Urban Water Journal

ISSN: 1573-062X (Print) 1744-9006 (Online) Journal homepage: https://www.tandfonline.com/loi/nurw20

A review of data-driven approaches for burst detection in water distribution systems

Yipeng Wu & Shuming Liu

URBAN WATER JOURNAL, 2017
VOL. 14 NO. 9, 972–983
https://doi.org/10.1080/1573062X.2017.1279191

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

A review of data-driven approaches for burst detection in water distribution systems

Yipeng Wu and Shuming Liu

School of Environment, Tsinghua University, Beijing, China

ABSTRACT
This study focuses on data-driven approaches for burst detection and classifies them into three categories: classification method, prediction-classification method and statistical method. The performance of these methods is discussed. By analysing uncertainty in burst detection, this paper revealed that non-stationary monitoring data and limitations present in these methods challenge the reliability of detection results. Data pre-processing and probabilistic solutions to deal with the uncertainty are summarised. From these findings and discussions, this paper concludes and recommends that: a) data-driven approaches are promising in real-life burst detection and reducing false alarms is an important issue; b) more comprehensive performance evaluation might be necessary, in particular regarding detectable burst size; c) further research on new methods employing multivariate analysis and a new category based on clustering analysis would be beneficial to tackle uncertainty; d) more focus on the use of pressure data might facilitate burst location and reduce investment in burst detection.

ARTICLE HISTORY
Received 25 May 2016
Accepted 18 December 2016

KEYWORDS
Burst detection; data-driven; water distribution system; district metering area; uncertainty

1. Introduction

Water loss in water distribution systems (WDSs) is a global challenge that has been attracting extensive attention over the past two decades. IWA/AWWA standard water balance methodology (AWWA 2009) considers leakage an important form of water loss and defines bursts as a form of leakage characterized by short duration but typically high flow. Although the overall water loss of a burst is typically less than that of a leak which has a small flow rate but a much longer run time, the negative effects of bursts are non-negligible. Other than the significant waste of water associated with bursts, the abrupt release of pressure during bursts causes energy dissipation and may also lead to contamination intrusion through broken pipes (Karim *et al*. 2003; Fox *et al*. 2016). For water companies, this causes economic losses. From a regulatory point of view, bursts are significant because, when supply is interrupted or water quality is affected, consumers become dissatisfied. Many countries have developed performance indicators or policies to measure and regulate the service levels of water companies. Customer Minutes Lost (CML), which is defined as the average number of minutes per year that a customer does not receive any water or water of a quality that doesn't meet the legal standards, is a performance indicator proposed in the Netherlands (Bakker *et al*. 2012). The service incentive mechanism (SIM) in the UK is based on a service level evaluation for water companies and can affect a company's earnings.

Burst duration is determined by unawareness, awareness, location, isolation and repair periods (Mounce and Boxall 2010; Bakker *et al*. 2012). Many large bursts are reported by consumers whereas some bursts that are relatively small or happen late in the evening may stay unnoticed (Bakker *et al*. 2014a; Tao *et al*. 2014). This causes a long period of burst unawareness and intensifies the negative influences of bursts. Bursts can occur both on mains and service connections (Farley and Trow 2003) and various factors including poor pipe condition, inappropriate operation and extreme weather may cause bursts (e.g., freezing weather may lead to the cracking of pipes). Consequently, it is difficult to know when and where bursts happened. This indicates that technologies and strategies are needed to detect bursts in a timely manner. Timely burst identification will facilitate location, isolation and repair, which reduces the run time of the burst.

Various methods have been studied and developed to shorten the unawareness time by detecting bursts quickly. A great deal of research has been conducted into data-driven approaches to detect bursts since Mounce *et al*. (2002) introduced artificial neural networks (ANNs) for burst detection. Mutikanga *et al*. (2013) reviewed tools and methods that vary from simple performance indicators to highly sophisticated optimization algorithms for water loss management. Puust *et al*. (2010) considered the overall leakage management process starting from when the leak is identified through to controlling the level of leakage. Li *et al*. (2015) focused on burst detection/location and divided all methods into two classes: hardware-based methods and software-based methods. Although data-driven approaches have been identified as promising tools to detect bursts in the literature, a comprehensive review of their performance capabilities and limitations has not yet been presented in any review paper.

CONTACT Shuming Liu shumingliu@tsinghua.edu.cn

Practicality is vital for a burst detection method to be launched in the field. Hardware methods using acoustic or non-acoustic equipment are available to detect and locate leakage in a WDS (Li *et al*. 2015). However, this kind of method is expensive, time consuming, and labour intensive, which means application for fast burst detection can be difficult. Colombo *et al*. (2009) reviewed and classified transient-based methods. One technique focuses on directly detecting negative pressure waves caused by pipe bursts (Misiunas *et al*. 2005; Srirangarajan *et al*. 2013; Lee *et al*. 2016), which relies on frequent data collection and has high transmission costs. However, the burst-induced pressure signals may be masked by background noise and other events in complex networks. The other technique analyses subsequent behaviour of burst-induced transient signals (Liggett and Chen 1996; Mpesha *et al*. 2001), but studies were generally based on numerical simulations or heavily controlled laboratory work (e.g. a single pipe with consistent diameter). Although a few field tests were carried out and promising results were presented (e.g., successful detection and location of some simulated leakages), real networks used for testing were small and pipe materials within these networks was homogenous. More field investigations in complex networks with high uncertainty should be carried out. Some methods based on hydraulic models were also developed to detect leakage and bursts (Bicik *et al*. 2011; Islam *et al*. 2011; Meseguer *et al*. 2014). The core of model-based methods is calibration and even the burst detection itself is regarded as part of calibration (Wu *et al*. 2010; Sanz *et al*. 2016). However, well-calibrated hydraulic models are not widespread in water companies and maintaining the models is also one of the main issues. Furthermore, these kinds of methods are mainly tested on simple networks using synthetic data. Data-driven approaches, on the other hand, have been validated in real-life WDSs with the development of hydraulic monitoring sensors and computer technologies. These techniques, based on empirical observation of behaviours in networks, have the ability to mine useful information from monitoring data that are too complex or numerous for humans to process and analyse in a timely manner (Romano *et al*. 2014a). For example, Mounce *et al*. (2010) applied an online artificial intelligence system on 144 district metering areas (DMAs) in the UK for 2 months and achieved promising results.

Differing from previous reviews, this paper summarizes different methods completely based on hydraulic data collected from Supervisory Control and Data Acquisition (SCADA) systems and aims to discuss capabilities and limitations of these data-driven methods. Section 2 classifies all data-driven methods into three categories and this classification work has not been implemented in other papers. Section 3 introduces performance evaluation criteria and compares the performance of different methods. Then, Section 4 analyses uncertainty present in monitoring data and methods themselves, which form the limitations of data-driven approaches. Techniques that attempt to reduce uncertainty or minimize effect of uncertainty are also discussed. Finally, some future work is recommended and conclusions drawn in Section 5 in order to promote development in the field of burst detection.

2. Categories of data-driven approaches

Masses of hydraulic data are collected every day from WDSs. In an ideal scenario (without any other events such as fire hydrant use and bursts), all these data follow specific periodic patterns (daily, weekly, and yearly). In a typical diurnal demand pattern, two water usage peaks are in the morning and evening and the minimum demand presents at night. When a burst happens, additional demand will generate a spike in the diurnal demand pattern (flow data follows a similar pattern and more details are provided in Section 4.1). Therefore, data representing bursts are deemed as outliers because of significant distinction with other data generated by normal consumers' water usage (i.e., normal data). Note that the data caused by weather changes or festivals are also normal data, because these data are generated by consumers and have different features compared with those caused by fire hydrant use or bursts (more details are provided in Section 4.1). There is an abundance of normal data in a monitoring database, whereas outliers are scarce (Mounce *et al*. 2011). Taking that into consideration, anomaly detection or deviation detection, a common data mining task, is proposed to identify outliers (Tan *et al*. 2005). Various data-driven approaches applied in burst detection also belong to anomaly detection.

