically limited and the number of sensors required to distinguish all events is impractical. Thus, an alternative and more realistic objective is to maximize the number of distinguishable event pairs with at most β sensors, i.e., $|\mathcal{S}| \leq \beta$. This formulation is denoted as SP-EI2, and it is equivalent to the *maximum covering test*, which is defined as follows [69]:

Definition 2. Maximum test covering (MCT): Consider a finite set E, of which all item pairs comprise the set $E^P = \{(e_h, e_j) \mid \forall e_h, e_j \in E, h \neq j\}$, and a set of *tests* $T = \{T_i \mid T_i \subseteq E^P, \forall i\}$. A test T_i *covers* or differentiate the item pair (e_h, e_j) if $(e_h, e_j) \in T_i$. Given the number $\beta > 0$ of tests, the maximum test covering is to find $\mathcal{T} \subseteq T$ that maximized the number of item pairs can be covered, subject to the constraints that $|\mathcal{T}| \leq \beta$.

As previously, the tests refer to the sensors and the objective of SP-EI2 is to maximize the number of distinguishable pairs with at most β sensors, i.e., $\arg\max_{\mathcal{S}} \left(\mathcal{F}_I(\mathcal{S}) \mid |\mathcal{S}| \leq \beta\right)$. The SP-EI1 and SP-EI2, abstracted as the MTC and MCT, respectively, are then cast as MILP problems, as described in the following subsections. In Section 5, the solution and the performance of the solution, when considering different event representations, i.e., boolean, one and two features, are demonstrated.

4.4.1. Robust sensor placement with unlimited number of sensors

We formulate the SP-EI1 as a MILP problem, where the objective is to minimize the number of sensors under the constraints that all pair-wise events can be distinguished. Firstly, we introduce a binary decision variable x_j for each plausible sensor s_j to indicate whether it is selected in the subset, i.e., $x_j = 1$ if $s_j \in \mathcal{S}$ and $x_j = 0$ otherwise. Subsequently, the optimization problem is formulated in Eq. (5). The objective of the optimization problem is to minimize the total cost of the selected sensors $\sum c_j x_j$, where c_j represents the cost of selecting sensor s_j. For each set of pair-wise events, the set of linear constraints guarantees that any distinguishable event pair (e_k, e_l) can be distinguished by at least α selected sensors. Typically, α is set to 1, indicating that the pair-wise events can be distinguished if they have different signatures in at least one selected sensor; however, for the purpose of improving robustness (e.g., under data uncertainties, sensor failures), α can be set to a greater value. The coefficients in the linear constraints are determined using Eq. (3), which is controlled by the identification tolerance δ.

$$\begin{aligned} \min_{x_j} \quad & \sum_{j=1}^{N_S} c_j x_j \\ \text{s.t.} \quad & \sum_{j=1,(e_k,e_l)\in C_j^I}^{N_S} x_j \geq \alpha \qquad \forall k, l = 1, 2, \ldots, N_E, k \neq l \\ & x_j \in \{0, 1\} \qquad \forall j = 1, 2, \ldots, N_S \end{aligned} \tag{5}$$

The model in Eq. (5) can be solved using any modern MILP solver, e.g., [70–72].

4.4.2. Robust sensor placement with limited number of sensors

In this section, we examine the more practical problem, in which given a limited budget the objective is to maximize the number of identified events. We first define a binary decision variable y_{kl} for each set of pair-wise events, where $y_{kl} = 1$ indicates that the two events, e_k and e_l, are distinguishable based on the signature matrix and a given tolerance level δ, and $y_{kl} = 0$ otherwise. The objective is then to maximize the number of pair-wise events that can be identified by the subset of sensors, subject to two types of constraints. The first set of constraint indicates whether pair-wise events (e_k, e_l) are distinguishable (i.e., have different event features), and the second constraint limits the available budget to be at most I_{cost}. Combining the objective and constraints, the maximum coverage test problem can be formulated as:

$$\begin{aligned} \max_{x_j, y_{kl}} \quad & \sum_{e_k, e_l \in E} y_{kl} \\ \text{s.t.} \quad & \sum_{j=1,(e_k,e_l)\in C_j^I}^{N_S} x_j \geq y_{kl} \qquad \forall k, l = 1, 2, \ldots, N_E, k \neq l \\ & \sum_{j=1}^{N_S} c_j x_j \leq I_{cost} \\ & 0 \leq y_{kl} \leq 1 \qquad \forall k, l = 1, 2, \ldots, N_E, k \neq l \\ & x_j \in \{0, 1\} \qquad \forall j = 1, 2, \ldots, N_S \end{aligned} \tag{6}$$

4.5. Performance evaluation metrics

We define four metrics to evaluate the performance of the sensor placement: detection, distinction, identification, and homogeneity scores. These four metrics provide evaluation from different and complementary perspectives, and are defined as follows:

Detection score represents the likelihood of detecting the EoIs, which is evaluated as the number of events that can be detected by the subset of sensors $\mathcal{S}$ normalized by the total number of events, and is defined as:

$$I_{De}(\mathcal{S}) = \frac{\mathcal{F}_D(\mathcal{S})}{N_E} \tag{7}$$

The detection score ranges between 0 and 1, where a higher detection score indicates that more EoIs are detected, and a lower detection score indicates otherwise.

Distinction score measures the number of pair-wise events that can be distinguished from each other, as defined in Eq. (3), normalized by the total number of detectable event pairs.

$$I_{Di}(\mathcal{S}) = \frac{\mathcal{F}_I(\mathcal{S})}{\binom{|E_D|}{2}} \tag{8}$$

The distinction score ranges between 0 and 1, where higher distinction score indicates that more pair-wise events can be distinguished by the set of sensors $\mathcal{S}$.

Identification score is the number of identifiable events normalized by the number of detectable events:

$$I_I(\mathcal{S}) = \frac{|E_I|}{|E_D|} \tag{9}$$

Intuitively, identification score quantifies the number of events that can be detected and result in a unique pressure signal, hence can be uniquely identified. The identification score is a conservative metric that represents only perfect event identification. As with detection and distinction scores, the identification scores ranges between 0 and 1, where an identification scores of 1 indicates that every detectable event can be uniquely identified.

Homogeneity score represents the size of largest group of events that cannot be distinguished from each other based on the sensor inputs normalized by the number of detectable events. Let $G_i = \{e_j, i \neq j \mid (e_i, e_j) \notin C^I_k, \ \forall s_k \in S\}$ be a subset of events for which the pair-wise events (e_i, e_j) that cannot be distinguished by any sensor s_k in

S, i.e., the output of all chosen sensors for events $e_j \in G_i$ is identical to that for event e_i. G_i is termed as the *homogeneity group* for event i. An event with larger homogeneity group indicates that more events exhibit similar and indistinguishable impact on the WDS; thus, the identification of the events is not possible by observing the extracted features. We then define the size of the largest homogeneity group normalized by the total number of events as the homogeneity score (I_H):

$$I_H(S) = \frac{max_i \mid G_i \mid}{|E_D|} \tag{10}$$

The homogeneity score ranges between 0 and 1, but unlike the previous scores, $I_H = 0$, implies a perfect one-to-one identification, such that all pairs of detectable events can be distinguished from each other.

