Monitoring Support for Water Distribution Systems based on Pressure Sensor Data



Caspar V. C. Geelen¹ • Doekle R. Yntema² • Jaap Molenaar³ • Karel J. Keesman^{1,2,3}

Received: 12 November 2018 / Accepted: 19 March 2019 / Published online: 6 July 2019 © The Author(s) 2019

Abstract

The increasing age and deterioration of drinking water mains is causing an increasing frequency of pipe bursts. Not only are pipe repairs costly, bursts might also lead to contamination of the Dutch non-chlorinated drinking water, as well as damage to other above- and underground infrastructure. Detection and localization of pipe bursts have long been priorities for water distribution companies. Here we present a method for proactive leakage control, referred to as Monitoring Support. Contrary to most leak prevention methods, our method is based on real-time pressure sensor measurements and focuses on detection of recurring pressure anomalies, which are assumed to be indicative of misuse or malfunctioning of the water distribution network. The method visualizes and warns for both recurring and one-time anomalous events and offers monitoring experts an unsupervised decision support tool that requires no training data or manual labeling. Additionally, our method supports any time series data source and can be applied to other types of distribution networks, such as those for gas, electricity and oil. The performance of our method, including both instance-based and featurebased clustering, was validated on two pressure sensor data sets. Results indicate that featurebased clustering is the best method for detection of recurring pressure anomalies, with accuracy F1-scores of 92% and 94% for a 2013 and 2017 data set, respectively.

Keywords Water distribution network (WDS) · Proactive leakage control · Unsupervised learning · Real-time learning · Early warning system

Karel J. Keesman karel.keesman@wur.nl

¹ Biobased Chemistry & Technology, Wageningen University, PO Box 17, 6700 AA Wageningen, The Netherlands

² European Center of Excellence for Sustainable Water Technology, Wetsus, Oostergoweg 9, P.O. Box 1113, 8900 CC Leeuwarden, The Netherlands

³ Biometris, Wageningen University, PO Box 16, 6700 AA Wageningen, The Netherlands

1 Introduction

The Netherlands has an excellent drinking water distribution system (WDS), with water losses of only 6%, compared to 25% and 16% for the US and UK, respectively (Rosario-Ortiz et al. 2016). The relatively good state of the Dutch drinking water infrastructure is in part caused by the replacement of at least half of the distribution network since 1970, resulting in an average pipe age of 33 to 37 years, compared to an estimated 75 to 80 years in the UK. Although the pipes are relatively new, the actual state of the water mains is largely unknown. Pipe bursts regularly occur, causing damage to other above- and underground infrastructure as well as requiring costly repairs. The Dutch drinking water is not chlorinated, which means that contamination as a consequence of bursts will not be neutralized by chlorine, therefore introducing more risk to consumers. In order to ensure proper functioning, water companies need to assess the probability of failure and apply leakage control.

Currently, the probability of pipe failure is estimated based on pipe properties, historical (failure) data and external conditions, with emphasis on reactive leakage control in the form of leak detection and localization (Mounce et al. 2003; Puust et al. 2010; Bakker et al. 2014; Gelazanskas and Gamage 2014; Okeya et al. 2015; Wu et al. 2016). However, to deal with the unknown state and continuous degradation of pipes, a proactive strategy, with a focused on leak prevention, is required. The objective of this study is therefore to present and evaluate a method for proactive leakage control.

Although various leak detection methods have been developed and tested, leak prevention methods have only recently been published (Wang et al. 2012; Xu et al. 2013; Kabir et al. 2015; Leu and Bui 2016; Kakoudakis et al. 2017). Although powerful, these methods frequently rely on supervised machine learning, requiring extensive data on pipe properties and external conditions. However, these methods often do not incorporate available real-time pressure and flow sensor data. Moreover, internal pipe conditions and grid management can also play a role in asset failure. In addition to extensive data sets, for the training of supervised models, classification labels are also required. Lastly, since these methods mostly use historical data, real-time implementation was not considered.