Table 1 summarises data-driven approaches based on different techniques and all the methods are divided into three categories. Most of these methods were validated on DMA-level in real life. As shown in Table 1, flow and pressure (the two variables that are most commonly under scrutiny in WDSs) are extensively requested and water demand calculated from flow is also used in some studies.

2.1 Classification method

In the first category, a classification model is constructed to distinguish bursts from normal data. Mounce and Machell (2006) applied static and time-delay ANNs to detect bursts and these two ANNs have different neural architectures. The time-delay ANN adds a temporal dimension to express a relationship between inputs in time with a sliding window (i.e., the input of the network contains both the current and previous signals). The dynamic property made the time-delay ANN outperform the static ANN when working on time series (like hydraulic monitoring data in WDSs). Self-organizing map (SOM) ANN facilitated by a leak function was used in leakage detection (Aksela *et al*. 2009). Outputs of these ANNs are a positive value that ranges from 0 to 1. A burst or leakage alarm will be raised if the value exceeds a defined level (e.g., 0.75), which achieves the purpose of classifying monitoring data. For ANNs used for classification, the process of training is crucial and both normal data and outliers (i.e., burst or leakage) are indispensable in a training set, in order to give adequate information for ANNs to learn. The data requirement is also the main limitation of these approaches (more details are provided in Section 4.2).

2.2. Prediction-classification method

In this category, classification is carried out after a stage of prediction. Differing from the classification models in the first category, only normal hydraulic data are needed for prediction models and a data selection process to exclude outliers is an important issue in this category (more details are provided in Section 4.2).

Table 1. Summary of data-driven approaches.

Reference	Data requested	Technique adopted	Category
Mounce and Machell 2006	Flow	Static/time-delay ANN	Classification method
Aksela *et al.* 2009	Flow	Self-organizing map ANN	
Mounce *et al.* 2002 Mounce *et al.* 2003	Flow/pressure	MDN and classification module	Prediction-classification method
Mounce *et al.* 2007 Mounce *et al.* 2010	Flow	MDN and FIS	
Mounce *et al.* 2011	Flow/pressure	Support vector regression	
Eliades and Polycarpou 2012	Flow	Fourier series and CUSUM	
Romano *et al.* 2010 Romano *et al.* 2014a Romano *et al.* 2014b	Flow/pressure	ANN, SPC and BIS	
Bakker *et al.* 2014a Bakker *et al.* 2014b	Demand/pressure	Adaptive forecasting model and deviation analysis	
Hutton and Kapelan 2015a	Demand	Probabilistic demand forecasting model	
Ye and Fenner 2011	Flow/pressure	LKF	
Ye and Fenner 2014b	Flow	Polynomial function based on weighted least squares with EM algorithm	
Jung and Lansey 2015	Nodal group demand	NKF and SPC	
Laucelli *et al.* 2015	Flow/pressure	EPR Paradigm	
Palau *et al.* 2012	Flow	PCA and SPC	Statistical method
Jung *et al.* 2015	Flow/pressure	SPC	
Loureiro *et al.* 2016	Flow	Modified SPC	

Notes: ANN = artificial neural network; MDN = mixture density network; FIS = fuzzy inference network; CUSUM = cumulative sum; SPC = statistical process control; BIS = Bayesian inference system; LNF = Linear Kalman filter; NKF = Nonlinear Kalman filter; EM = expectation maximization; EPR = Evolutionary Polynomial Regression.

Ye and Fenner (2011) utilized a linear Kalman filter (LKF) to detect bursts. The filter provides a statistical description of the current state of dynamic systems with consideration of all historical data. Trained by normal data, LKF gives an optimal estimation at each time step. Furthermore, when the filter converges using training data with system knowledge, it only needs current observation for a prediction, making it an efficient method. Ye and Fenner (2014b) fitted historical data using polynomial functions based on weighted least squares so that normal flow or pressure values could be forecasted. Additionally, the performance of fitting was improved after an expectation maximization (EM) algorithm was introduced. When a burst occurs in a DMA, sensor measurements will differ from the predicted values significantly because the prediction relies on normal data. Though attempts have been made within these methods to predict normal hydraulic conditions, the classification is relatively simple and involves computing the absolute difference between predicted and observed values. Except for the simplest method, diverse classification strategies were developed.

Bakker *et al*. (2013) succeeded in forecasting water demand with a model that adaptively learns normal demand patterns for DMAs. Deviation analysis (i.e., the classification stage) was used in burst detection supported by the prediction results (Bakker *et al*. 2014b). A threshold value was determined offline through analysing the deviations (i.e., the difference between predicted values and actual observations) in the year prior to the detection period. When it comes to real-time detection, an alarm was triggered as long as the deviation exceeded the threshold value. A similar classification method was employed by Mounce *et al.* (2011). After the prediction stage using support vector regression (SVR), deviations are also analysed and only if several abnormal observations occur consecutively can an event be determined. ANNs are still extensively used in the stage of prediction (Mounce *et al*. 2002; Mounce *et al*. 2003; Mounce *et al*. 2007, 2010; Romano *et al*. 2010; Romano *et al*. 2014a, b). Instead of a value that varies from 0 to 1, the outputs of ANNs here are predicted values of flow or pressure. That is to say, the prediction results provide no information about classification. After predicting a sensor's readings using an ANN called mixture density network (MDN), a classification module was applied to raise a burst alarm (Mounce *et al*. 2002; Mounce *et al*. 2003). By employing fuzzy inference system (FIS) in which knowledge is encoded explicitly in a rule base, Mounce *et al*. (2007) improved the classification stage in the form of mimicking human reasoning. A sophisticated Bayesian inference system (BIS) was also proposed to identify bursts or other events (Romano *et al*. 2010; Romano *et al*. 2014a, 2014b). In this study, ANN was used to forecast short-term hydraulic values and statistical process control (SPC) was used to analyse event-induced measurement variations, generating evidence for the BIS that comprised of a Bayesian network. For these classification strategies, more reliable detection results are acquired. On one hand, domain experts' knowledge can give instructions to identify a burst. The FIS and the BIS are both expert systems. On the other hand, some methods employ historical data (including burst data) to extract useful information. For example, Bakker *et al*. (2014b) set alarm thresholds after analysing yearlong historical deviations. Romano *et al*. (2014b) fine-tuned some algorithm's parameters using historical event data to improve the performance of BIS. Furthermore, the FIS and BIS both generate probabilistic results to deal with uncertainty. Hutton and Kapelan (2015a) also compared a forecasting result with the corresponding monitoring value expressed in the form of probability (more details are provided in Section 4.2).

There is one thing to note. All aforementioned methods implement classification by evaluating the difference between predicted and actual observed values. Differing from all these methods, Eliades and Polycarpou (2012) used adaptive inflow approximation and fault detection to realize prediction and classification respectively. By updating coefficients of a Fourier series, dynamic DMA inflows that include both normal flow conditions and leakage faults are learned, indicating that the deviation is also minimized even with the presence of a burst or leakage in the stage of prediction. However, leakage can lead to changes in some parameters of Fourier coefficients. Consequently, the cumulative sum (CUSUM) algorithm is employed to detect changes in the mean value of a parameter. An alarm occurs when a threshold selected offline is surpassed. Laucelli *et al.* (2015) forecasted hydraulic values using an evolutionary polynomial regression

(EPR) paradigm. When it comes to classification, the prediction values are not compared with actual observations. Cumulative probabilities of prediction values are calculated under a predefined conditional probability density function (PDF) and a big enough cumulative probability indicates an event.