Continuing the illustrative example with the change duration signature matrix shown in Fig. 1(e), for the two-sensor design $S = \{s_1, s_2\}$ and identification tolerance $\delta = 0.3$, all three events can be detected. Moreover, e_3 can be distinguished from e_1 and e_2, but e_1 and e_2 cannot be distinguished from each other, because the difference between the extracted features in the signature matrix are below the tolerance value. For example, the change duration for e_1 and e_2 as recorded by s_1 is identical (and equals to 2), and recorded by s_2 is 1.90 and 1.70, respectively, the difference between which is less than $\delta = 0.3$. The homogeneity groups for each event are $G_1 = \{e_2\}$, $G_2 = \{e_1\}$, and $G_3 = \emptyset$. The corresponding group sizes are $|G_1| = 1$, $|G_2| = 1$, and $|G_3| = 0$, indicating that e_3 is an identifiable event, and the event pair (e_1, e_2) is not distinguishable. Thus, the detection score is $I_{De} = 1$, the distinction score is $I_{Di} = 2/3$, the identification score is $I_I = 1/3$, and the homogeneity score is $I_H = 1/3$.

4.6. Multiple criteria decision analysis

The goal of this step to choose the optimal sensor placement out from various options obtained by using different event representation scheme (SM^B or SM) and selecting different number of sensors (in SP-EI2). It is important to systematically consider multiple criteria, including the information gain, as quantified by the four performance metrics, and the cost for installing and maintaining the sensors. To achieve this goal, we implemented the PROMETHEE method for multi-criteria decision making and selecting the best compromise between different alternatives [50]. Intuitively, the PROMETHEE method involves: (1) performing pair-wise comparisons between the alternatives, (2) represent decision-makers preferences by weighting the different criteria, and (3) aggregating the performance based on outranking and outranked performance. This approach allows the decision maker to include their preferences for the different performance metrics and select a single best design option from the different alternatives. In our approach, we explored the best selection of the number of sensor by weighing the total information gain (as measured by the four performance metrics) and cost of sensors, thus enabling a flexible decision making framework.

A single ranking in PROMETHEE can be obtained following the procedure presented below. First, a partial preference function $P_j(S_1, S_2)$ for each of the five criteria, i.e., $I_j \in [I_{De}, I_{Di}, I_I, I_H, I_{cost}]$, is defined as:

$$P_j(S_1, S_2) = \begin{cases} 1 & \text{if} \quad I_j(S_1) - I_j(S_2) \geq r_j \\ \frac{I_j(S_1) - I_j(S_2) - q_j}{r_j - q_j} & \text{if} \quad q_j < I_j(S_1) - I_j(S_2) < r_j \\ 0 & \text{if} \quad I_j(S_1) - I_j(S_2) \leq q_j \end{cases} \tag{11}$$

where S_1, S_2 represents two different sensor placement designs, r_j is the preference threshold representing the smallest performance deviation, which is considered sufficient to generate a full preference of one design over the other, and q_j is the indifference threshold representing the largest performance deviation that is considered negligible based on criteria I_j. Hence, $P_j(S_1, S_2)$ represents the degree of preference between sensor design S_1 and S_2 on criteria I_j, where higher value of $P_j(S_1, S_2)$ indicates S_1 is preferred over S_2 based on criteria I_j.

Second, a weight (w_j) needs to be specified for each criteria j to reflect how the decision maker values each criteria. In this study, we specify $w_j = \omega/4$ for all four performance metrics, which sum up to ω as they collectively measure the information gain, and $w_j = 1 - \omega$ for cost, such that $\sum w_j = 1$. Here, ω is a parameter that represents the value of information, where if information gain and cost are valued equally, then $\omega = 0.5$; if information gain is valued more than cost, then $\omega > 0.5$; otherwise, $\omega < 0.5$.

Third, the comprehensive preference function $\pi(S_1, S_2)$ that represents the degree of preference of design S_1 over design S_1 taking into account all criteria simultaneously is calculated as:

$$\pi(S_1, S_2) = \sum_{j=1}^{n} w_j P_j(S_1, S_2) \tag{12}$$

Consequently, for each design $S_1 \in A$, where A is the set of all design alternatives, the positive $\phi^+(S_1)$, negative $\phi^-(S_1)$, and net $\phi(S_1)$ outranking flows can be calculated:

$$\phi^+(S_1) = \frac{1}{|A| - 1} \sum_{S_2 \in A, S_1 \neq S_2} \pi(S_1, S_2) \tag{13}$$

$$\phi^-(S_1) = \frac{1}{|A| - 1} \sum_{S_2 \in A, S_1 \neq S_2} \pi(S_2, S_1) \tag{14}$$

$$\phi(S_1) = \phi^+(S_1) - \phi^-(S_1) \tag{15}$$

where $\phi^+(S_1)$ quantifies how much design S_1 is outranking other designs (i.e., the power of S_1), $\phi^-(S_1)$ measures how much design S_1 is outranked by others (i.e., the weakness of S_1), and $\phi(S_1)$ is the balance between $\phi^+(S_1)$ and $\phi^1(S_1)$, expressing the overall strength of design S_1 with respect to the set of all alternative designs. Then the design alternatives are ranked by the net outranking flows (ϕ), and the design with highest ϕ should be selected.

5. Application and results

In this section, we test the proposed sensor placement approach to identify the location of bursts in a medium-size WDS. The example network comprises 126 junctions, 1 reservoir, 2 tanks, 168 pipes, 2 pumps, and 8 valves, and its topology is depicted in Figure 1(a) in the Supporting Information (SI). The complete information and hydraulic model can be found in Ostfeld et al. [73]. In this application, each junction was considered as candidate location for sensor placement and the cost of each sensors was assumed to be identical regardless of the location. Also, $N_E = 106$ representative burst events of same size were simulated at each location in the network that is not directly connected to a valve or pump. Hence, in the context of this case study, event identification refers to identifying the location of the burst.

5.1. Event simulation

The transient simulations were carried out to model the impacts of different burst events at all potential sensing locations in the WDS. The bursts start at the beginning of the simulation and take 1s to fully develop to the final state, with lumped burst coefficient set to 0.002 $m^3/s/(mH_2O)^{1/2}$, resulting in a final burst discharge of approximately 20 l/s. Fig. 3 illustrates the pressure responses at Junction-2 and Junction-18 to the burst occurring at Junction-18. The pressure head at the burst node (Junction-18) decreases as the burst develops from 0 to 1 s and then, recovers gradually to a pressure lower than the original pressure. Subsequently, the pressure wave induced by the burst arrives to Junction-2 at approximately 0.5 s, generating a pressure drop of approximately 10 m, which is greater than the pressure drop observed at the burst node due to the complex topology of the network, in which pressure waves are reflected, transmitted, and interacted. Junction-2 also experiences additional pressure fluctuations due to the reflections of pressure waves in the network. After approximately 40 s, the pressure at both stations stabilizes and reaches the new steady state.