Our method focuses on proactive leakage control and offers an early warning and decision support system for proactive management of the WDS, which helps to prevent future bursts and malfunctioning. Contrary to the previously mentioned leak prevention studies, our method is based on real-time sensor data only, detecting recurring pressure anomalies which are indicative of misuse or malfunctioning within the WDS. Additionally, our method provides monitoring experts with an unsupervised decision support tool that requires no training data or manual labeling. Unsupervised learning is particularly suited for recurring pattern detection due to its robustness regarding detection of novel recurring patterns (Kotsiantis and Pintelas 2004). Clustering of anomalies allows detection of clusters containing a common recurring pattern. In this paper, both instance-based and feature-based clustering is applied to two pressure data sets from the Dutch drinking water company Vitens. Lastly, our method supports any time series data and can be applied to other distribution networks, such as those for gas, electricity or oil.

2 Materials & Methods

The detection of anomalous and recurring pressure patterns is divided into three steps: detection of anomalous events (Fig. 1a), clustering of events (Fig. 1b) and visualization of recurrence history (Fig. 1c).



Fig. 1 Flowchart of Monitoring Support. a Measured time series subjected to anomaly detection (six anomalous events). Whenever a new anomalous event is detected, windows of a preset number of preceding events are created (solid, dashed and dotted windows with four events per window in this example). These windows are then subjected to **b** moving window clustering, resulting in two clusters (star and circle) and outliers (black cross). **c** Clustering results are then summarized in fingerprint graphs, stacked area plots with time of event detection on the horizontal axis and frequency of pattern occurrence per cluster on the vertical axis

2.1 Data Sets

Access to actual and historical pressure sensor data was provided by Vitens, a Dutch drinking water company. A known case of recurring anomalous pressure patterns followed by a pipe burst was investigated from 1/6/2012 to 1/6/2013, hereafter referred to as the 2013 data set. In addition, a recent data set from another pressure sensor is used, with measurements from 18/5/2017 to 17/11/2017, hereafter referred to as the 2017 data set. Both pressure sensors were situated close to water reservoirs.

As a preprocessing step, erratic measurements were removed. Resampling and linear interpolation in time were used to obtain a constant sampling interval of one second.

2.2 Event Detection

Anomalous events were detected using a moving window range statistic, defined as the difference between maximum and minimum values of every ten-seconds moving window, divided by the window size of ten seconds. A ten-seconds window range statistic was used instead of the derivative, so as to avoid problems associated with noise present in the pressure measurements. Measurements with a range statistic of more than two kPa/s were flagged as anomalous (Fig. 1a), since rapid pressure changes of this magnitude are most often caused by events that are relevant for the purposes of this study. Although quite simple, the range statistic and absolute range threshold were found to be able to detect all relevant anomalous events. Since anomaly detection is an important and complicated process, a more extensive definition of anomaly detection will most likely improve performance (Branisavljević et al. 2011; Mounce et al. 2014; Scozzari and Brozzo 2017). However, for illustration of our method on the aforementioned data sets, the current metric is sufficient and suitable.

The anomalies were combined into events, where anomalous measurements within a 15 min duration were considered to be part of one event (Fig. 1a). Next, each event was extended with two minutes of preceding and two minutes of succeeding measurements to ensure the entire anomaly and context were captured as a single event.

2.3 Event Clustering

Recurrence of anomalous pressure patterns was defined as the repetition of similar anomalous events. Events were clustered in order to detect which events are similar and probably have the same cause. Clustering is an unsupervised method for grouping of similar events based on the distances between events. For this, events were represented by vectors, after which the distance between these vectors can be calculated. Events with a low distance between them are deemed similar and were included in the same cluster. Each cluster corresponds to a specific recurring and anomalous pattern (Fig. 1b). The vectors assigned to each event were based either on event measurements (instance-based) or on each event's characteristic features (feature-based) (Fulcher and Jones 2014).

In this study, clustering was performed using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al. 2017), which clusters events based on their density within a vector space. Unlike similar clustering methods, such as DBSCAN (Ester et al. 1996) and Mean Shift (Ray and Benammar 2002), HDBSCAN uses a hierarchical minimum density threshold and is better in detecting varying cluster shapes. HDBSCAN also allows clustering with a precomputed distance matrix and has the capacity to distinguish core samples from outliers.