2.3. Statistical method

In the third category, no prediction or classification model is constructed and burst detection completely relies on statistical theory. Specifically, SPC is the key technique in this category. SPC utilizes control charts, a graphical and analytic tool with a set of control limits, for monitoring process variation (i.e., outliers caused by bursts among a sensor's measurements).

Jung *et al*. (2015) compared different SPC methods, including three univariate and three multivariate methods. For univariate methods, Western Electric Company (WEC) rules can identify one outlier or some consecutive outliers beyond multifold standard deviations. CUSUM and exponentially weighted moving average (EWMA) control charts both incorporate past data, making them have a longer memory than WEC rules. For multivariate methods, CUSUM and EWMA control charts have multivariate versions to deal with measurement from multiple sensors and the Hotelling T^2 control chart with an elliptical control limit uses Mahalanobis distance for hydraulic data from different sensors to identify a burst. Palau *et al*. (2012) performed principal component analysis (PCA), building a PCA model to transform a flow data matrix to a lower dimension. This model then applied Hotelling T^2 and distance to model (DMOD) to identify severe and moderate outliers for both historical data and data to be detected. Loureiro *et al*. (2016) developed a modified SPC method with a purpose to reduce false alarms (more details are provided in Section 4.2)

Actually, SPC methods prevail in the field of burst detection though they only contribute to one module or one stage in other categories. For example, CUSUM algorithm was used to identify leakage faults in the stage of classification (Eliades and Polycarpou 2012). Evidence for events is provided by WEC rules, insuring the implementation of BIS (Romano *et al*. 2010; Romano *et al*. 2014a, 2014b). Jung and Lansey (2015) employed CUSUM and Hotelling T^2 in the stage of classification to detect bursts after estimating nodal group demands. SPC methods are also used for data pre-processing (more details are provided in Section 4.2).

3. Performance evaluation

3.1. Criteria of performance evaluation

To evaluate the performance of data-driven approaches for burst detection, this paper details three of the performance evaluation criteria found in the literature that are deemed to be the most appropriate for the burst detection problem.

No matter what method is employed, the result of burst detection only has two possibilities: burst or non-burst. Therefore, true positive rate (TPR) and false positive rate (FPR) are widely used in this field to evaluate accuracy (Mounce and Machell 2006). In addition, to shorten the unawareness period of a burst, detection time (DT) is also an important indicator for the performance of a method (Mounce *et al*. 2010).

When a burst detection method is used in real WDSs, water companies will pay significant attention to the detection accuracy. For a binary classifier, accuracy is composed of two parts: sensitivity and specificity. Sensitivity means the ability to detect real bursts, while specificity is the ability to correctly exclude normal data from being classified as bursts (McKenna *et al*. 2008). It is obvious that sensitivity has a negative correlation with specificity; namely, a higher sensitivity with more bursts detected will have a greater number of false alarms (i.e., lower specificity). To quantify the two aspects of accuracy, TPR and FPR are calculated:

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$1 \text{ - specificity} = \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{2}$$

where TP (true positive) is the detection of an actual burst; FP (false positive) refers to an incorrectly classified burst; TN (true negative) refers to a correctly identified normal situation; FN (false negative) is the situation in which an actual burst is not detected.

According to the above analysis, every burst detection method needs to balance sensitivity and specificity, which can be depicted using a receiver operating characteristics (ROC) curve (Egan 1975). The ROC curve defines the TPR as a function of the corresponding FPR. Furthermore, the area under the curve (AUC) gives quantified information about the effectiveness of a method (Hanley and Mcneil 1982). When comparing two burst detection methods on the same data-set, the one with a greater AUC is recommended. Considering that bursts happen rarely in a WDS, a good detection method must keep FPR small even at the expense of reducing TPR to within an acceptable range (Metz 1978). Otherwise false alarms will burden workers in water companies and there is a risk that workers will start ignoring alerts.

Detection time (DT) can be defined as a period between the beginning and the detection of a burst. The value of DT determines the duration of awareness time of a burst and relates to the damage caused by a burst. When calculating the DT for a burst detection method, it is always necessary to carry out some simulated bursts by opening fire hydrants or drain valves (Romano *et al*. 2014a). The reason is that the burst information recorded by water companies may not contain the exact start time of the burst, which causes DT acquired through historical bursts to be inaccurate. Not all studies have conducted burst simulations to evaluate performance and some studies used information in water company records to directly calculate the DT (Bakker *et al*. 2014b; Loureiro *et al*. 2016). This is partly because burst simulations can result in unsteady conditions (e.g., pressure drop and deterioration of water quality) for consumers and potentially lead to dissatisfaction. In addition, some studies have used synthetic data (e.g., leakage simulated by adding random values to real flow data or data generated by hydraulic models) to evaluate the performance of proposed methods (Eliades and Polycarpou 2012; Jung *et al*. 2015). Therefore, not all DTs calculated in these studies are accurate and the DT might be longer when used in real life.

Other than the three commonly used criteria discussed above, detectable burst size is an important aspect of a burst detection method. It is a supplementary to TPR and FPR because some methods may only have the capability to detect relatively large

Table 2. Performance of some data-driven approaches.

Reference	TPR (%)	FPR (%)	DT	Burst size in test** (% of average DMA inflow)	Testing data	Sampling rate (min)	Category
Mounce and Machell 2006	75	0	-	2%-10%	Offline flow data	1	Classification method
Mounce *et al.* 2010	-	15	Several hours	-	Online flow data	15	Prediction-classification method
Romano *et al.* 2014a	-	8	Mostly within 1 h	5%-16%	Offline flow and pressure data	15	
Ye and Fenner 2014b	-	4	Mostly within 15 min	10%-50%	Offline flow data	15	
Bakker *et al.* 2014b*	-	-	5-10 min	Over 20%	Offline demand data	5	
Eliades and Polycarpou 2012	94.5	0	9.8 days (average)	1.5%-10%	Synthetic flow data	5	
Loureiro *et al.* 2016	80	10	At least 30 min	-	Offline flow data	15	Statistical method

Notes: *The TPR and FPR in this study cannot be compared with those in other studies, because the definitions of TPR and FPR are different; **This paper defines bursts with volume over 20% of average DMA inflow as relatively large bursts and those below 5% as small leaks; TPR = true positive rate; FPR = false positive rate; DT = detection time.

bursts (Bakker *et al*. 2014b; Ye and Fenner 2014b) and some other methods can even capture small leaks (Aksela *et al*. 2009; Eliades and Polycarpou 2012). The TPR can be small when many known small leaks exist in the testing data-set, even if the method being tested is effective for detecting relatively large leakages (Bakker *et al*. 2014b). On the contrary, the FPR might be large if the method can perceive small leaks. Furthermore, the size of an area (a DMA) affects the detectable burst size. Bakker *et al*. (2014a) found that the detectable burst size varies with different area sizes of a DMA and gave a function to describe this correlation. However, this criterion has not been widely used and properly discussed in most previous studies. Consequently, the existing performance evaluation is one-sided and no adequate information is provided to water companies for real life application.

3.2. Performance of data-driven approaches

Table 2 summarizes the performance of some data-driven approaches from the three categories. Noticeably, burst size used for testing, data sampling rate and testing data for evaluation are also used to compare the performance of different methods.