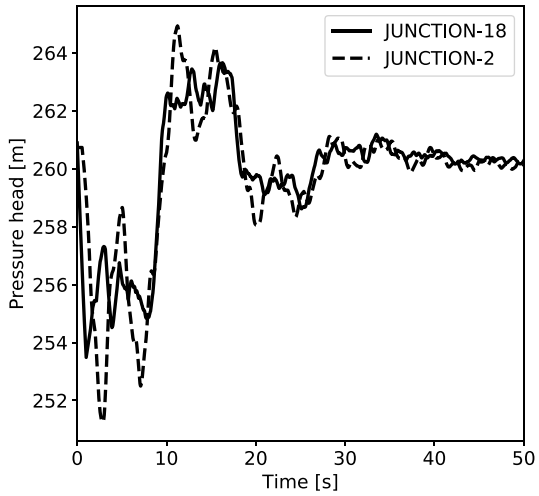

Fig. 3. Pressure responses at Junction-2 and Junction-18 to the burst occurring at Junction-18.

5.1.1. Comparing transient, steady-state, and distance-based approximation models

Fig. 4 shows the results comparing system response to a burst occurring at Junction-18 using transient, steady-state, and distance-based approximation models. To illustrate the limitation of using steady-state analysis for modeling pipe bursts, we performed simulations using EPANET (following the modeling approach suggested in [14,36]) and TSNet (as proposed in this work). Fig. 4(a) presents the comparison of simulation results in response to a burst occurring at Junction-18 using the transient model — TSNet [58], and the steady-state model — EPANET [56]. The x-axis represents the steady-state pressure difference at the different nodes in the network induced by a burst occurring at Junction-18 modeled in EPANET using an emitter with the coefficient set to 0.002 $m^3/s/(mH_2O)^{1/2}$. The y-axis shows the amplitude of pressure changes at the different nodes in the network induced by the same burst event modeled in TSNet using the same emitter coefficient, as discussed in Sections Section 4.1. The black circles show the magnitude of the pressure changes at different junctions in the network that can be detected by CUSUM with 5 m threshold (as described in Section 5.2), while the red crosses represent the locations in the network at which the burst cannot be detected. Noticeably, the steady state nodal pressure differences are one magnitude smaller than the transient amplitude. Additionally, the impact of the burst is more distinguishable based on the results of the transient model, with pressure amplitudes changing between 5 and 12 m, while the majority of junctions experience a pressure drop of approximately 0.4 m, based on the results of the steady state model. The difference in pressure changes in the transient and steady state models can be explained by observing the pressure signals in Fig. 3. Immediately following the occurrence of the burst event the pressure drops significantly, which then fluctuates until reaching a new steady state. The differences in pressure head between the initial and the new steady states, which are captured by steady state simulations, are much smaller than amplitude of the pressure transient, which is captured by the transient dynamics.

Furthermore, Fig. 4(b) illustrates the amplitude of the first transient wave simulated using TSNet as a function of the distance on the fastest path from the origin of the burst to each node. Although after the distance from the burst origin exceeds approximately 1500 m, most of the junctions cannot sense the impact of the burst (as suggested by Sela Perelman et al. [40]), it can be noticed that the change in transient pressure does not necessarily decreases as the distance from the burst origin increases. This limitation of the distance-based models can also be observed from the transient pressure signals shown in Fig. 3. For example, the amplitude of the first transient wave at the burst origin (green marker in Fig. 4(b) and solid line in Fig. 3) is smaller than that at Junction-2 (blue marker in Fig. 4(b) and dashed line in Fig. 3), which is approximately 629 m away from the burst origin.

5.2. Feature extraction

The magnitude and duration of the pressure changes are key features of the pressure transients that can help distinguish between the different burst events. Various techniques have been developed to detect changes in time series signals, such as CUSUM [66], discrete wavelet transformation (DWT) [74], and singular spectrum transformation (SST) [75]. In this paper, a modified CUSUM algorithm, which was previously applied for pressure transient detection in a real WDS, is adopted as the change detection technique due to its efficiency, interpretability and generalizability [31]. The CUSUM algorithm, originally proposed by Page [66] as two repeated uses of sequential probability tests, tracks the characteristics of the changes, i.e., rate and magnitude, and compares these characteristics with control limits. The modified CUSUM detects the start ts and end time te of all the changes occurred during the period of interest and the characteristic features of the changes, such as the amplitude dp and duration dt of the first wave front, can be extracted from the signal to characterize additional features of EoIs as $dt = te - ts$, $dp = p(s_j, te \mid e_i) - p(s_j, ts \mid e_i)$.

The modified CUSUM algorithm is applied to all pressure time-series to detect the pressure changes originated from the burst events. The following results are demonstrated for setting the threshold in the CUSUM algorithm to 5 m, i.e, only pressure changes of amplitude greater than 5 m can be detected, resulting in 90 detectable events out of the 106 possible events. The characteristics of the detected changes, i.e., amplitude dp and duration dt, are then recorded to formulate the signature matrices.

5.3. Signature matrix

Based on the transient simulations and change detection algorithm, four signature matrices are constructed: (1) boolean, (2) change duration, (3) change amplitude, and (4) joint change duration and amplitude. As an example, the signature matrix with continuous change amplitude (dp) of the first change as the single feature is shown in Figure 2 in the SI. In this matrix, pixel (i, j) denotes the amplitude of the first detectable wave front in the pressure response at junction j to a burst event e_i. A positive change represents pressure rise (red), while a negative change represent pressure drop (green). It can be noticed that the dark green elements are mostly located on or near main diagonal, revealing that the amplitude of the pressure change at the burst node is generally greater than that at other nodes. Nevertheless, this is not always the case, as illustrated previously in Fig. 3. Moreover, according to energy conservation, a pipe burst typically results in a pressure drop; however, positive pressure changes are also observed in Figure 2 in the SI. It is plausible that the magnitude of the first front of the wave is smaller than the threshold, and hence, undetectable; however, the following cycle of the wave exhibits a greater pressure rise than the threshold and, thus, be detected by the algorithm and counted as the first detected pressure change.