Since the presented method is intended for real-time application, clustering needs to be performed anew when novel events are detected. Clustering was performed over the most recent 150 events using a moving window of these events whenever a new anomalous event was detected (Fig. 1b). This moving window approach ensures real-time applicability and detection of distinct clusters for the different recurring patterns present in the investigated data. The window size can be adjusted if requested. However, larger window sizes potentially result in merging of clusters due to a higher overall density of events, making the distinguishing of local denser areas more difficult. Smaller window sizes might result in failure to detect recurring patterns with a low frequency of occurrence.

2.4 Distances for Instance-Based Clustering

In order to calculate the distance between two event vectors of different lengths, the vectors are clipped to equal lengths. Clipping was done based on the maximum cross-correlation between both events (Fig. 2b). For every pair of event time series, the lag related to the maximum cross-correlation was removed (Fig. 2a), followed by clipping of the non-overlapping tails of both events (Fig. 2a) to obtain events of equal length (Fig. 2c).

Optionally, Dynamic Time Warping (DTW) can then be applied in order to correct for temporal drift, which increases the accuracy of the succeeding distance calculations (Fig. 2d) (Aghabozorgi et al. 2015). In this study, DTW was limited to warping of up to 5% of the total event duration in both directions. After clipping and DTW, the Euclidean distance between events was calculated and corrected by dividing by the length of the events before being subjected to clustering.

2.5 Distances for Feature-Based Clustering

For each event, 43 features were calculated (Appendix 1, Table 1). In each clustering window of 150 events, the features of these events were scaled by median subtraction followed by interquartile range division, ensuring that scaling was robust for outliers.

The features were chosen so as to be robust for distinguishing between a limited number of recurring patterns. After scaling, the distances between each event pair's feature vectors were calculated and the resulting distance matrix was subjected to the clustering method.

2.6 Fingerprint Graphs

Fingerprint graphs (Fig. 3) present an effective overview of the periods of recurrence for different type of patterns and their respective frequency of occurrence. When a new anomalous event is detected, the clustering results of the corresponding 150-event window is added to the fingerprint graph as a vertical white slice. Each colored area depicts a recurring pattern, where each pattern's height depicts its frequency of occurrence within the 150-event window and its length corresponds to the duration of the pattern recurrence (Fig. 1c).

2.7 Validation Report

The validation report depicts the precision (fraction true positives among detected positives), recall (fraction of true positives among actual positives) and F_1 -score (2 * *precision* * *recall*/(*precision* + *recall*)) for each true recurring pattern present in the



Fig. 2 a Time series clipping based on maximum cross correlation of standardized events. b Cross correlation between both events. c Events A and B after clipping and c) before and d after Dynamic Time Warping

manually labeled validation data (van Rijsbergen 1979). In order to calculate these scores, cluster ID numbers were mapped to the validation labels. Clusters mapping to the same pattern were deemed a single cluster for the sake of accuracy scores calculation only.



Fig. 3 Example fingerprint graph, showing the recurring anomalous pressure patterns as different areas in the stacked area graph. Each legend entry number matches a separate cluster. The bottom, black area with legend entry -1 represents outlier events, which are deemed non-recurring

3 Results

The method was applied to pressure data of the WDS of Vitens. In order to validate the method, a known case of pressure pattern recursion leading to a pipe burst was investigated, as well as a more recent data set from 2017. The data set from 2013 contains a rapid crack propagation event at 2013/03/12 18:03 (Fig. 4). The pipe in question was already under strain due to angular displacement and sub-zero temperatures. However, afterwards it was concluded that the burst probably occurred due to pressure oscillations caused by the interaction of two upstream pumps connected in parallel. Repeated activation and deactivation of these pumps led to these recurring oscillations, which had been occurring for over two months before the coincidence with sub-zero temperatures and additional pipe stress caused by traffic led to a burst.

To prevent future malfunctioning and to obtain more insight into the behavior inside the pipes, we developed a method functioning as a real-time decision support and early warning system for recurring unwanted pressure patterns. By timely detection of recurring anomalous pressure patterns, the 2013 pump malfunction could have been identified earlier and the pipe burst might have been prevented. As a proof of concept, our method has been applied to the 2013 (Fig. 5 and 6) and 2017 data sets (Fig. II-1, Fig. II-2, Appendix 2) using instance-based clustering with and without DTW and feature-based clustering. In order to assess the real-time performance of the method, it was applied to the 2013 and 2017 data sets with moving windows, as a stand-in for real-time application.

