Asset-Risk–Weighted Spectral Partitioning to Improve the Resilience of Water Distribution Networks

Daniel Kozelj†
Faculty of Civil and Geodetic Engineering
University of Ljubljana
Ljubljana, Slovenia
daniel.kozelj@fgg.uni-lj.si

Abstract

District Metered Areas (DMAs) are central to leakage control, but partitions based only on topology or demand ignore local failure risks. We propose an asset risk weighted spectral partitioning XGBoost-derived pipe failure probabilities (PFp) are integrated into a generalized normalized cut (GNC) framework. The pipe level PFp are length-weighted and aggregated into nodes to form vertex weights, that guide the spectral solver to balance clusters by state risk. Using the Ljubljana-Šentvid case, we compare PFp-weighted GNC with a demand-balanced baseline across representative edge weighting cases (unweighted u, diameter d, length l, minimum-cost C_{min}, and topological case w_1), identical clustering (squared-euclidean). The condition dependent balancing favors fewer, more compact DMAs, and concentrates pipes networks poor-condition pipes on smaller DMAs. The method is reproducible and data-based and embeds PFp directly into the partitioning to provide hydraulically coherent, operationally tractable, and risk-oriented DMAs for aging WDNs.

Keywords

Water loss, District Metered Areas, Spectral Graph Partitioning, Pipe Failure Probability, XGBoost

1 Introduction

Public water supplies are energy-intensive to extract, transport, treat and deliver. As a recent comprehensive US analysis showed, energy intensity has increased from 2001 to 2020 - by 12% for large (0.49 kWh/m³), 8% for medium (0.53 kWh/m³), and 28% for small (0.67 kWh/m³) utilities. These trends confirm that water supply is becoming increasingly energy intensive, underlining the need for sustainable, integrated water–energy management [1]. Although utilities cannot control final energy, leakages in water distribution systems are an increasingly critical problem [2], exacerbated by post-World War II pipes that are in poor condition [3]. In Slovenia, annual water abstraction for water supply amounts to about 168 million m³, with reported losses of about 50 million m³; leakages vary widely between utilities (about 20–70%) [4]. Water leaks are recognized as one of the driving forces for higher costs, as the water is treated and pumped into the distribution network. The loss of this valuable resource is not only an economic concern but is also increasingly jeopardizing water security as droughts have become more severe in recent years [5]. The main cause of inefficiency in water distribution is aging infrastructure, which is a key challenge for utilities, municipalities and customers alike.

Scientific and technical literature has established and confirmed that one of the most effective measures to identify water losses is the establishment of District Metered Areas (DMAs) – sub-areas in which inflow and/or outflow are measured simultaneously to equalize water volumes [6]. DMAs are realized through the installation of flow meters and valves and also enable the calculation of the water balance and its components for the respective network. As important as DMAs are in combating water loss, they require significant investment. So while DMAs reduce inefficiencies in the WDN, they are themselves prone to inefficiencies in design and investment costs, as well as in the insights they can provide through their mass balancing.

As one of the most common approaches to control and reduce real water losses for water utilities, DMA design reduces unearned water and clarifies hydraulic conditions in the WDN. Partitioning the WDN is complex and prone to weighting case selection to guide the partitioning method for efficiency and expediency [7], [8]. The method used in this study employs spectral partitioning algorithms, which are classified into three coding types, namely the Ratio Cut (RC), the Normalized Cut (NC), and the Generalized Normalized Cut (GNC). The latter, GNC, decouples the weighting of vertex (balance) and edge (cut) and thus enables balanced partitions and flexible weighting [9]. The GNC method can further control the network partition by balancing and controlling the network partition objectives such as water loss control and reduction [7], [9], [10], [11].

The probability of pipe bursts and their subsequent impact on water loss is a promising proposal to balance the DMA and improve the efficiency and effectiveness of leak detection, localization and remediation. Recent advances in machine learning (ML) have significantly improved pipe failure prediction and leak detection capabilities [12], [13], [14], [15], [16]. However, ML models for pipeline failures are still

*Article Title Footnote needs to be captured as Title Note
†Author Footnote to be captured as Author Note

Information Society 2025, 6–10 October 2025, Ljubljana, Slovenia

http://doi.org/DOI_RECEIVED_AFTER_REVIEW

constrained by limited data, heterogeneous degradation, utility-specific calibration, and weak data integration [17], [18].