As shown in Table 2, the classification method developed by Mounce and Machell (2006) acquired an ideal FPR-zero. However, the offline flow data used for cross validation and testing both contained the same simulated burst data (created by opening hydrants), which reduces confidence in the performance of this method. Eliades and Polycarpou (2012) also reported that the FPR was zero, but the result was generated by using synthetic data (adding random values to real flow data to simulate leakage). There was no information provided to demonstrate the performance when tackling real-life data. For other methods in Table 2, all data used to evaluate the performance are real and the value of FPR is relatively high. As an automated online method, Mounce *et al*. (2010) achieved an FPR as large as 15%, showing great challenges for burst detection in real life application, particularly due to uncertainty in network information (activities and events) when monitoring and analysing a large number of sensors. Therefore, the relatively high FPR is still a major issue for data-driven approaches. More discussion on this is provided in Section 4.

The DT is influenced by several aspects. Alarm rules of a method and data sampling and communication frequencies directly affect the value of DT. To keep the FPR small, many methods won't raise an alarm whilst detecting a data anomaly. Mounce *et al*. (2010) set a 12-h window of analysis for the FIS in online use, leading to a long DT of up to several hours. In addition, the sampling interval for burst detection in previous studies ranges from 1 min to 15 min. Though a fifteen-minute sampling rate is standard in the UK, Mounce *et al*. (2012) found that high-frequency data does improve the performance of burst detection. Loureiro *et al*. (2016) made a rule that an alarm is activated after three abnormal observations. The DT is at least 30 min because the sampling interval was 15 min in this study. If the sampling rate is 5 min, the shortest DT will be reduced to 10 min. Actually, the DT could be longer when a method is used in real life application, because data communication frequency may be lower than the sampling rate. For example, it could be every 30 min or daily in the UK. Using the same sampling frequency (e.g., 15 min) to detect a relatively large burst, a mature data-driven method will definitely require a shorter DT when the data communication frequency is 30 min rather than one day. However, high data sampling rate and data communication frequency always means a big investment in telemetry facilities. For important areas in mega cities or water mains, both high sampling rate and communication frequency (e.g., 5 min) might be necessary.

Detectable burst size also has some relationship with the DT. It is difficult to compare the detectable burst size in different studies because the water demand of tested DMAs, occurrence time of bursts and burst duration are different. Considering that, this paper takes percentage of average DMA inflow as a criterion and defines bursts with a volume over 20% of average DMA inflow as relatively large bursts and those below 5% as small leaks. Some methods in Table 2 (references 3 to 6 in the table) show that the smaller the burst size a method can detect, the longer the DT is. The DT may take several days to detect small leaks, while relatively large bursts can be quickly detected in a period of 5–10 min. The reason is that detecting small leaks always needs much more newly collected data to identify whether a trend that deviates from the original data series is present or not.

It is noteworthy that the data type highly impacts on the performance of data-driven approaches. In practice, flow and pressure sensors are installed at inlets of a DMA. Some other pressure sensors are located at critical points (e.g., a location with highest elevation or the point farthest from the inlet) in this area (Ye and Fenner 2011; Romano *et al*. 2014a). Pudar and Liggett (1992)

Table 3. Uncertainty of data in burst detection.

Source of uncertainty		Reason	Methods to tackle
Measurement errors		Defect of sensors	Sensor improvement and maintenance
			Data denoising
Missing data		Sensor fault	Hardware improvement
		Transmission failure	ARIMA, Linear interpolation
Demand variation	Periodic variations	Daily/weekly fluctuations	Data transformation
			Data decomposition
		Seasonal changes	Online retraining/updating
	Non-stationary changes	Weather changes	Multivariate analysis
		Festivals	Pattern matching
		Unexpected activities	Using nodal demand
		Maintenance work	Additional metering
		Operational condition changes	

concluded that the sensitivity of pressure data to a burst depends on local pipe friction parameters and a pressure sensor's location, causing pressure measurements to be less sensitive when the number of pressure sensors is limited. Under this condition, flow is proven to be more sensitive to bursts that happen in a DMA (Mounce *et al*. 2011; Ye and Fenner, 2011). Therefore, the data-driven approaches summarised in Table 1 did not utilize pressure data alone to detect bursts. Demand data were also used in some studies, but the drawback is that demand is only sensitive to relatively large bursts (Bakker *et al*. 2014b; Jung and Lansey 2015).

4. Uncertainty in burst detection

For data-driven approaches aimed at burst detection, relatively high FPR is still a challenge in real life application. A great number of false alarms are raised because of high uncertainty and this pervades all fields of scientific research and hinders effective decision-making (Samson *et al*. 2009). It was reported that uncertainty is often driven by particular activities, such as measurement or modelling (Walker *et al*. 2003). Similarly, uncertainty in burst detection mainly derives from monitoring data, and limitations of adopted methods introduce new uncertainty and aggravate the situation.

4.1. Uncertainty in monitoring data

As discussed in Section 2, burst detection is regarded as a data mining task of anomaly detection. However, WDSs cannot operate under an ideal situation in which all data follow some known rules. In the process of monitoring, dirty data, which could mean missing data, incorrect data or data that behaves stochastically, is always present. Water demand in a network varies due to a diversity of events, making the monitoring non-stationary. As a consequence, outliers caused by bursts are prone to be shadowed by non-stationary monitoring data. Data with high uncertainty pose an obstacle to precise and effective burst detection. Analysing uncertainty in monitoring data is meanwhile a process of understanding data and this process is the primary phase for any data mining task (Han *et al*. 2011). To fully understand monitoring data in WDSs, various sources of uncertainty are presented in Table 3.

Measurement errors are generated during data collection and are mainly dependent on the performance of sensors. In pipes with low flow velocity, especially during the early morning with little water usage, some sensors can hardly collect actual flow through a pipe because the sensors are designed to measure values within a certain range. Some older sensors in poor condition may lead to abnormal fluctuations, making the measurements deviate from true values. Therefore, more accurate sensors should be developed and this is important for water audits. Water companies should select proper meters (e.g., zone meters) and pay more attention to equipment maintenance. For pressure monitoring, signal noise is present in measurements and restricts the use of pressure data. In order to retain important information in the original data and improve the accuracy of the ANN prediction model, Romano *et al*. (2014a, 2014b) employed wavelet analysis to remove noise in pressure data.

In the process of data collection, data are often absent during a period of time because of sensor errors or transmission failure. In a WDS, many sensors are powered by batteries with limited lifetime and they cannot work without electricity supply (Mounce *et al*. 2012; Ye and Fenner 2014a). Furthermore, the signal is weak in rural areas and monitoring data are prone to transmission problems (Ye and Fenner 2014b). Missing data can violate the continuity of monitoring data, which has a negative impact on online use of data-driven approaches. A great deal of missing data even leads to insufficient historical data for the implementation of a burst detection method. To work out this problem, more reliable sensors and expanded coverage of communication networks are radical solutions. However, significant cost is associated with improving hardware facilities. Alternatively, Romano *et al*. (2014a) used linear interpolation to fill in missing values after abandoning data with severe deficiency. Auto-regressive integrated moving-average (ARIMA) modelling is also applied to deal with missing data (Mounce *et al*. 2010). Although these techniques are more feasible, the filled values have relatively high uncertainty, which may reduce the reliability of burst detection results.