5.4. Sensor placement with unlimited number of sensors

In this section, we solve the sensor placement problem with unlimited number of sensors, i.e., the SP-EP1 problem, to find the smallest subset of sensors that can identify all identifiable events. Our objective is to explore the sensitivity of the performance of the sensor placement designs to: (1) increasing level of information, and (2) different levels of robustness to data and model uncertainty. To address the first scenario, four different levels of information (i.e., extracted features) are tested. We expect that with increasing level of information the performance of the sensor placement strategies will improve, however at the expense

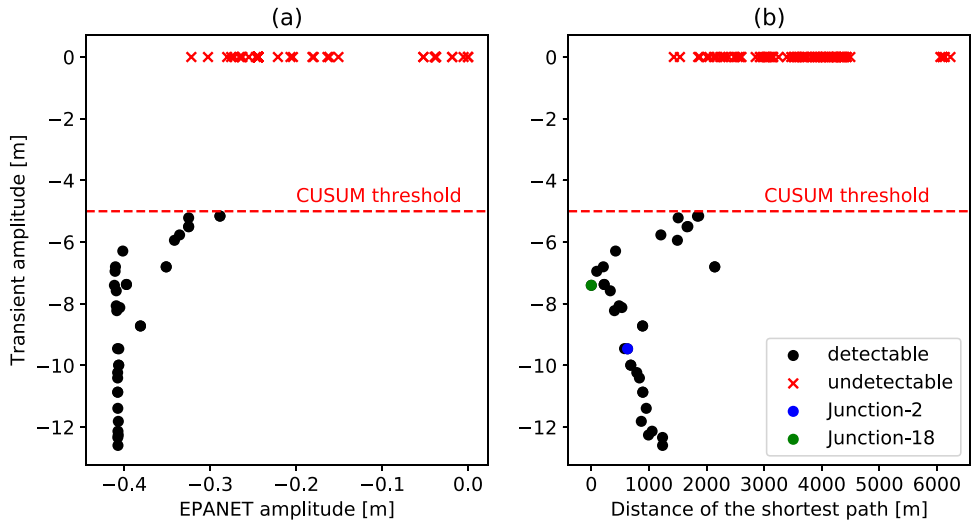

Fig. 4. Results comparison between different simulation models with the burst occurring at Junction-18: (a) the transient model versus the steady state model (crosses represent locations where the burst is undetectable, while dots represent locations where the burst is detectable), and (b) the transient model versus the distance-based model, where the green dot represents Junction-18 and blue dot represents Junction-2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of needing more complex and reliable modeling and change detection techniques. Another interesting aspect to explore is which feature, i.e., amplitude or duration, contributes the most unique information. To address the second question, we vary the level of robustness by resolving the sensor placement problem for different values of δ. In all the results reported next, the number of sensors required to distinguish two events is set to be one, i.e., $\alpha = 1$. IBM-CPLEX 12.9 commercial MILP solver is used to solve the optimization problems [71], each of which can be solved within 2–3 s on a 2.9 GHz Dual-Core Intel Core i5 processor.

5.4.1. Testing the sensitivity to incorporating more information

To begin with, four optimization problems, as formulated in Eq. (5), are solved with the different signature matrices mentioned above. The identification tolerance, as defined in Eq. (3), for change duration and amplitude set to be 1 s and 3.5 m, respectively, i.e., $\delta_{dp} = 3.5$ m and $\delta_{dt} = 1$ s, which is the expected accuracy from the transient simulations and CUSUM change detection algorithm. The solutions of the four optimization problems with different inputs were evaluated based on the three performance metrics described in Section 4.5: distinction, identification, and homogeneity scores. In the MTC problem, all detectable events have to be detected by the selected sensors, as enforced by the constraints, hence, all solutions detect all the events.

Fig. 5 shows the performance metrics for each case, where darker color represent better results. The first column of Fig. 5 reports the performance of the sensor placement based on only boolean information. Out of the 106 burst events that were generated, 90 events are detectable, and 45 unique signatures are observed. The solution of the optimization problem suggests that 29 sensors are needed to capture all the possible unique signatures, which enable to distinguish 3847 out of the $\binom{90}{2} = 4005$ pair-wise detectable events, i.e., the distinction score equals 0.96, and uniquely identify 28 burst events, i.e., the identification score equals $28/90 \approx 0.31$. For the remaining events, the largest homogeneity group contains 12 burst events sharing the same signature, i.e., with homogeneity score of $12/90 \approx 0.13$. When additional information, either event amplitude or duration, is incorporated in the signature matrix, the performance improves. Using the event amplitude signature matrix, the number of required sensors decreases from 29 to 27, and the distinction and identification scores increase from 0.96 to 0.98, and 0.31 to 0.41, respectively. Even better performance is achieved when the event duration is utilized to enrich the information in signature matrix: the distinction and identification

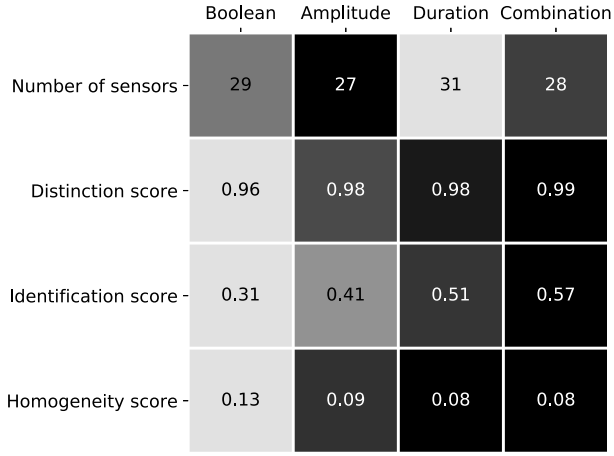

Fig. 5. Performance evaluation using different extracted information.

scores further increase to 0.98 and 0.51 respectively, although the number of sensors required increases to 31. This observation indicates that in this WDS, duration signature matrix with $\delta_{dt} = 1s$ can provide more information for event identification compared to the amplitude signature matrix with $\delta_{dp} = 3.5$ m, and more sensors are required to reveal the additional information.

Furthermore, using signature matrix with combined features, i.e., coupling duration and amplitude, best performance across all metrics is achieved: with 28 sensors, 3951 out of the 4005 pair-wise detectable events can be distinguished from each other, i.e., the distinction score increases to 0.99, and 51 out of the 90 events can be uniquely identified, i.e., the identification score increases to 0.57. Additionally, the size of the largest homogeneity group decreases to 7, i.e., homogeneity score reduces to 0.08. Ultimately, the results in Fig. 5 demonstrate that the inclusion of more information can improve the performance of the sensor placement and provide more accurate event identification.

In the context of WDSs, the homogeneity score has practical implications for identifying the location of the bursts and informing more localized inspection efforts. Figure 1(b) in the SI shows an example of different homogeneity groups using signature matrices containing

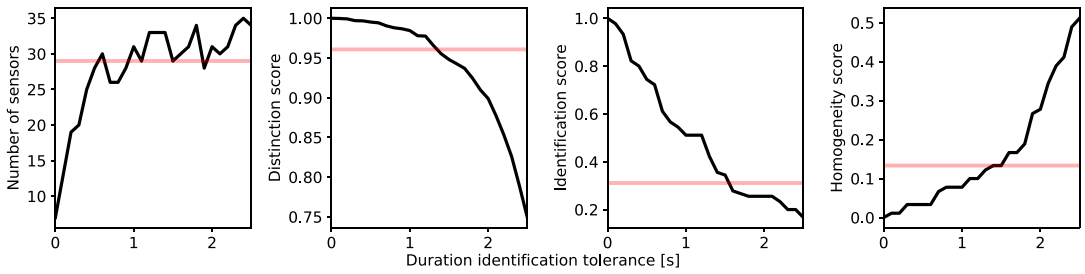

Fig. 6. Sensitivity of sensor placement performance to duration identification tolerance (δ_{dt}).

different levels of information for burst event generated at Junction-79 (black circle node). With only boolean information, the size of the homogeneity group is 12, indicating that 12 burst events, as depicted by all the square markers in Figure 1(b) in the SI, result in the same signature as the burst at Junction-79 and thus cannot be distinguished within the group. Adding information about change amplitudes, the homogeneity group reduced to contain 9 burst events, as shown using square markers with dotted and striped filling. Using signature matrix with combined information about change duration and amplitude, the size of the homogeneity group further shrinks to 7 events, shown as the markers with striped filling. The decreasing homogeneity group reveals that as more information is included in the optimization problem, the number of bursts that cannot be distinguished decreases.