When a novel anomalous event was detected in the pressure time series data of a sensor, the most recent 150 events time window was again clustered. Events that belong to the same cluster were assumed to be part of the same recurring anomalous pressure pattern. Based on the manually labeled validation data (Fig. 5d), there are five main types of recurring patterns present in the 2013 data set (labeled as fast oscillation, oscillation, slope, spike and valley) (see Fig. 6).

As mentioned before, the 2013 burst (Fig. 4) probably happened because of recurring pressure oscillations (Fig. 6, Oscillation), which in turn were caused by erroneous behavior of two pumps upstream of the sensor. Without having this prior knowledge, our method detects these oscillations and so would have provided an early warning of the problem months in advance of the eventual burst.

Besides the oscillations, four other recurring patterns are detected. The fast oscillation events most probably occurred as a consequence of rapid pump activation and deactivation. The slope pattern (Fig. 6: Slope) consists of rapid pressure increases due to increased pumping activity. The slope events occur mostly in the early morning, where rapid pump activations cause the pressure to rise to a higher pressure than is necessary, before gradually decreasing again. The



Fig. 4 Pressure sensor data from the 2013 data set containing the pipe burst at 12-03-2013 18:03



Fig. 5 Results of the 2013 data set for instance-based clustering without DTW (A), with DTW (B), feature-based clustering (C) and validation using manual labeling (D). As can be seen from the validation labeling (D), five different recurring patterns are found in the data set, referred to as fast oscillation, oscillation, slope, spike and valley. The legend lists the cluster ID number and which pattern matches best with that clusters ID, as derived from the manually labeled validation data. Clusters numbers mapping to the same recurring pattern, were deemed a single cluster for the sake of accuracy scores calculation only

spike pattern (Fig. 6: Spike) consists of pressure transients, caused by rapid pump, valve or water consumption changes. Pressure transients may cause (gradual) degradation and deformation of pipes, connections or valves (National Research Council 2006). Lastly, the valley patterns (Fig. 6: Valley) consists of short pressure drops where for a short period of time water diversion or increased water consumption causes temporary but considerable pressure drops.

4 Discussion and Conclusions

The fast oscillations events show a large variation between them (Fig. 6). Consequently, both instance-based methods and the feature-based method show a lower recall for fast oscillations compared to other patterns (Fig. 5, validation reports). Besides a lower recall, our method often detects multiple clusters matching the fast oscillation recurring pattern, due to the large variation between various fast oscillation events. (Fig. 5b fingerprint graph clusters 1,2 and 8 all correspond to the fast oscillation pattern. The same is true for Fig. 5c clusters 3 and 6.)

4.1 Method Comparison

Most spike events closely resemble half a period of an oscillation event, resulting in a small instance-based clustering distance between these events, especially after event clipping. This phenomenon is reflected in the low accuracy of spike detection for instance-based clustering without and with DTW (Fig. 5ab: F_1 -scores of 0.62 and 0.00, respectively), as opposed to the high accuracy using feature-based clustering (Fig. 5c: F_1 -score of 0.93). To some extent, the same occurs for slope events resembling parts of valley events (Fig 5abc: slope F_1 -scores of 0.61 and



Fig. 6 Examples of the recurring patterns present in the 2013 data set

0.62 for instance-based with and without DTW, respectively, versus 0.92 for feature-based clustering). Because of this low distance between parts of both patterns, instance-based clustering is less suitable for distinguishing oscillation and spike events compared to feature-based clustering, which does not rely on the distances between events as calculated for instance-based clustering.

Like fast oscillations, there is a large variation between the valley events. Additionally, only 17 out of the 334 events in the 2013 data set represent valleys. As a result, instance-based clustering is unable to detect the valley recurring pattern (Fig. 5ab: F_1 -scores 0.00 and 0.00 for with DTW and without DTW instance-based clustering) and feature-based clustering shows a low recall of 0.69 for valley detection (Fig. 5c).

Since an unsupervised approach was taken in this study, novel patterns that did not occur in the past could still be detected successfully, such as the oscillation pattern seen in the 2017 data

set (Fig. II-1, Fig. II-2, Appendix 2). Not only do new patterns occur as time progresses, the types of patterns detected also differ widely between sensors, as can be seen when comparing the 2013 and 2017 data set results. Consequently, an unsupervised method is considered the most suitable approach for detecting pattern recurrence in sensor data.