Considering that most methods employed flow or demand data, elaborate discussions on demand variation are provided here. Demand variation is the most complex part of data uncertainty. Figure 1 presents typical data over 24 h from a flow meter in a real DMA. In Figure 1A, pipeline flushing (downstream from the DMA) was carried out between 0:45 and 4:30, while Figure 1B displays the flow data for another 24 h in which a real burst event was recorded between 9:35 to 10:05. From Figure 1, it is clear that flow meter readings change due to variation in user demand, which forms a diurnal pattern. However, this kind of periodic fluctuation varies significantly and nonlinearly with the change of time, making it difficult to model the diurnal pattern. By contrast,

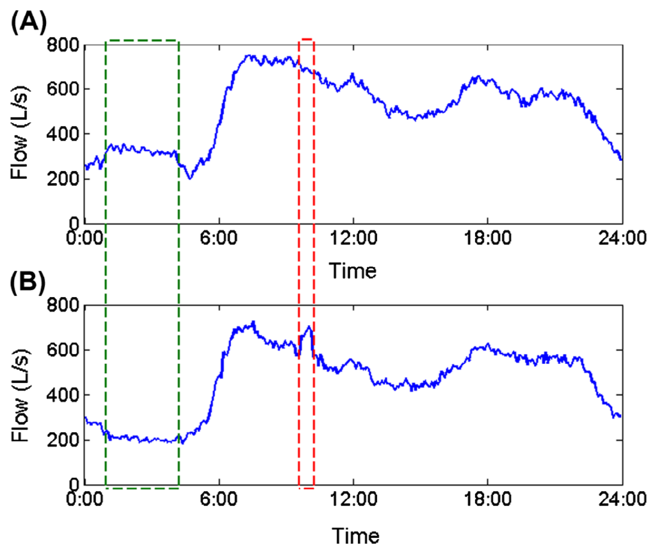

Figure 1. Two days of flow data from a DMA. (A) A day that includes pipe flushing (carried out downstream from the DMA); (B) A day that includes a real burst.

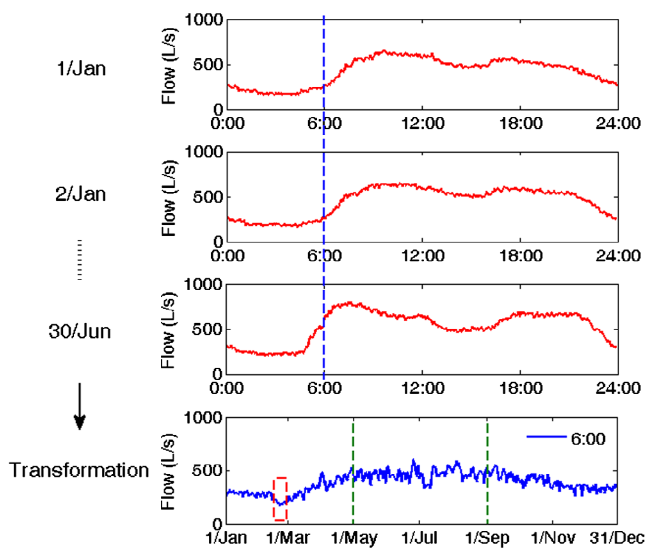

Figure 2. Demonstration of the transformation for time series data.

the flow measurements at a specific time (e.g. 6:00) do not vary significantly with the change of days. Taking that into consideration, data from a specific time in a consecutive time series was extracted to form daily series (Mounce *et al*. 2011; Jung *et al*. 2015). The transformation process is demonstrated in Figure 2 (data are collected in a WDS in South China). The diurnal pattern varies between weekdays and weekends. To further reduce variation, Ye and Fenner (2014b) utilized weekly series where the time interval between two flow measurements is one week instead of one day. As presented in Figure 2, the diurnal patterns in different months are distinctive, making the transformed flow data non-stationary with changes in season. It is obvious that the data are quasi-stationary between 1 May and 1 September when the climate is warm in South China. In other months, water demand decreases and forms a trough in February (highlighted by the red rectangle with dash line in Figure 2) because people leave cities to return to their hometown during Chinese New Year. For the yearly seasonal changes, some data-driven approaches implement continual online retraining/updating to weaken this long-term trend (Mounce *et al*. 2010).

Apart from data transformation, Eliades and Polycarpou (2012) decomposed the original flow time series into two components according to statistical theories. The first component involves long-term trends and yearly seasonal changes, while the second one is weekly periodic changes. Taking the two components into account, a multiplicative formulation was applied to model the flow data.

Outliers caused by bursts are supposed to distinguish themselves after data transformation or data decomposition. However, data fluctuations can be caused by different reasons apart from the above periodic variations. Figure 1 shows that both maintenance work (e.g., pipe flushing) and pipe bursts result in an increase of flow measurements. Ye and Fenner (2014b) pointed out that water usage in a WDS may increase suddenly because of weather changes and unexpected activities (e.g., unplanned industrial demand or fire fighting). In addition, abnormal fluctuations caused by operational condition changes (e.g., the number of operating pumps and closure and opening of valves) were not identified from detected outliers in most above-mentioned methods. Some festivals with mass population movements also significantly affect the water usage in a WDS (like Chinese New Year). All these non-stationary changes in water demand will enhance the difficulty of burst detection and increase the FPR.

Note that periodic variations and some non-stationary changes (i.e., weather changes and festivals) are caused by normal consumers' water usage. Therefore, corresponding monitoring data, which have different features compared with other non-stationary data (i.e., true outliers are caused by unexpected activities, maintenance work, operation condition changes and bursts), are normal data. For these normal data in disguise, data transformation or decomposition can reduce periodic variations. Some identification techniques or ideas have been proposed to tackle the data generated by weather changes and festivals. Romano *et al*. (2014a) concluded that a hot day can cause a global anomaly in a WDS (i.e., simultaneous increase of water usage in all DMAs), making it possible to exclude false alarms raised by weather changes (bursts only cause increase of water usage in a certain area). In fact, it is the correlation/comparison across a number of sensors. Pattern matching is also an effective solution, Mounce *et al*. (2014) applied this approach to automatically recognise specific waveforms in a time series based on their shapes and set up a system called Advanced Uncertain Reasoning Architecture (AURA) Alert in order to distinguish between different types of events.

The true outliers caused by unexpected activities, maintenance work and changes in operating conditions, will be important contributors to false alarms. To reduce FPR, corresponding strategies are also proposed. Considering that demands are independent of operating conditions, nodal group demands were proposed by Jung *et al*. (2015). These demands are calculated using measurements from multiple flow meters. When large consumers with consumption that accounts for a significant proportion of total demand in a DMA are continuously metered, unexpected consumer demand can be identified (Loureiro *et al*. 2016).

Table 4. Limitations of data-driven approaches.

Category	Limitation	Method improvement
Classification methods	Lacking labels of hydraulic data to train and test models Easily affected by unbalanced class sizes	Unsupervised models
Prediction-classification methods	Propagation of data uncertainty	Ensuring stationary conditions in historical data using statistical tests before model construction Removing abnormal data in an unsupervised manner during model construction
	Misleading results because of deterministic model outputs	Developing probabilistic methods to express the degree of conviction in model outputs
Statistical methods	Inappropriate distribution assumptions	Selecting robust statistics Using asymmetric control limits to fit imperfect data

4.2. Method issues and limitations

As discussed in Section 2, data requirement is a main limitation of classification methods. Considering that monitoring data can vary irregularly for different reasons (summarised in Table 3), water companies failed to record every abnormal water usage and the corresponding reason. Consequently, hydraulic data lack labels (i.e., burst or non-burst), which are essential for training and testing classification models (Mounce *et al*. 2010). From another perspective, the amount of burst-induced data is far less than that of normal data and the imbalance of different classes of data will affect the accuracy of classification (Oliker and Ostfeld 2014). Therefore, classification methods are impractical in real life and related studies are rarely presented. However, Aksela *et al*. (2009) reported that the unsupervised SOM ANN is capable of distinguishing leak data from normal data by finding similarities among all input data, although the training data are not explicitly labelled. That is to say, unsupervised models might be worthy of further study to overcome the limitations of classification methods.