5.4.2. Testing the sensitivity to tolerance threshold

Next, the sensitivity of performance of the sensor placement to different values of duration and amplitude identification tolerance is tested, by resolving the MTC problem and evaluating the different performance scores. Fig. 6 shows the number of sensors and the performance scores, including the distinction, identification, homogeneity scores, as functions of duration identification tolerance (δ_{dt}). As a benchmark, the red horizontal lines show the performance with boolean signature matrix. A solution that includes the duration of the event outperforms a solution with only boolean information, if the performance for the former (black lines) is above the performance of the latter (red lines) for the distinction and identification scores, and otherwise for the homogeneity score. It can be observed that all the performance metrics degrade as δ_{dt} increases: more sensors are required, smaller number of pair-wise bursts can be distinguished from each other, less bursts can be uniquely defined, and number of bursts sharing the same signature increases. Additionally, when δ_{dt} exceeds approximately 1.4 s, the benefits of incorporating the change duration information are overshadowed by the uncertainties represented by the identification tolerance. This observation indicates that sufficient temporal precision is required to make the inclusion of change duration information meaningful. Furthermore, we observe that the number of required sensors does not increase considerably after δ_{dt} exceeds approximately 0.7 s. This is because δ_{dt} essentially affects the total amount of information about the EoIs that can be extracted; hence, the increasing level of uncertainty as represented by increasing δ_{dt} inherently reduces the amount of information, which cannot be regained even with increasing amount of sensors.

Similarly, the sensitivity of the solutions to the amplitude tolerance (δ_{dp}) is tested, as shown in Figure 3 in the SI. Similar results are observed when using the change amplitude as the characteristic feature of the burst events. When δ_{dp} exceeds 5 m, the performance of using amplitude signature matrix is again worse than that of using only boolean signature matrix, which emphasizes the importance of ensuring adequate accuracy of the information incorporated in the signature matrices. For this WDS, we demonstrated that it is essential to guarantee that the uncertainties associated with duration and amplitude are less than 1.4 s and 5 m, respectively. Otherwise, the incorporation of this information can potentially degrade the performance of the sensor placement design.

5.5. Sensor placement with limited number of sensors

In this section, we solve the more realistic MCT problem as formulated in Eq. (6), in which the number of sensors is limited. We test the sensitivity of the solution to an increasing number of sensors and to an increasing level of information. The identification tolerance for duration and amplitude are set to be $\delta_{dt} = 1$ s and $\delta_{dp} = 3.5$ m for the following results. Fig. 7 compares the performance metrics using different signature matrices as functions of the number of sensors. As expected, all performance scores improve as the number of available sensors increases and more information is incorporated in the optimization problem. It should be emphasized that only distinction score, as shown in Fig. 7(b), is the objective function in the optimization problem, and the rest of the performance metrics are evaluated after solving the optimization problem. Hence, with increasing number of sensors, only improved distinction score is guaranteed, while for the rest of the metrics the general trend is increasing but not guaranteed at every solution point. Additionally, we observe from Fig. 7(a) that similar detection scores are obtained with all four different signature matrices, because only boolean information is needed for detection. Fig. 7(b) indicates that boolean signature matrix yields lower distinction score compared with the other three matrices, which behave similarly compared to each other. Moreover, detection and distinction scores exhibit a diminishing return trend with the number of available sensors, i.e., the scores increase steeply when the number sensors increases from 2 to 9, but result in only marginal improvement as more sensors are added. Additionally, only 15 sensors are sufficient to achieve the near optimal performance in terms of detection and distinction. However, the identification score increases steadily as the number of sensors increases from 2 to 25, as shown in Fig. 7(c). In fact, a closer look at Fig. 7(c) reveals that the benefit of adding more sensors only become apparent after more than 5 sensors are already included, and 15 sensors are apparently not enough for the purpose of uniquely identifying the bursts. Furthermore, with the same number of sensors, better identification performance can be achieved when more information is contained in the signature matrix: the combination signature matrix contains the most information and thus yields the highest identification score, followed by duration, amplitude, and boolean signature matrices. The difference is magnified as the number of sensors increases. Finally, the homogeneity scores are presented in Fig. 7(d), where again the combination signature matrix gives the best results, i.e., lowest homogeneity scores, and the overall trend of diminishing return as with detection and distinction scores is observed.

We applied the PROMETHEE method to determine the number of sensors that should be deployed based on the information gain from the four performance metrics (I_{De}, I_{Di}, I_I, I_H) and cost of the sensors (I_{cost}). We show results for sensor design based on the combination signature matrix. The parameters used in PROMETHEE are summarized in Table 1 in SI.

Fig. 8(a) shows the net outranking flows (ϕ) considering only information gain (i.e., $\omega = 1$, green squares), only cost (i.e., $\omega = 0$, red triangles), and combined performance with equal weight assign to information gain and cost (i.e., $\omega = 0.5$, black circles). Observing

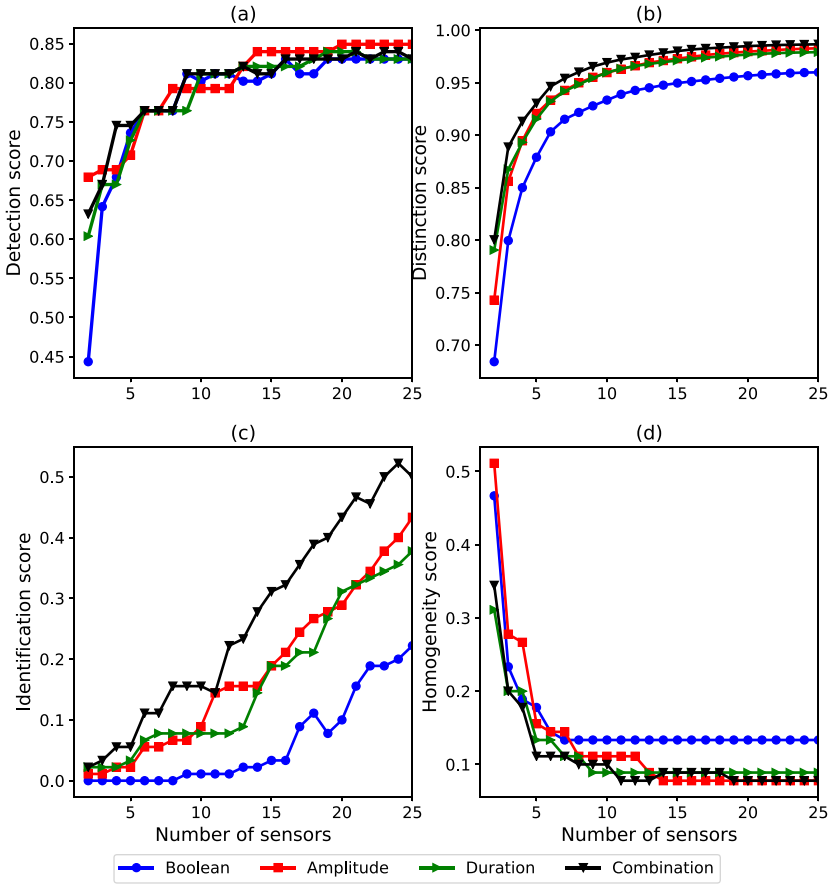

Fig. 7. Performance evaluation using different extracted information and increasing number of sensors.