Greater integration of various infrastructure data has shown improvements in quantifying the probability of pipeline failure and subsequent condition assessment for strategic planning and proactive asset management. Kozelj and Abert Fernández [19] have shown that ML models are very effective when they incorporate a multidimensional approach to quantitatively assess non-WSS features (i.e., neighboring infrastructure systems) in the prediction of pipeline failures, such as construction activity, operational loads from nearby transportation infrastructure, and environmental impacts from neighboring utilities. The multi-system interdependency of buried supply infrastructures can restructure the occurrence of water losses and operational prioritization.

As Slovenian utilities report high water losses (~20-70% in NRW), budgets are limited and assets are outdated, the limited investment in monitoring and analysis must provide the greatest benefit to the goal of combating water losses. Therefore, our study focuses on whether condition-based vertex weighting – using ML-based pipe break probabilities (PFp) – can make spectral partitioning and DMA design more feasible by prioritizing high-risk zones without compromising hydraulic performance or increasing implementation costs. Embedding PFp as vertex weights in the GNC formulation and comparison with on- demand GNC in the same case of Ljubljana [9].

2 Spectral Graph Partitioning – GNC method

A WDN strongly resembles the structure of a graph $G = (V, E)$, where V is a set of n vertices and E is a set of undirected edges between these n vertices. The constat nodes and pipes correspond to vertices and edges, respectively. In this study, spectral graph partitioning is used to partition graph G into subgraphs $G_1, G_2, \dots, G_p$ where $p \leq n$. In a subgraph $G_k = (V_k, E_k)$, where $k = 1, \dots, p$, all the edges connecting the vertices V_k are referred to as intracluster E_k, while edges connecting vertices from different subgraphs are referred to as intercluster edges B, which represent links between different subgraphs. A complete partition is therefore referred to as [9]:

$$\mathcal{P} := \{G_1, G_2, \dots, G_p\}. \tag{1}$$

To achieve a more balanced partitioning, we can use different objective functions for our partitioning problem. Our research uses the generalized normalized cut (GNC) since it balances the sum of vertex weights (w_v) within each cluster while minimizing the sum of intercluster edge weights ($w_{vv'}$). For a partition $\mathcal{P} := \{G_1, G_2, \dots, G_p\}$ the objective therefore is [9]:

$$\eta(\mathcal{P}) := \min_{V_1, V_2, \dots, V_p} \sum_{k=1}^{p} \frac{\mathrm{vol}(\partial(V_k))}{\mathrm{vol}(V_k)}, \tag{2}$$

where $\mathrm{vol}(\partial(V_k))$ is the sum of the weights of all intercluster edges in $\partial(V_k)$; and $\mathrm{vol}(V_k)$ is the sum of the weights of the vertices in V_k [9]:

$$\begin{aligned} \mathrm{vol}(\partial(V_k)) &= \sum_{vv' \in \partial(V_k)} w_{uv}, \\ \mathrm{vol}(V_k) &= \sum_{v \in V_k} w_v. \end{aligned} \tag{3}$$

The corresponding generalized eigenvalue problem is written as follows:

$$LU = WUL, \tag{4}$$

where L is the Laplacian matrix; U is the eigenvector matrix; and $W = D_V$ is the diagonal matrix of vertex weights:

$$D_{V_{ij}} := \begin{cases} w_{v_i}, & i = j, \\ 0, & \text{sicer} \end{cases} \tag{5}$$

with $i, j = 1, \dots, n$ [9]. By solving equation (4) we obtain the p smallest eigenvalues and their corresponding eigenvectors, which are then clustered by rows into p clusters using the k-means++ clustering algorithm, where the cosine or squared Euclidean distance is used to determine the cluster centroids [20]. The clustering algorithm assigns each node to a corresponding DMA. After partitioning the spectral graph, the characteristics of each established DMA can be extracted. The determined subgraphs $G_1, G_2, \dots, G_p$ are then interconnected by any combination of the edges from the intercluster set B to obtain the final graph.