Feature-based clustering requires a suitable selection of features capable of distinguishing recurring patterns. As a consequence of the unsupervised approach, it is not possible to automatically choose a set of features most suited for grouping pressure anomalies or to weigh features based on suitability. Therefore, additional care is required for initial feature selection. However, even though the 2013 and 2017 data sets differ widely in recurring patterns present, the currently selected features show high accuracies detecting and distinguishing between recurring patterns (Fig. 5, Fig II-1, Appendix 2). Feature-based clustering also outperforms instance-based clustering, as can be seen from the F1-scores of 0.93 and 0.94 for feature-based clustering of the 2013 and 2017 data sets, compared to 0.49/0.82 and 0.52/0.80 for the no DTW/DTW instance-based clustering of 2013 and 2017 data sets, respectively (Fig. 5, Fig II-1, Appendix 2). This indicates that the currently chosen set of features are robust for clustering 150 event windows (Fig. I-1, Appendix 1).

4.2 Method Performance

Our method fills the gap for real-time sensor-based and proactive leakage control methods. Besides recurrence detection, the method offers an easy framework for monitoring pressure measurements. Our method finds all anomalous pressure events and detects which contain a recurring pattern. The method can isolate, visualize and summarize both recurring and onetime events and so helps to determine the cause and potential consequences of the aberrant pressure events. Combined with an unsupervised approach, our method represents a powerful tool that alleviates the grid monitoring workload of monitoring experts.

Overall, our method shows promising results regarding recurrence detection and visualization. Although only the performance with time series data from pressure sensors was investigated, flow data or data from other distribution systems can also be used. By choosing a suitable anomaly detection method, our method can be applied to any time series data where recurrence of unwanted or artificial patterns might occur.

Our application to real data shows that feature-based clustering is the preferred method for detecting recurring pressure anomalies. This implies that selection of these features is a crucial ingredient of this approach. Implementation of our method and/or testing more data sets will allow reevaluation of chosen features over time, if required. However, since an average accuracy F₁-score of 93.5% was achieved with the current feature-based unsupervised method, current features show robustness for clustering of 150 event windows.

Acknowledgements This work was performed in the cooperation framework of Wetsus, European Centre of Excellence for Sustainable Water Technology (www.wetsus.nl). Wetsus is co-funded by the Dutch Ministry of Economic Affairs and Ministry of Infrastructure and Environment, the European Union Regional Development Fund, the Province of Fryslan and the Northern Netherlands Provinces. The authors would like to thank the participants of the research theme "Smart Water Grids" for fruitful discussions and financial support, especially the Dutch drinking water company Vitens for providing the data sets that made this research possible.

Compliance with Ethical Standards

Conflict of Interest None.

Appendix 1

Table 1	Features us	sed for feat	ure-based c	lustering (Schreiber a	and Schmitz	1997; Fulche	r and Jones	2014; C	.'hrist
et al. 20)18)									