the outranking flows of information gain, the general increasing trend indicates that more sensors are preferred if a decision maker is only concerned about information gain. However, it should be noted that there is only a marginal increase in ϕ when the number of sensors exceeds 20. On the other hand, ϕ for sensor costs, as represented by the red line, decreases monotonically as the number of sensors increases, indicating that, as expected, a design with fewer sensors is preferred if only cost is considered. Furthermore, when information gain and sensor cost are valued equally maximum ϕ is achieved at 9 sensors, indicating that the optimal sensor placement design is with 9 sensors. We further investigated the number of optimal sensors as a function of the weight that a decision-maker assigns to information gain (ω) and cost ($1-\omega$), as shown in Fig. 8(b). It can be observed that when the cost is valued more than information gain (i.e., $\omega < 0.5$), designs with fewer sensors are preferred, thus reflecting budget-constrained water utilities preferences. On the other hand, when information gain is valued more than cost (i.e., $\omega > 0.5$), a water utility will benefit from installing more sensors to obtain a better burst identification performance. This approach enables incorporating decision makers preferences for balancing information gain and cost and can facilitate in selecting the number of sensors.

6. Conclusions and future work

In this work, the sensor placement problem for event identification under model and measurement uncertainties was investigated. The main contributions of this paper include: (1) we rely on transient hydraulics to model pipe burst events and demonstrate that the surrogates may not provide a good approximation for transient system response, (2) we propose a reduced representation of the continuous system response combined with tolerance analysis that is more robust to data and modeling uncertainties, and (3) we demonstrate the sensitivity of the performance to different levels and accuracy of extracted information. Ultimately, the results show that using transient modeling and incorporating more information in the signature matrix can improve the performance of burst identification. However, the analysis also suggests that if information is uncertain, including more information in signal representation can be suboptimal to parsimony signal representation. Thus, sufficient accuracy of the extracted information is required to accrue the benefits of incorporating the additional information for event identification.

To increase the usability of the proposed approach, further research is required to address some of the limitations that were not included in the scope of the current work. The proposed method requires a hydraulic model that accurately represents the WDS; however, building such hydraulic model typically require more exhaust than readily available information. The exact network topology, hydraulic connectivity, and user demands in a WDS are usually difficult to obtain, thus introducing additional uncertainties in the sensor placement process. Hence, this work emphasizes the need for hydraulic models to improve decision-making for system monitoring and burst detection. Additionally, constrained by the network size limitations of TSNet, the proposed method was tested on a mid-size WDS. Future work should investigate the performance of the proposed method on large-scale WDSs using more advance modeling frameworks for speeding up transient simulations using parallel computing [76]. Moreover, the sensitivity of the proposed approach should be tested using more advanced techniques for extracting features from transient pressure signals [77]. Finally, this paper investigated sensor placement for detecting pipe bursts before these develop into persistent leaks. The approach is based on analyzing the transient pressure signals that are induced by bursts, while leaks are preexisting condition and do not induce transients. While the data

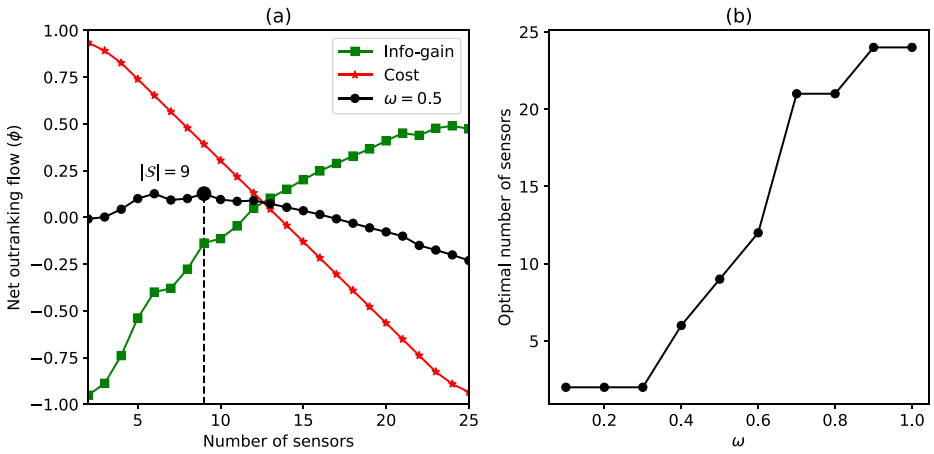

Fig. 8. PROMETHEE results to determine the optimal number of sensors: (a) Net outranking flow (ϕ) for information gain (green), sensor costs (red), and combined performance with equal weights (black) of designs with different number of sensors; and (b) number of optimal sensors as a function of the weight that a decision-maker assigns to information gain (ω) and cost ($1-\omega$).

collected from the pressure sensors that we consider in this paper can be used for leak detection, it is not the objective of the optimization problem that we consider, since system dynamics for steady-state and transient hydraulics are different. Despite these limitations, the current study demonstrated the potential of incorporating transient modeling for the sensor placement problem for pipe burst detection.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the University of Texas at Austin, United States of America Startup Grant and by the National Science Foundation, United States of America under Grant 1943428. All data generated in this work was simulated using Example 3 in TSNet https://tsnet.readthedocs.io/en/latest/examples.html.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.aei.2021.101484.