The efficiency of connecting subgraphs is ensured by using spanning trees, which identify all topologically valid possibilities for connecting subgraphs from the partition $\mathcal{P}$ with the smallest possible number of edges [9]. A connected unweighted multigraph $H = (V_H, E_H)$ is constructed, where V_H are vertices representing one of the subgraphs G_p, and E_H are the intercluster edges in B. The spanning tree method over k-shortest weighted (hydraulic resistances) paths prioritizes the identified water mains, reduces the combinatorial space, and simultaneously preserves hydraulic sufficiency [9].

As described in the previous subsection, the spectral graph partitioning algorithms provide balanced partitions with the lowest cut values, i.e., the sums of the intercluster edge weights. The GNC partitioning method uses two different sets of weights, i.e., edge (cut) weights for the weighted Laplacian matrix L, and an independent set of vertex (balancing) weighting for the matrix D_v. The weight cases are as follows:

- Laplacian matrix L: unweighted (u), pipe diameter (d), pipe length (l), minimum edge costs (uC), and a topological weighting ($w1$) which includes additional topological characteristics (i.e., gate valves, bridges in the graph, and water mains).
- Diagonal matrix D_v : considers the pipe failure probabilities (PFp).

The selected weight cases are compared with the published results of Zevnik et al. [9]. The most important comparison is the introduction of pipe failure probabilities (PFp) as balancing weights to make spectral partitioning more suitable for its

primary purpose of efficiently detecting and reducing water losses.

2.1 Pipe failure probability (PFp) modeling

The PFp weights come from a recent study by Kozelj and Abert Fernández [19], which are obtained by training an ML algorithm with data from the Ljubljana WSS (Ljubljana, Slovenia). Three datasets were used for the study, namely 52,605 records of pipe sections from the national utility cadastre managed by GURS [21], a register of 2,281 documented pipe bursts from the utility for the period from 2010 to 2025 [22], and several cross-utility infrastructure networks that showed meaningful predictive power for pipe burst risk [21]. The ML model was built using the XGBoost algorithm and validated by stratified K-fold cross-validation, and achieved high performance (ROC AUC: 0.9102; recall: 0.7750; accuracy: 0.8750; F1 score: 0.2261; and LogLoss: 0.2500), confirming its reliability [19].

The XGBoost algorithm [23] was validated by stratified five-fold cross-validation[24], which yields out-of-fold PFp for each pipe, which were later grouped into five classes $[0-0.2]$, $(0.2-0.4]$, $(0.4-0.6]$, $(0.6-0.8]$, and $(0.8-1]$ (Figure 1). The metric of feature importance showed that the most influential predictors of the model were pipe material, installation year, and pipe diameter, but also influential gains from adjacent infrastructure systems, such as electricity grids, sewage systems, and roads.

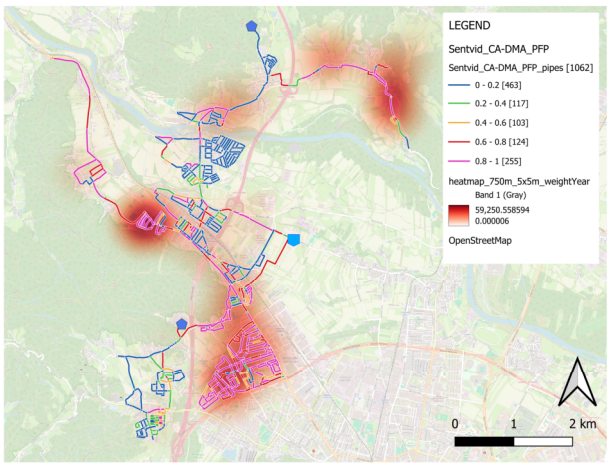

Figure 1: Pipe failure probabilities (PFp) and heatmap the spatial distribution of historical failures

In addition, to the spatial representation of the probabilistic results generated by the XGBoost algorithm (Figure 1), we can view the pipeline segments in a statistical analysis by looking at the distribution of pipe-specific characteristics, such as material and diameter, in the classes of failure probability. Figure 2 illustrates the probability of pipe failure as a function of pipe and age.