As discussed in Section 2, most prediction-classification methods fit normal hydraulic conditions and detect bursts by evaluating the difference between predicted and observed values. However, misjudgements may appear because of prediction and classification limitations.

In one aspect, data uncertainty, in particular non-stationary changes in historical data, will propagate to predicted values (Hutton *et al*. 2014) and affect the accuracy of burst detection. To minimize the adverse impact, selection of historical data is needed to assist with model construction in most studies. Data selection is in the data pre-processing module and statistical methods are widely used in this process (Mounce *et al*. 2002; Mounce *et al*. 2003; Mounce *et al*. 2010; Romano *et al*. 2014a). An assurance test which compares mean and variance of data is indispensable to assessing if historical data are almost stationary before training the ANNs (Mounce *et al*. 2010). Similarly, Romano *et al*. (2014a) employed three statistical tests to filter out event-induced outliers in historical data. In these tests, SPC methods were used to assess how the mean of hydraulic data changes over time. Exceptionally, Ye and Fenner (2014b) used an expectation-maximization algorithm to automatically exclude abnormal data in the process of modelling historical data. In other words, previous data selection was abandoned.

From another aspect, many studies presented a deterministic forecast of hydraulic data and no information about prediction errors was provided (Ye and Fenner 2011, 2014b; Bakker *et al.* 2014a, 2014b). However, in the presence of scientific uncertainty, single-valued predictions are misleading (Handmer *et al*., 2001). When it comes to the stage of classification, a deterministic result also misleads decision makers. Brown (2004) defined uncertainty as a state of confidence, which means degree of trust or conviction in knowledge. To consider the uncertainty in prediction and classification and express the degree of conviction, some probabilistic methods were proposed. An MDN is a mixture density model combined with ANN and Mounce *et al*. (2002) used this model to produce a PDF for each input. Differing from deterministic forecast, the PDF expresses uncertainty in prediction results. Then the FIS can evaluate the PDF and observations to detect bursts and generate a subjective probability to describe the degree of trust in the logic judgement (Mounce *et al*. 2010). Hutton and Kapelan (2015b) developed a probabilistic demand forecasting model, which presents a statistically robust description of prediction errors. Similar to the MDN, the output of the model is a PDF of future demand under normal conditions. By comparing with the PDF, a probability that actual demand is greater than the prediction is acquired and an alarm is raised when the probability is large enough (Hutton and Kapelan 2015a). Romano *et al*. (2010) introduced a Bayesian network, which allows reasoning under uncertainty, to infer whether an event happens in a DMA. With evidence generated by the SPC methods, a probabilistic output of the network can provide enough information for making decisions. Furthermore, parameters in the network were recalibrated using the EM algorithm (Romano *et al*. 2014b). As a result, the network gained knowledge of historical event records and the inference results became more convincing.

Statistical methods are concise because no sophisticated prediction or classification models are needed, but their results have high uncertainty. In this category, SPC methods imply that process variables are normally distributed (Jung *et al*. 2015). In other words, distributions of hydraulic data are always assumed to be Gaussian in most statistical methods. However, the distributions are easily coloured by non-stationary changes summarised in Table 3, forming asymmetric features in these distributions. Under this condition, Loureiro *et al*. (2016) renewed outlier regions (i.e., control limits) by selecting some robust statistics instead of mean and standard deviation. Most importantly, an outlier region considering asymmetric behaviour was proposed to accommodate the imperfect data. Furthermore, the decision threshold for each outlier region was determined by computing FPR and TPR. This modified SPC method has been implemented in commercial software (Loureiro *et al*. 2016).

Table 4 summarises the limitations of different methods and the corresponding improvements. According to the above discussions, classification methods are rarely studied due to the limit of real-life monitoring data. As for statistical methods, inappropriate distribution assumptions limit their application to burst detection. However, SPC techniques are indispensable and do play an important role in prediction-classification methods. With the development of probabilistic solutions, prediction-classification methods take uncertainty into consideration and generate more reliable results to support decision-making.

5. Conclusions and recommendations

5.1. Suggestions on prediction-classification methods

In this paper, burst detection is defined as a task of anomaly detection. All data-driven approaches for burst detection were classified into three categories: classification methods, prediction-classification methods and statistical methods. Some methods show great promise and have been used or validated in real applications (Mounce *et al*. 2010; Romano *et al*. 2012; Loureiro *et al*. 2016). As discussed in Section 4, classification methods and statistical methods both have limitations. Prediction-classification methods, on the other hand, deal carefully with uncertainty and are the most popular methods for use in this field.

Current studies mainly utilized monitoring data for burst detection and omitted factors that can provide additional information to evaluate network behaviours. Exogenous information (e.g., weather forecast and pipe attributes) could be included as additional inputs of a prediction model to improve its performance. Taking this into consideration, ANN-based prediction-classification approaches may be more capable of accurately detecting bursts because ANNs can process different kinds of inputs (e.g. day of week, temperature and hydraulic measurements) and model any function without explicit knowledge of the parameters involved (Romano *et al*. 2014a). Furthermore, the core of data-driven approaches is the knowledge mining of data. Consequently, model construction and calibration using historical data is preferable in the classification stage. Romano *et al*. (2014b) have shown that calibration-based BIS outperforms a purely expert-based system.

5.2. Comprehensive performance evaluation

TPR, FPR and DT are widely used to evaluate the performance of these data-driven methods. For burst detection, a small FPR should be emphasized and a short DT is preferred. However, the performance evaluation is not exhaustive. As discussed in Section 3, there is a correlation between the detectable burst size and the size of an area. Considering that not all DMAs are constructed based on standard IWA recommendations (e.g., DMAs in China are larger and more complicated and may not always be hydraulically isolated), this correlation should be considered. When applying a data-driven method in a large area with high water consumption, the absolute volume of detectable burst size can be quite high even if the percentage of average inflow is still relatively small. Under this condition, additional flow or pressure sensors need installing to reach a reasonable detectable burst size (in terms of absolute value).

In addition, a method that can detect small leaks (i.e., below 5% of average DMA inflow) might trigger more false alarms and make the detection system unconvincing (discussed in Section 3.1). The DT is always long for small leaks using data-driven approaches and Eliades and Polycarpou (2012) found that about 15% of simulated small leaks were only detected after 3 weeks. In a water company with DMAs, the workers can detect and locate small leaks much more precisely using acoustic or non-acoustic equipment in routine inspections that could be conducted monthly. Consequently, an acceptable detectable burst size of data-driven approaches should be determined and a low FPR might be more important than low detectable burst size.

5.3. New strategies to deal with uncertainty and reduce false alarms

Uncertainty in monitoring data is discussed and false alarms can be raised by these non-stationary data. Water companies should aim to improve hardware and do maintenance work, as this can provide more reliable data to implement data-driven approaches. With the presence of method limitations, burst detection results may be misleading because of the high uncertainty associated with predictions, classifications and distribution assumptions. Wu *et al*. (2016) developed a clustering based method without any distribution assumption or prediction process, which can exclude significant uncertainty. Flow data (from multiple sensors in a single DMA) were transformed into matrices where each row was considered as a vector. Each matrix contains the flow data from all sensors at a specific time (e.g. 0:00, 0:15, …, 23:45) and every row (vector) corresponds to the data in one day. When implementing the clustering analysis, similarities of vectors in the same matrix, which were calculated using Euclidean distances, were compared to detect abnormal fluctuations (i.e., the comparison of monitoring data at a specific time on different days). As a popular technique in the field of anomaly detection, clustering analysis and a new category based on this method could be the focus of future studies.