References

[1] A. Sedki, D. Ouazar, Hybrid particle swarm optimization and differential evolution for optimal design of water distribution systems, Adv. Eng. Inform. 26 (3) (2012) 582–591.
[2] T. Chan, C. Chin, X. Zhong, Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection, IEEE Access 6 (2018) 78846–78867.
[3] J. Makar, Y. Kleiner, Maintaining water pipeline integrity, in: AWWA Infrastructure Conference and Exhibition, Vol. 12, 2000, pp. 1–13.
[4] K.J. Brothers, Water leakage and sustainable supply—truth or consequences? J. Amer. Water Works Assoc. 93 (4) (2001) 150–152.
[5] W. Guo, L. Soibelman, J. Garrett Jr., Visual pattern recognition supporting defect reporting and condition assessment of wastewater collection systems, J. Comput. Civ. Eng. 23 (3) (2009) 160–169.
[6] M.R. Halfawy, J. Hengmeechai, Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine, Autom. Constr. 38 (2014) 1–13.
[7] M. Ahadi, M.S. Bakhtiar, Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to acoustic emission signals, Appl. Acoust. 71 (7) (2010) 634–639.
[8] S. Li, Y. Song, G. Zhou, Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition, Measurement 115 (2018) 39–44.
[9] A. Martini, A. Rivola, M. Troncossi, Autocorrelation analysis of vibro-acoustic signals measured in a test field for water leak detection, Appl. Sci. 8 (12) (2018) 2450.
[10] R. Müller, S. Illium, F. Ritz, T. Schröder, C. Platschek, J. Ochs, C. Linnhoff-Popoien, Acoustic leak detection in water networks, 2020, arXiv preprint arXiv:2012.06280.
[11] L. Crocco, F. Soldovieri, T. Millington, N.J. Cassidy, Bistatic tomographic GPR imaging for incipient pipeline leakage evaluation, Prog. Electromagn. Res. 101 (2010) 307–321.
[12] S. Demirci, E. Yigit, I.H. Eskidemir, C. Ozdemir, Ground penetrating radar imaging of water leaks from buried pipes based on back-projection method, Ndt E Int. 47 (2012) 35–42.
[13] P.M. Bach, J.K. Kodikara, Reliability of infrared thermography in detecting leaks in buried water reticulation pipes, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (9) (2017) 4210–4224.
[14] R. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, A. Peralta, Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks, Control Eng. Pract. 19 (10) (2011) 1157–1167.
[15] M.V. Casillas Ponce, L.E. Garza Castanon, V.P. Cayuela, Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities, J. Hydroinform. 16 (3) (2014) 649–670.
[16] G. Moser, S.G. Paal, I.F. Smith, Performance comparison of reduced models for leak detection in water distribution networks, Adv. Eng. Inform. 29 (3) (2015) 714–726.
[17] G. Moser, S.G. Paal, I.F. Smith, Leak detection of water supply networks using error-domain model falsification, J. Comput. Civ. Eng. 32 (2) (2017) 04017077.
[18] X. Wang, M.F. Lambert, A.R. Simpson, J.A. Liggett, J.P. Vítkovský, Leak detection in pipelines using the damping of fluid transients, J. Hydraul. Eng. 128 (7) (2002) 697–711.
[19] A.F. Colombo, P. Lee, B.W. Karney, A selective literature review of transient-based leak detection methods, J. Hydro-Environ. Res. 2 (4) (2009) 212–227.
[20] J. Bohorquez, A.R. Simpson, M.F. Lambert, B. Alexander, Merging fluid transient waves and artificial neural networks for burst detection and identification in pipelines, J. Water Resour. Plan. Manage. 147 (1) (2021) 04020097.
[21] D. Levinas, G. Perelman, A. Ostfeld, Water leak localization using high-resolution pressure sensors, Water 13 (5) (2021) 591.
[22] R. Ebina, K. Nakamura, S. Oyanagi, A real-time burst detection method, in: 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, IEEE, 2011, pp. 1040–1046.
[23] X. Du, M.F. Lambert, L. Chen, E. Jing Hu, W. Xi, Pipe burst detection, localization, and quantification using the transient pressure damping method, J. Hydraul. Eng. 146 (11) (2020) 04020077.
[24] R.A. Cody, S. Narasimhan, A field implementation of linear prediction for leak-monitoring in water distribution networks, Adv. Eng. Inform. 45 (2020) 101103.
[25] X. Du, W. Zeng, M.F. Lambert, L. Chen, E. Jing Hu, Approach for near-real-time pipe burst detection, localization, and quantification with low data transmission and sampling rates, J. Water Resour. Plan. Manage. 147 (7) (2021) 04021032.
[26] A. Lambert, T.G. Brown, M. Takizawa, D. Weimer, A review of performance indicators for real losses from water supply systems, J. Water Supply: Res. Technol.—AQUA 48 (6) (1999) 227–237.
[27] M.S. Ghidaoui, M. Zhao, D.A. McInnis, D.H. Axworthy, A review of water hammer theory and practice, Appl. Mech. Rev. 58 (1) (2005) 49–76.

[28] R. Puust, Z. Kapelan, D. Savic, T. Koppel, A review of methods for leakage management in pipe networks, Urban Water J. 7 (1) (2010) 25–45.

[29] X. Xu, B. Karney, An overview of transient fault detection techniques, in: Modeling and Monitoring of Pipelines and Networks, Springer, 2017, pp. 13–37.

[30] M. Allen, A. Prels, M. Lqbal, S. Srirangarajan, H.B. Llm, L. Glrod, A.J. Whittle, Real-time in-network distribution system monitoring to improve operational efficiency, J. Amer. Water Works Assoc. 103 (7) (2011) 63–75.

[31] L. Xing, L. Sela, Unsteady pressure patterns discovery from high-frequency sensing in water distribution systems, Water Res. 158 (2019) 291–300.

[32] S. Meniconi, H. Duan, P. Lee, B. Brunone, M. Ghidaoui, M. Ferrante, Experimental investigation of coupled frequency and time-domain transient test–based techniques for partial blockage detection in pipelines, J. Hydraul. Eng. 139 (10) (2013) 1033–1040.

[33] M. Hagos, D. Jung, K.E. Lansey, Optimal meter placement for pipe burst detection in water distribution systems, J. Hydroinform. 18 (4) (2016) 741–756.

[34] W. Cheng, Y. Chen, G. Xu, Optimizing sensor placement and quantity for pipe burst detection in a water distribution network, J. Water Resour. Plann. Manage. 146 (11) (2020) 04020088.

[35] M. Zhao, C. Zhang, H. Liu, G. Fu, Y. Wang, Optimal sensor placement for pipe burst detection in water distribution systems using cost–benefit analysis, J. Hydroinform. 22 (3) (2020) 606–618.

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

[37] J.-A. Goulet, S. Coutu, I.F. Smith, Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks, Adv. Eng. Inform. 27 (2) (2013) 261–269.

[38] A. Deshpande, S.E. Sarma, K. Youcef-Toumi, S. Mekid, Optimal coverage of an infrastructure network using sensors with distance-decaying sensing quality, Automatica 49 (11) (2013) 3351–3358.

[39] W. Abbas, L. Sela Perelman, S. Amin, X. Koutsoukos, An efficient approach to fault identification in urban water networks using multi-level sensing, in: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, in: BuildSys '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 147–156.

[40] L. Sela Perelman, W. Abbas, X. Koutsoukos, S. Amin, Sensor placement for fault location identification in water networks: A minimum test cover approach, Automatica 72 (2016) 166–176.

[41] Y. Douek-Pinkovich, I. Ben-Gal, T. Raviv, The generalized test collection problem, TOP (2020).