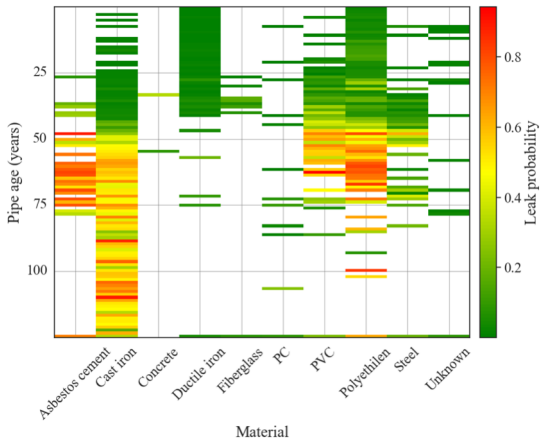

Figure 2: Heat map of pipe failure probabilities (PFp - leak) as a function of pipe material and age

The assignment of PFps for the GNC method was carried out by embedding the condition risk in the weighting term of GNC, whereby the PFp probabilities of the pipes $pf_e \in [0,1]$ were transferred to the vertex weights w_v via an incident pipe aggregation:

$$w_v = \frac{\text{Agg\_}e \in \delta(v)(pf_e \cdot l_e)}{\text{Agg\_}e \in \delta(v)(l_e)}, \tag{6}$$

where $\delta(v)$ is the set of pipes belonging to vertex (node) v, l_e is edge (pipe) length, and Agg is the (length-weighted) mean value. This concentrates the compensation probabilities for possible current or future water losses in areas with many or long high-risk pipes. The edge-weighted cases (u, d, l, C_{min}, and w_1) remain unchanged from the baseline [9].

3 Results and discussion of the GNC Spectral Graph Partitioning

Following the baseline study of Zevnik et al. [9] the computation k -means++ clustering was performed using squared Euclidean distance [20], although the cosine distance was also investigated, but yielded poorer results compared to the squared Euclidean distance, as noted in the previous study. Each partition $\mathcal{P}$ was subjected to: (i) an intrernal connectivity testing; (ii) efficient connection selection using spanning trees, and (iii) hydraulic screening using the generalized resilience index I_r [25] and tank-flow. The final alternatives are evaluated using the six-criteria model and the journalistic weights mentioned above [9]. The values of the selected criteria are normalized according to Liu and Han [26] considering a positive or negative influence. The final score of each partition is calculated as a weighted sum of the scores for a particular criterion. The higher the values of the final score, the better the alternatives of the DMA design.

The evaluation of the PFp-balanced GNC, i.e, vertex weights corresponding to the aggregated PFp risk of the node, w_v (length-weighted), was evaluated over the same five

representative cases of weighting: unweighted (u), pipe diameter (d), pipe length (l), minimum edge costs (C_{min}), and a topological weight case (w_1). Clustering by squared Euclidean distance was performed for $p = 2, \ldots, 20$, with summaries of performance, best-p, and evaluation criteria (final score) provided for each case.

p	u	d	l	C_{min}	w_1
2	0.397	0.573	0.364	0.393	0.294
3	0.592	0.432	0.595	0.588	0.514
4	0.631	0.711	0.618	0.623	0.506
5	0.715	0.572	0.637	0.709	0.604
6	0.675	0.643	0.709	0.670	0.708
7	0.672	0.647	0.829	0.671	0.710
8	0.849	0.680	0.809	**0.859**	0.837
9	0.848	0.688	0.815	**0.864**	0.692
10	0.841	0.688	0.826	0.851	0.828
11	0.816	0.678	**0.861**	0.818	0.836
12	0.797	0.665	**0.853**	0.800	0.817
13	0.561	0.622	0.826	0.590	0.572
14	0.544	0.611	0.823	0.548	0.555
15	0.737	0.596	0.804	0.549	0.547
16	0.721	0.574	0.542	0.744	0.550
17	0.704	0.561	0.731	0.710	0.539
18	0.696	0.554	0.720	0.706	0.700
19	0.690	0.543	0.699	0.704	0.700
20	0.523	0.487	0.632	0.692	0.687
Q_2	0.696	0.611	**0.731**	0.704	0.690
Max	0.849	0.711	0.861	**0.864**	0.837

Figure 3: The final scores of the PFp- balancing GNC partitioning for all edge weights