It is particularly worth mentioning that almost all studies in the three categories analysed data based on an independent sensor's measurements. To reduce false alarms, analysing the data correlation across multiple sensors is beneficial. As discussed in Section 4.1, the simultaneous increase in several flow meters' measurements probably represent more water usage caused by a hot day rather than a burst. In the field of contamination detection for WDSs, multivariate analysis has been widely used and has improved detection performance significantly (Liu *et al*. 2014; Liu *et al*. 2015). Unlike water quality monitoring, which collects multiple parameters in one site, hydraulic data only consists of flow and pressure and the two parameters may not be monitored in the same site. However, the comparison across a number of sensors in a DMA or several DMAs is also a kind of multivariate analysis. Jung *et al*. (2015) applied multivariate SPC methods for analysing hydraulic data collected from multiple sensors, but the results were not convincing because all data were generated using a hydraulic model. Mounce *et al*. (2014) employed binary correlation matrix memory to combine quantized data from different flow and pressure sensors, giving a potential solution for a DMA with multiple sensors. The system proposed by Romano *et al*. (2014a, 2014b) also enables analysing multiple DMA signals (pressure and flow) in a synergistic manner aiming at, *inter alia*, reducing false alarms. Wu *et al*. (2016) analysed the data correlation between inlets and outlets of a DMA in China and found that bursts lead to an increase in inflows and decrease in outflows. Based on the findings, this study succeeded to distinguish between bursts and some other events (i.e. weather changes and festivals). According to the above discussion, the research situation calls for further research on multivariate analysis for burst detection with the aim of reducing FPR.

5.4. Online burst detection

Any burst detection method developed ultimately must be able to work in an online environment. Monitoring devices and transmission facilities in a water company play an important role in burst detection because they determine the quality of data and the sampling and transmitting frequencies. Consequently, the TPR, FPR and DT are directly influenced by hardware basis. That is to say, hardware facilities that are well constructed and maintained will ensure good burst detection performance.

Burst detection methods must be able to deal with complex online data. This data could either be missing (e.g., due to a failure in signal transmission) or it could be extremely abnormal due to a sensor fault (possible abnormalities include signals that are extremely large or constant over a time). When this occurs over a short time (e.g., 1 h), a data pre-processing module must be used to fill or replace values before applying a burst detection method. When this occurs over a long period of time, the data pre-processing module will trigger an alarm to warn of a hardware problem. Many methods, such as ANN-based prediction-classification methods (Mounce *et al*. 2010; Romano *et al*. 2014a, 2014b), need continual online retraining and updating to reduce the uncertainty caused by periodic variations in the data (e.g., seasonal changes) and ensure the accuracy of the model. This calls for an additional updating module. Therefore, the detection method itself should only be considered the core module of an online system. Further modules (e.g., data pre-processing and updating modules) are necessary for the system to work in real-life application.

In practice, the computing power of hardware might be an issue for a method with sophisticated models (e.g., the BIS) when tackling quantities of data across a whole WDS, especially when the sampling and communication frequencies are high. To overcome this problem, concise models that need less computing power should be developed. Furthermore, computer speed is constantly improving and analysis of big datasets may not be a problem in the near future.

5.5. Burst location

Most data-driven approaches provide little information on the location of a burst. Determining a burst within a DMA is supposed to be realized by these DMA-level burst detection methods. Nevertheless, it was reported that the same burst event may lead to repeat alarms in different DMAs (Mounce *et al*. 2010; Romano *et al*. 2014b). Therefore, current burst detection methods should be improved so that they not only identify the existence of a burst when applied at the level of a whole WDS, but also identify the particular DMA where the burst is occurring. In order to better identify the location of a burst, Mounce *et al*. (2003) and Romano *et al*. (2014c) have attempted to determine the occurrence of an event in a particular DMA in the case of cascading DMAs. For a more precise burst location, Romano *et al*. (2013) employed geostatistical models to generate a probability value of a burst associated with each DMA pipe and locate the burst within a small group of pipes, which presents a promising solution totally based on data-driven techniques. Furthermore, when adequate pressure sensors are installed at appropriate sites in DMAs, it is possible to approximately locate a burst. Farley *et al*. (2013) employed the data-driven method developed by Mounce *et al*. (2010) to justify that this method can provide location information with a proper number of pressure meters. As discussed in Section 3.2, many current data-driven methods are more effective when using flow data, but the flow monitoring is more expensive than pressure monitoring. If the layout of pressure meters is optimized in an area, the investment in burst detection will reduce and burst location information can be acquired at the same time. The multivariate analysis will take full advantage of a situation where multiple pressure meters are installed in a single area (DMA).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Water Pollution Control and Treatment Science and Technology Major Project under Grant [2014ZX07406003].

References

Aksela, K., Aksela, M. and Vahala, R., 2009. Leakage detection in a real distribution network using a SOM. *Urban Water Journal*, 6 (4), 279–289.

AWWA, 2009. *Water audits and loss control programs*. 3rd ed. Denver, CO: American Water Works Association.

Bakker, M., *et al.*, 2012. Reducing customer minutes lost by anomaly detection? *In: 14th Water Distribution Systems Analysis Conference*, 24–27 September 2012 Adelaide, Australia.

Bakker, M., *et al.*, 2013. A fully adaptive forecasting model for short-term drinking water demand. *Environmental Modelling & Software*, 48 (5), 141–151.

Bakker, M., *et al.*, 2014a. Analysis of historic bursts and burst detection in water supply areas of different size. *Water Science & Technology: Water Supply*, 14 (6), 1035–1044.

Bakker, M., *et al.*, 2014b. Heuristic burst detection method using flow and pressure measurements. *Journal of Hydroinformatics*, 16 (5), 1194–1209.

Bicik, J., *et al.*, 2011. Pipe burst diagnostics using evidence theory. *Journal of Hydroinformatics*, 13 (4), 596–608.

Brown, J.D., 2004. Knowledge, uncertainty and physical geography: Towards the development of methodologies for questioning belief. *Transactions of the Institute of British Geographers*, 29 (3), 367–381.

Colombo, A.F., Lee, P. and Karney, B.W., 2009. A selective literature review of transient-based leak detection methods. *Journal of Hydro-environment Research*, 2 (4), 212–227.

Egan, J.P., 1975. *Signal detection theory and ROC analysis*. New York, NY: Academic Press.

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

Farley, M. and Trow, S., 2003. *Losses in water distribution networks: A practitioner's guide to assessment, monitoring and control*. London: IWA Publishing.

Farley, B., Mounce, S.R. and Boxall, J.B., 2013. Development and field validation of a burst localization methodology. *Journal of Water Resources Planning and Management*, 139 (6), 604–613.

Fox, S., *et al.*, 2016. Experimental quantification of contaminant ingress into a buried leaking pipe during transient events. *Journal of Hydraulic Engineering*, 142 (1), 04015036.

Han, J., Kamber, M. and Pei, J., 2011. *Data mining: Concepts and techniques*. 3rd ed. Burlington, MA: Morgan Kaufmann.

Hanley, J.A. and Mcneil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143 (1), 29–36.

Handmer, J., Norton, T. and Dovers, S., 2001. *Uncertainty, ecology and policy: Managing ecosystems for sustainability*. Harlow: Prentice-Hall.

Hutton, C. and Kapelan, Z., 2015a. Real-time burst detection in water distribution systems using a Bayesian demand forecasting methodology. *Procedia Engineering*, 119, 13–18.

Hutton, C.J. and Kapelan, Z., 2015b. A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting. *Environmental Modelling & Software*, 66, 87–97.

Hutton, C.J., *et al.*, 2014. Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation. *Journal of Water Resources Planning and Management*, 140 (2), 169–183.