Features	Description						
max	Maximum of the event measurements						
min	Minimum of the event measurements						
std	Standard deviation of the event measurements						
skew	Unbiased skewness, normalized by N-1						
kurt	Unbiased kurtosis using Fisher's definition of kurtosis, normalized by N-1.						
diff mean	Mean of the first order derivative of the event measurements						
diff_skew	Unbiased skewness, normalized by N-1, of the first order derivative of the event measurements						
diff std	Standard deviation of the first order derivative of the event measurements						
lin appr slope	Signe estimate n_{y} from a linear least squares fit ($y = n_{x} + n_{y}$)						
lin_appr_stope	Intercent estimate p_2 from a linear least squares fit $(y = p_1 x + p_2)$						
cub appr_mercept	Polynomial coefficient estimate p_1 from a cubic least squares fit through the sorted data						
cub_appi_xi	$(y - p_1 x^3 + p_2 x^2 + p_3 x + p_3)$						
cub appr x?	$(y - p_1x + p_2x + p_3x + p_4)$ Polynomial coefficient estimate n from a cubic least squares fit through the sorted data						
cub_appi_x2	$(y = p x^2 + p x + p)$						
out oppr y?	$(y = p_1x^2 + p_2x^2 + p_3x + p_4)$ Baluramial apofficiant assimption from a subia losst squares fit through the sorted data						
cub_appi_x5	$(y = p_1 x^3 + p_2 x^2 + p_3 x + p_4)$						
cub_appr_intercept	Intercept estimate p ₄ from a cubic least squares fit through the sorted data						
	$(y = p_1 x^3 + p_2 x^2 + p_3 x + p_4)$						
slope_appr	Difference between initial and final event value divided by the number of seconds between						
std_0.5	Fraction of event measurements larger than 0.5 times the standard deviation						
std_1.0	Fraction of event measurements larger than 1 times the standard deviation						
std_1.5	Fraction of event measurements larger than 1.5 times the standard deviation						
domfreq_0.000	Dominant Power Spectral Density frequency						
domfreq_0.003	Dominant Power Spectral Density frequency above 1/300 s ⁻¹						
domfreq_0.008	Dominant Power Spectral Density frequency above 1/120 s ⁻¹						
domfreq_0.017	Dominant Power Spectral Density frequency above 1/60 s ⁻¹						
psdbin_1	Power Spectral Density average between frequencies 1/600 and 1/103 s ⁻¹						
psdbin_2	Power Spectral Density average between frequencies 1/103 and 1/56 s ⁻¹						
psdbin 3	Power Spectral Density average between frequencies 1/57 and 1/39 s ⁻¹						
psdbin 4	Power Spectral Density average between frequencies 1/39 and 1/30 s ⁻¹						
psdbin 5	Power Spectral Density average between frequencies 1/30 and 1/24 s ⁻¹						
acorr 5	Autocorrelation with a lag of 5 s						
acorr 10	Autocorrelation with a lag of 10 s						
acorr 15	Autocorrelation with a lag of 15 s						
acorr 30	Autocorrelation with a lag of 30 s						
acorr 60	Autocorrelation with a lag of 60 s						
domacorr 0.50	Fraction of autocorrelation function above than 50%						
domacorr 0.65	Fraction of autocorrelation function above than 65%						
domacorr 0.80	Fraction of autocorrelation function above than 80%						
domacorr 0.90	Fraction of autocorrelation function above than 90%						
domacorr 0.95	Fraction of autocorrelation function above than 95%						
c3 lag 5	Time series non-linearity measure using a lag operator of 5 s						
c3 lag 10	Time series non-linearity measure using a lag operator of 10 s						
$c_3 \log 15$	Time series non-linearity measure using a lag operator of 15 s						
tras lag 5	Time reversal asymmetry statistic using a lag operator of 5 s						
tras lag 10	Time reversal asymmetry statistic using a lag operator of 10 s						
tras lag 15	Time reversal asymmetry statistic using a lag operator of 15 s						



Fig. I-1 Feature performance evaluation for the 2013 and 2017 data set. After feature calculation for all events within a single data set, each feature was standardized. For each pattern present in the manually labeled validation, a box plot of feature values was made per pattern. Features showing well-separated pattern-specific box plots with low variation within the events of a single pattern are most suitable for separating the patterns present in the investigated data set. As can be derived from the box plots and validation reports of the 2013 and 2017 data sets (Fig. 5, Fig. II-1, Appendix 2), current features are deemed suitable and robust for the intended goal of detecting recurrence of anomalous patterns



Appendix 2

Fig. II-1 Results of the 2017 data set for instance-based clustering without DTW (A), with DTW (B), featurebased clustering (C) and validation using manual labeling (D). As can be seen from the validation labeling (D), three different recurring patterns are found in the data set, referred to as rise, fall and oscillation. The legend lists the cluster ID number and which pattern matches best with that clusters ID, as derived from the manually labeled validation data. Clusters numbers mapping to the same recurring pattern, were deemed a single cluster for the sake of accuracy scores calculation only