As you can see from Figure 3, the best-p under the PFp-balancing GNC reached peak scores close to 7(and occasionally p = 4 or p = 6) for many weight classes. Averaging the final scores over p for each edge weighting case shows that PFp-balanced GNC produces a stable topological clustering with hydraulically viable solutions, as indicated by the hydraulically sound solutions found. Figure 3 shows the final scores for all edge weighting cases, with length edge weighting (l) leading (Q_2 = 0.731), followed by C_{min} (0.704) and unweighted (u) (0.696), while diameter (d) lags behind (0.611). The maxima reinforce this hierarchy: C_{min} achieves the best overall score (max = 0.864 at p = 9), with l almost on a par (0.861 at p = 11). The supplementary design metrics for l and C_{min}. Case p_{best} explain these patterns. The weighting case l produces the strongest mean value and almost the best maximum, albeit at the highest cost (€45,363), with a good balance ($\tilde{Q}_{DMA}$ = 7.43; σ_l = 4,003.8). C_{min} combines a balanced performance with moderate cost (€31,494) and a solid hydraulic performance of quantities $\tilde{Q}_{DMA}$ = 8.82; σ_l = 4,982.2 (Table 1, p = 9). Overall, C_{min} and l define the efficient frontier of the best alternatives among the graph partitions found, with C_{min} offering the most balanced compromise.

Table 1: Best solution of PFp-balanced GNC spectral partitioning (p=9)

p	Ir [−]	Cost [€]	$\tilde{Q}_{DMA}$ [l/s]	σ_l [m]	Score [−]
9	0.832	31,494	8.82	4,982.2	0.864

In the solution, C_{min}, p = 9 solution (Figure 4), the stacked risk–length profiles show that the PFp-balanced segmentation has effectively concentrated pipes with poorer condition (orange/red, p > 0.6) in smaller DMAs, while larger DMAs are comparatively dominated by lower-risk classes (green/yellow, p < 0.4). Compact districts such as DMA-2, DMA-4, and DMA-9 have a high percentage of orange/red segments despite their modest overall pipe length, suggesting deliberately carved 'maintenance cells' where targeted leak detection and renewal can be efficiently staggered. In contrast, DMA-5, DMA-6, and DMA-8 have a much greater absolute length, but also a significantly higher proportion of green/yellow, segments. This indicates more homogeneous, lower-risk operating zones that are more suitable for monitoring rather than immediate rehabilitation.

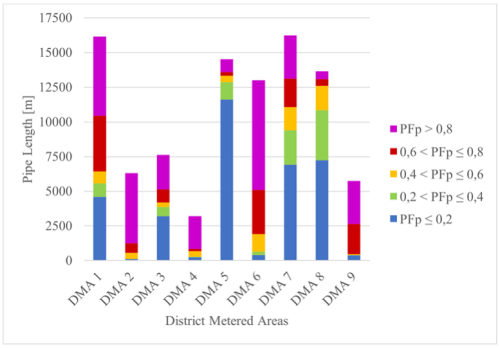

Figure 4: Length of pipes classified by DMA and pipe failure probabilities

A notable caveat is DMA-1, which combines a large overall length with a sizable red component, making it as a priority corridor where incremantal renewal or boundary secondary refinement may be warranted. Overall, the distribution confirms the intended behavior of risk-weighted vertex balancing: it locates high-risk assets in smaller, more manageable DMAs, while retaining hydraulically coherent, lower-risk areas at a larger scale. In this way, operational monitoring is aligned with condition-based maintenance and CAPEX/OPEX is concentrated where the expected returns are highest.

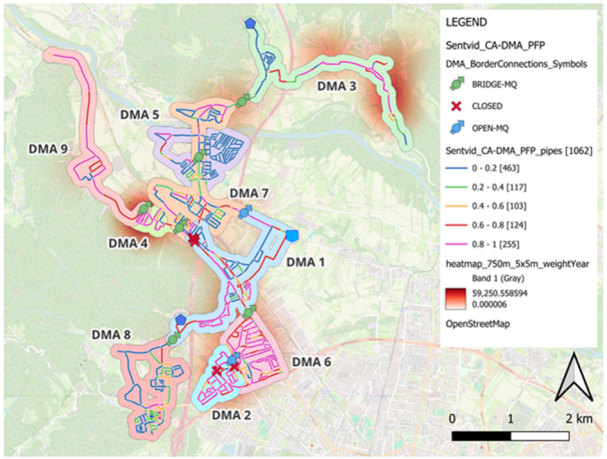

Figure 5: Heat map of the pipe failure probabilities (PFp - leak) depending on the pipe material and age

The map in Figure 5 shows the DMA layout of Ljubljana-Šentvid (DMAs 1–9) with symbols (open/closed points and bridge-MQ). The pipe segments are color-coded by PFp: green (≤0.2) to magenta (≥0.8), indicating higher-risk corridors around DMAs 1 and 6 and generally lower-risk networks in areas such as DMAs 3, 5, and 9. A red colored background heatmap shows the spatial distribution of historical failures (2010-2025).