Islam, M.S., *et al.*, 2011. Leakage detection and location in water distribution systems using a fuzzy-based methodology. *Urban Water Journal*, 8 (6), 351–365.

Jung, D.H. and Lansey, K., 2015. Water distribution system burst detection using a nonlinear kalman filter. *Journal of Water Resources Planning and Management*, 141 (5), 04014070.

Jung, D., *et al.*, 2015. Improving the rapidity of responses to pipe burst in water distribution systems: A comparison of statistical process control methods. *Journal of Hydroinformatics*, 17 (2), 307–328.

Karim, M.R., Abbaszadegan, M. and Lechevallier, M., 2003. Potential for pathogen intrusion during pressure transients. *Journal American Water Works Association*, 95 (5), 134–146.

Laucelli, D.R., *et al.*, 2015. Detecting anomalies in water distribution networks using EPR modelling paradigm. *Journal of Hydroinformatics*, 18 (3), 409–427.

Lee, S.J., *et al.*, 2016. Online burst detection and location of water distribution systems and its practical applications. *Journal of Water Resources Planning and Management*, 142 (1), 04015033.

Li, R., *et al.*, 2015. A review of methods for burst/leakage detection and location in water distribution systems. *Water Science & Technology: Water Supply*, 15 (3), 429–441.

Liggett, J.A. and Chen, L.C., 1996. Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering*, 122 (8), 288–289.

Liu, S., *et al.*, 2014. Contamination event detection using multiple types of conventional water quality sensors in source water. *Environmental Science-Processes & Impacts*, 16 (8), 2028–2038.

Liu, S.M., Smith, K. and Che, H., 2015. A multivariate based event detection method and performance comparison with two baseline methods. *Water Research*, 80, 109–118.

Loureiro, D., *et al.*, 2016. Water distribution systems flow monitoring and anomalous event detection: A practical approach. *Urban Water Journal*, 13 (3), 242–252.

Mckenna, S.A., Wilson, M. and Klise, K.A., 2008. Detecting changes in water quality data. *Journal American Water Works Association*, 100 (1), 74–85.

Meseguer, J., *et al.*, 2014. A decision support system for on-line leakage localization. *Environmental Modelling & Software*, 60, 331–345.

Metz, C.E., 1978. Basic principles of ROC analysis. *Seminars In Nuclear Medicine*, 8 (4), 283–298.

Misiunas, D., *et al.*, 2005. Burst detection and location in water distribution networks. *Water Science & Technology: Water Supply*, 5(3), 3–4.

Mounce, S.R. and Boxall, J.B., 2010. Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: A case study. *Water Science & Technology: Water Supply*, 10 (3), 437–444.

Mounce, S.R. and Machell, J., 2006. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal*, 3 (1), 21–31.

Mounce, S.R., *et al.*, 2002. A neural network approach to burst detection. *Water Science and Technology*, 45 (4–5), 237–246.

Mounce, S.R., *et al.*, 2003. Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. *Information Fusion*, 4 (3), 217–229.

Mounce, S.R., Boxall, J.B. and Machell, J., 2007. An artificial neural network/fuzzy logic system for dma flow meter data analysis providing burst identification and size estimation. *In*: B. Ulanicki, K. Vairavamoorthy, D. Butler, P.L.M. Bounds, & F.A. Memon,eds. *Water management challenges in global change*. London: Taylor and Francis, 313–320.

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

Mounce, S.R., Mounce, R.B. and Boxall, J.B., 2011. Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics*, 13 (4), 672–686.

Mounce, S.R., Mounce, R.B. and Boxall, J.B., 2012. Identifying sampling interval for event detection in water distribution networks. *Journal of Water Resources Planning and Management*, 138 (2), 187–191.

Mounce, S.R., *et al.*, 2014. Pattern matching and associative artificial neural networks for water distribution system time series data analysis. *Journal of Hydroinformatics*, 16 (3), 617–632.

Mpesha, W., Chaudhry, M.H. and Gassman, S.L., 2001. Leak detection in pipes by frequency response method. *Journal of Hydraulic Engineering*, 127 (2), 134–147.

Mutikanga, H.E., Sharma, S.K. and Vairavamoorthy, K., 2013. Methods and tools for managing losses in water distribution systems. *Journal of Water Resources Planning and Management*, 139 (2), 166–174.

Oliker, N. and Ostfeld, A., 2014. A coupled classification - Evolutionary optimization model for contamination event detection in water distribution systems. *Water Research*, 51, 234-245. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24268294

Palau, C.V., Arregui, F.J. and Carlos, M., 2012. Burst detection in water networks using principal component analysis. *Journal of Water Resources Planning and Management*, 138 (1), 47–54.

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

Puust, R., *et al.*, 2010. A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7 (1), 25–45.

Romano, M., Kapelan, Z. and Savic, D., 2010. Real-time leak detection in water distribution systems. *In: 12th Water Distribution Systems Analysis Conference*, 12–15 September 2010 Tucson, USA.

Romano, M., Kapelan, Z. and Savic, D.A., 2012. Testing of the system for detection of pipe bursts and other events in a UK water distribution system. *In: 14th Water Distribution Systems Analysis Conference*, 24–27 September 2012 Adelaide, Australia.

Romano, M., Kapelan, Z. and Savic, D.A., 2013. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics*, 15 (3), 634–651.

Romano, M., Kapelan, Z. and Savic, D.A., 2014a. Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management*, 140 (4), 457–467.

Romano, M., Kapelan, Z. and Savic, D.A., 2014b. Evolutionary algorithm and expectation maximization strategies for improved detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management*, 140 (5), 572–584.

Romano, M., *et al.*, 2014c. Near real-time detection of pipe bursts events in cascading district metered areas. *In: 11th International Conference on Hydroinformatics*, New York, USA.

Samson, S., Reneke, J.A. and Wiecek, M.M., 2009. A review of different perspectives on uncertainty and risk and an alternative modeling paradigm. *Reliability Engineering & System Safety*, 94 (2), 558–567.

Sanz, G., *et al.*, 2016. Leak detection and localization through demand components calibration. *Journal of Water Resources Planning and Management*, 142 (2), 1097–1098.

Srirangarajan, S., *et al.*, 2013. Wavelet-based burst event detection and localization in water distribution systems. *Journal of Signal Processing Systems for Signal Image and Video Technology*, 72 (1), 1–16.

Tan, P.N., Steinbach, M. and Kumar, V., 2005. *Introduction to data mining*. New Jersey: Addison-Wesley.

Tao, T., *et al.*, 2014. Burst detection using an artificial immune network in water-distribution systems. *Journal of Water Resources Planning and Management*, 140 (10), 04014027.

Walker, W.E., *et al.*, 2003. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4 (1), 5–17.

Wu, Z.Y., Sage, P. and Turtle, D., 2010. Pressure-dependent leak detection model and its application to a district water system. *Journal of Water Resources Planning and Management*, 136 (1), 116–128.

Wu, Y., *et al.*, 2016. Burst detection in district metering areas using a data driven clustering algorithm. *Water Research*, 100, 28–37. Available from: http://www.sciencedirect.com/science/article/pii/S0043135416303347

Ye, G.L. and Fenner, R.A., 2011. Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *Journal of Pipeline Systems Engineering and Practice*, 2 (1), 14–22.

Ye, G. and Fenner, R.A., 2014a. Study of burst alarming and data sampling frequency in water distribution networks. *Journal of Water Resources Planning and Management*, 140 (6), 06014001.

Ye, G. and Fenner, R.A., 2014b. Weighted least squares with expectation-maximization algorithm for burst detection in U.K. water distribution systems. *Journal of Water Resources Planning and Management*, 140 (4), 417–424.