Fig. II-2 Examples of the recurring patterns present in the 2017 data set

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T (2015) Time-series clustering A decade review. Inf Syst 53:16–38. https://doi.org/10.1016/j.is.2015.04.007
- Bakker M, Vreeburg JHG, Van De Roer M, Rietveld LC (2014) Heuristic burst detection method using flow and pressure measurements. J Hydroinf 16:1194. https://doi.org/10.2166/hydro.2014.120
- Branisavljević N, Kapelan Z, Prodanović D (2011) Improved real-time data anomaly detection using context classification. J Hydroinf 13:307. https://doi.org/10.2166/hydro.2011.042
- Christ M, Braun N, Neuffer J, Kempa-Liehr AW (2018) Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). Neurocomputing 307:72–77. https://doi.org/10.1016/j. neucom.2018.03.067
- Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second international conference on knowledge discovery and data mining. AAAI Press, Portland, pp 226–231
- Fulcher BD, Jones NS (2014) Highly comparative feature-based time-series classification. IEEE Trans Knowl Data Eng 26:3026–3037. https://doi.org/10.1109/TKDE.2014.2316504
- Gelazanskas L, Gamage KAA (2014) Demand side management in smart grid: A review and proposals for future direction. Sustain Cities Soc 11:22–30. https://doi.org/10.1016/j.scs.2013.11.001
- Kabir G, Tesfamariam S, Sadiq R (2015) Predicting water main failures using Bayesian model averaging and survival modelling approach. Reliab Eng Syst Saf 142. https://doi.org/10.1016/j.ress.2015.06.011
- Kakoudakis K, Behzadian K, Farmani R (2017) Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K -means clustering. Urban Water J 9006:1–6. https://doi.org/10.1080/1573062X.2016.1253755
- Kotsiantis SB, Pintelas PE (2004) Recent Advances in Clustering: A Brief Survey. Methods 1:73-81
- Leu SS, Bui QN (2016) Leak Prediction Model for Water Distribution Networks Created Using a Bayesian Network Learning Approach. Water Resour Manag 30:2719–2733. https://doi.org/10.1007/s11269-016-1316-8
- McInnes L, Healy J, Astels S (2017) hdbscan: Hierarchical density based clustering. J Open Source Softw 2:205. https://doi.org/10.21105/joss.00205
- Mounce SR, Khan A, Wood AS et al (2003) Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. Inf Fusion 4:217–229. https://doi.org/10.1016/S1566-2535(03)00034-4
- Mounce SR, Mounce RB, Jackson T et al (2014) Pattern matching and associative artificial neural networks for water distribution system time series data analysis. J Hydroinf 16:617. https://doi.org/10.2166 /hydro.2013.057
- National Research Council (2006) Drinking water distribution systems: Assessing and reducing risks. National Academies Press, Washington, DC
- Okeya I, Hutton C, Kapelan Z (2015) Locating pipe bursts in a District Metered Area via online hydraulic modelling. Proceedia Eng 119:101–110. https://doi.org/10.1016/j.proeng.2015.08.859
- Puust R, Kapelan Z, Savic DA, Koppel T (2010) A review of methods for leakage management in pipe networks. Urban Water J 7:25–45. https://doi.org/10.1080/15730621003610878
- Ray C, Benammar ASO (2002) Mean shift: A robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24:603–619
- Rosario-Ortiz F, Rose J, Speight V et al (2016) How do you like your tap water? Science 80(351):912–914. https://doi.org/10.1126/science.aaf0953
- Schreiber T, Schmitz A (1997) Discrimination power of measures for nonlinearity in a time series. Phys Rev E -Stat Physics, Plasmas, Fluids, Relat Interdiscip Top 55:5443–5447. https://doi.org/10.1103 /PhysRevE.55.5443
- Scozzari A, Brozzo G (2017) Making use of continuous measurements for change detection purposes: An application to water distribution networks. I2MTC 2017–2017 IEEE Int Instrum Meas Technol Conf Proc. https://doi.org/10.1109/I2MTC.2017.7969738
- van Rijsbergen CJ (1979) {I}nformation {R}etrieval, 2nd edn. Butterworths, London

- Wang L, Zhang H, Niu Z (2012) Leakage prediction model based on RBF neural network. Adv Intell Soft Comput 114:451–458. https://doi.org/10.1007/978-3-642-03718-4_56
- Wu Y, Liu S, Wu X et al (2016) Burst detection in district metering areas using a data driven clustering algorithm. Water Res 100:28–37. https://doi.org/10.1016/j.watres.2016.05.016
- Xu Q, Chen Q, Ma J, Blanckaert K (2013) Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network. J Hydro-Environment Res 7:134–140. https://doi.org/10.1016/j.jher.2013.03.003

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.