4 Conclusions

This paper, presents a unified framework that incorporates predictive failure risk into spectral DMA segmentation by using ML-based failure probabilities for pipes. PFps were mapped from pipes to node weights using an XGBoost model and a generalized normalized cut (GNC) over standard edge weight cases were solved as vertex weights within the GNC spectral partitioning method. In the case of Šentvid Ljubljana, the risk-weighted formulation resulted in superior composite scores compared to the demand-balanced baselines of Zevnik et al. [9]. The preferred solution – C_{min}. Weighting at $p = 9$ – achieved a composite score ~0.864, with moderate implementation costs (~€31.5k), and reasonable spatial uniformity ($\sigma_l \approx 5$ km). In practice, the split equalizes pipe failure probabilities (PFp) and concentrates pipes in poor condition on actionable DMAs, while reserving larger, lower-risk zones for routine monitoring. On a system scale, the approach strengthens proactive asset management, accelerates leak reduction in aging WDNs, and supports energy and emissions savings through avoided production and pumping operations.

Acknowledgments

This paper was reviewed and corrected for grammar and style using InstaText (instatext.io).

References

[1] R. B. Sowby and A. C. Siegel, 'The increasing energy intensity of drinking water supply', *Energy Reports*, vol. 11, pp. 6233–6237, Jun. 2024, doi: 10.1016/j.egyr.2024.06.014.

[2] T. AL-Washali, S. Sharma, and M. Kennedy, 'Methods of Assessment of Water Losses in Water Supply Systems: a Review', *Water Resour Manage*, vol. 30, no. 14, pp. 4985–5001, Nov. 2016, doi: 10.1007/s11269-016-1503-7.

[3] P. Rizzo, 'Water and Wastewater Pipe Nondestructive Evaluation and Health Monitoring: A Review', *Advances in Civil Engineering*, vol. 2010, pp. 1–13, 2010, doi: 10.1155/2010/818597.

[4] MNVP, 'Information System of Public Environmental Protection Services = Informacijski sistem za spremljanje gospodarskih javnih služb varstva okolja (IJSVO)', Ministry of Natural Resources and Spatial Planning, Ljubljana, Slovenia, 2025. Accessed: Jul. 15, 2025. [Online]. Available: https://ipi.eprostor.gov.si/jgp/data

[5] European Commission, 'European Water Resilience Strategy (COM(2025) 280 final)', European Commission, Brussels. [Online]. Available: https://environment.ec.europa.eu/publications/european-water-resilience-strategy_en

[6] J. Morrison, D. Rogers, and S. Tooms, *District metered areas guidance notes*, 1st ed. in Water Loss Task Force. London: IWA Publications, 2007.

[7] T. Zhang, H. Yao, S. Chu, T. Yu, and Y. Shao, 'Optimized DMA Partition to Reduce Background Leakage Rate in Water Distribution Networks', *J. Water Resour. Plann. Manage.*, vol. 147, no. 10, p. 04021071, Oct. 2021, doi: 10.1061/(ASCE)WR.1943-5452.0001465.

[8] C. Fan, Z. Cui, and X. Zhong, 'House Prices Prediction with Machine Learning Algorithms', in *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*, Macau China: ACM, Feb. 2018, pp. 6–10. doi: 10.1145/3195106.3195133.

[9] J. Zevnik, M. Kramar Fijavž, and D. Kozelj, 'Generalized Normalized Cut and Spanning Trees for Water Distribution Network Partitioning', *J. Water Resour. Plann. Manage.*, vol. 145, no. 10, p. 04019041, Oct. 2019, doi: 10.1061/(ASCE)WR.1943-5452.0001100.

[10] A. Di Nardo, M. Di Natale, C. Giudicianni, R. Greco, and G. F. Santonastaso, 'Weighted spectral clustering for water distribution network partitioning', *Appl Netw Sci*, vol. 2, no. 1, p. 19, Dec. 2017, doi: 10.1007/s41109-017-0033-4.