[42] T.Y. Berger-Wolf, W.E. Hart, J. Saia, Discrete sensor placement problems in distribution networks, Math. Comput. Modelling 42 (13) (2005) 1385–1396.

[43] V.R. Palleti, S. Narasimhan, R. Rengaswamy, Optimal sensor placement for contamination detection and identification in water distribution networks, in: Computer Aided Chemical Engineering, Vol. 33, Elsevier, 2014, pp. 1447–1452.

[44] F. Hooshmand, F. Amerehi, S. Mirhassani, Logic-based benders decomposition algorithm for contamination detection problem in water networks, Comput. Oper. Res. 115 (2020) 104840.

[45] M. Propato, F. Sarrazy, M. Tryby, Linear algebra and minimum relative entropy to investigate contamination events in drinking water systems, J. Water Resour. Plan. Manage. 136 (4) (2010) 483–492.

[46] V.R. Palleti, S. Narasimhan, R. Rengaswamy, R. Teja, S.M. Bhallamudi, Sensor network design for contaminant detection and identification in water distribution networks, Comput. Chem. Eng. 87 (2016) 246–256.

[47] W. Abbas, L. Sela Perelman, S. Amin, X. Koutsoukos, Resilient sensor placement for fault localization in water distribution networks, in: Proceedings of the 8th International Conference on Cyber-Physical Systems, 2017, pp. 165–174.

[48] S. Soloman, Sensors Handbook, McGraw-Hill, Inc., 2009.

[49] J.-Y. Dantan, N. Gayton, A.J. Qureshi, M. Lemaire, A. Etienne, Tolerance analysis approach based on the classification of uncertainty (aleatory/epistemic), Procedia CIRP 10 (2013) 287–293.

[50] M. Behzadian, R.B. Kazemzadeh, A. Albadvi, M. Aghdasi, PROMETHEE: A comprehensive literature review on methodologies and applications, European J. Oper. Res. 200 (1) (2010) 198–215.

[51] K.M. De Bontridder, B.V. Halldórsson, M.M. Halldórsson, C.A. Hurkens, J.K. Lenstra, R. Ravi, L. Stougie, Approximation algorithms for the test cover problem, Math. Program. 98 (1–3) (2003) 477–491.

[52] S.E. Christodoulou, A. Gagatsis, S. Xanthos, S. Kranioti, A. Agathokleous, M. Fragiadakis, Entropy-based sensor placement optimization for waterloss detection in water distribution networks, Water Resour. Manage. 27 (13) (2013) 4443–4468.

[53] R. Sarrate, J. Blesa, F. Nejjari, J. Quevedo, Sensor placement for leak detection and location in water distribution networks, Water Sci. Technol. Water Supply 14 (5) (2014) 795–803.

[54] E. Raei, M.E. Shafiee, M.R. Nikoo, E. Berglund, Placing an ensemble of pressure sensors for leak detection in water distribution networks under measurement uncertainty, J. Hydroinform. 21 (2) (2019) 223–239.

[55] M.S. Khorshidi, M.R. Nikoo, N. Taravatrooy, M. Sadegh, M. Al-Wardy, G.A. Al-Rawas, Pressure sensor placement in water distribution networks for leak detection using a hybrid information-entropy approach, Inform. Sci. 516 (2020) 56–71.

[56] L. Rossman, M.T. H. Woo, R.J. F. Shang, T. Haxton., EPANET 2.2 users manual, 2020.

[57] L. Sela, S. Amin, Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization, Adv. Eng. Inform. 36 (2018) 55–63.

[58] L. Xing, L. Sela, Transient simulations in water distribution networks: TSNet python package, Adv. Eng. Softw. 149 (2020) 102884.

[59] R. Debouk, S. Lafortune, D. Teneketzis, On an optimization problem in sensor selection, Discrete Event Dyn. Syst. 12 (4) (2002) 417–445.

[60] G. Chi, D. Wang, T. Le, M. Yu, M. Luo, Sensor placement for fault isolability using low complexity dynamic programming, IEEE Trans. Autom. Sci. Eng. 12 (3) (2015) 1080–1091.

[61] R. Sarrate, V. Puig, T. Escobet, A. Rosich, Optimal sensor placement for model-based fault detection and isolation, in: 2007 46th IEEE Conference on Decision and Control, IEEE, 2007, pp. 2584–2589.

[62] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, C. Faloutsos, Efficient sensor placement optimization for securing large water distribution networks, J. Water Resour. Plan. Manage. 134 (6) (2008) 516–526.

[63] E.B. Wylie, V.L. Streeter, L. Suo, Fluid Transients in Systems, first ed., Prentice Hall Englewood Cliffs, NJ, 1993.

[64] L. Xing, A. Rasekh, M.E. Shafiee, L. Sela, A. Preis, Streaming sensor data validation in networked infrastructure systems through synergic auto and cross similarity discovery and analysis, in: Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience, American Society of Civil Engineers Reston, VA, 2019, pp. 343–350.

[65] D. Lu, P. Mausel, E. Brondizio, E. Moran, Change detection techniques, Int. J. Remote Sens. 25 (12) (2004) 2365–2401.

[66] E.S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115.

[67] F. Moerchen, Algorithms for time series knowledge mining, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 668–673.

[68] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing SAX: a novel symbolic representation of time series, Data Min. Knowl. Discov. 15 (2) (2007) 107–144.

[69] M.G. Resende, Computing approximate solutions of the maximum covering problem with GRASP, J. Heuristics 4 (2) (1998) 161–177.

[70] A. Mosek, The MOSEK optimization software, 2010, p. 5, Online at http://www.mosek.com, 54.

[71] I.I. Cplex, V12. 1: User's manual for CPLEX, Int. Bus. Mach. Corpor. 46 (53) (2009) 157.

[72] I. Gurobi Optimization, Gurobi optimizer reference manual, 2018, URL http://www.gurobi.com.

[73] A. Ostfeld, J.G. Uber, E. Salomons, J.W. Berry, W.E. Hart, C.A. Phillips, J.-P. Watson, G. Dorini, P. Jonkergouw, Z. Kapelan, et al., The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms, J. Water Resour. Plan. Manage. 134 (6) (2008) 556–568.

[74] P. Chaovalit, A. Gangopadhyay, G. Karabatis, Z. Chen, Discrete wavelet transform-based time series analysis and mining, ACM Comput. Surv. 43 (2) (2011) 1–37.

[75] S. Aminikhanghahi, D.J. Cook, A survey of methods for time series change point detection, Knowl. Inf. Syst. 51 (2) (2017) 339–367, http://arxiv.org/abx/15334406.

[76] G. Riaño Briceño, L. Sela, B.R. Hodges, Distributed and vectorized method of characteristics for fast transient simulations in water distribution systems, Comput.-Aided Civ. Infrastruct. Eng. (2021) http://dx.doi.org/10.1111/mice.12709.

[77] T. Chan, C. Chin, Y. Li, E. Shafiee, L. Sela, A better estimation of wave arrival time in water distribution networks using wavelet knee (WANE), Adv. Eng. Inform. (ISSN: 1474-0346) 48 (2021) 101287.