[11] Q. Fang, H. Zhao, C. Xie, and T. Chen, 'A method for water supply network DMA partitioning planning based on improved spectral clustering', *Water Supply*, vol. 23, no. 8, pp. 3432–3452, Aug. 2023, doi: 10.2166/ws.2023.180.

[12] Y. Asadi, 'Employing machine learning in water infrastructure management: predicting pipeline failures for improved maintenance and sustainable operations', *Industrial Artificial Intelligence*, vol. 2, no. 1, p. 8, Nov. 2024, doi: 10.1007/s44244-024-00022-w.

[13] B. Bakhtawar, T. Zayed, and N. Elshaboury, 'Time-to-failure based deterioration factors of water networks: Systematic review and prioritization', *Reliability Engineering & System Safety*, vol. 263, p. 111246, 2025, doi: https://doi.org/10.1016/j.ress.2025.111246.

[14] M. Cabral, D. Gray, B. Brentan, and D. Covas, 'Assessing Pipe Condition in Water Distribution Networks', *Water*, vol. 16, no. 10, p. 1318, May 2024, doi: 10.3390/w16101318.

[15] A. A. M. Warad, K. Wassif, and N. R. Darwish, 'An ensemble learning model for forecasting water-pipe leakage', *Sci Rep*, vol. 14, no. 1, p. 10683, May 2024, doi: 10.1038/s41598-024-60840-x.

[16] M. Latifi, R. B. Zali, A. A. Javadi, and R. Farmani, 'Efficacy of Tree-Based Models for Pipe Failure Prediction and Condition Assessment: A Comprehensive Review', *Journal of Water Resources Planning and Management*, vol. 150, no. 7, p. 03124001, 2024, doi: 10.1061/JWRMD5.WRENG-6334.

[17] R. Jafar, I. Shahrour, and I. Juran, 'Application of Artificial Neural Networks (ANN) to model the failure of urban water mains', *Mathematical and Computer Modelling*, vol. 51, no. 9–10, pp. 1170–1180, May 2010, doi: 10.1016/j.mcm.2009.12.033.

[18] Y. Le Gat, C. Curt, C. Werey, K. Caillaud, B. Rulleau, and F. Taillandier, 'Water infrastructure asset management: state of the art and emerging research themes', *Structure and Infrastructure Engineering*, vol. 21, no. 4, pp. 539–562, Apr. 2025, doi: 10.1080/15732479.2023.2222030.

[19] D. Kozelj and D. A. Fernández, 'Predicting Water Distribution Pipe Failures Using Machine Learning and Cross-Infrastructure Data', *Acta hydrotechnica*, pp. 53–64, Jun. 2025, doi: 10.15292/acta.hydro.2025.05.

[20] D. Arthur and S. Vassilvitskii, 'K-means++: The advantages of careful seeding', in *Proc., 18th Annual ACM-SIAM Symp. on Discrete Algorithms*, Philadelphia (USA): Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[21] GURS, 'Data on public infrastructure = Zbirni kataster gospodarske javne infrastrukture', Surveying and Mapping Authority of the Republic of Slovenia (GURS), Ljubljana, Slovenia, 2025. Accessed: Jul. 15, 2025. [Online]. Available: https://ipi.eprostor.gov.si/jgp/data

[22] VOKAS, 'Pipe burst register', JP VOKA SNAGA d.o.o., 2025.

[23] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[24] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, 'Optuna: A Next-generation Hyperparameter Optimization Framework', in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA: ACM, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.

[25] E. Creaco, M. Franchini, and E. Todini, 'Generalized Resilience and Failure Indices for Use with Pressure-Driven Modeling and Leakage', *J. Water Resour. Plann. Manage.*, vol. 142, no. 8, p. 04016019, Aug. 2016, doi: 10.1061/(ASCE)WR.1943-5452.0000656.

[26] Z. Liu, Y. Kleiner, B. Rajani, L. Wang, and W. Condit, 'Condition assessment technologies for water transmission and distribution systems', U.S. Environmental Protection Agency, EPA/600/R-12/017, Apr. 2012. [Online]. Available: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=241510&Lab=NRMRL