Water Supply

Water Supply Vol 00 No 0, 1 doi: 10.2166/ws.2023.335

A study on pipe failure analysis in water distribution systems using logistic regression

I. Ethem Karadirek [a,*], Ebru Kaya-Basar [b] and Tugba Akdeniz [c]

[a] Department of Environmental Engineering, Faculty of Engineering, Akdeniz University, Antalya, Türkiye
[b] Akdeniz University, Statistical Consulting Application and Research Center, Antalya, Türkiye
[c] Antalya Water and Wastewater Administration, Antalya, Türkiye
*Corresponding author. E-mail: ethemkaradirek@akdeniz.edu.tr

IEK, 0000-0003-3689-4125; EK-B, 0000-0001-6204-3143

ABSTRACT

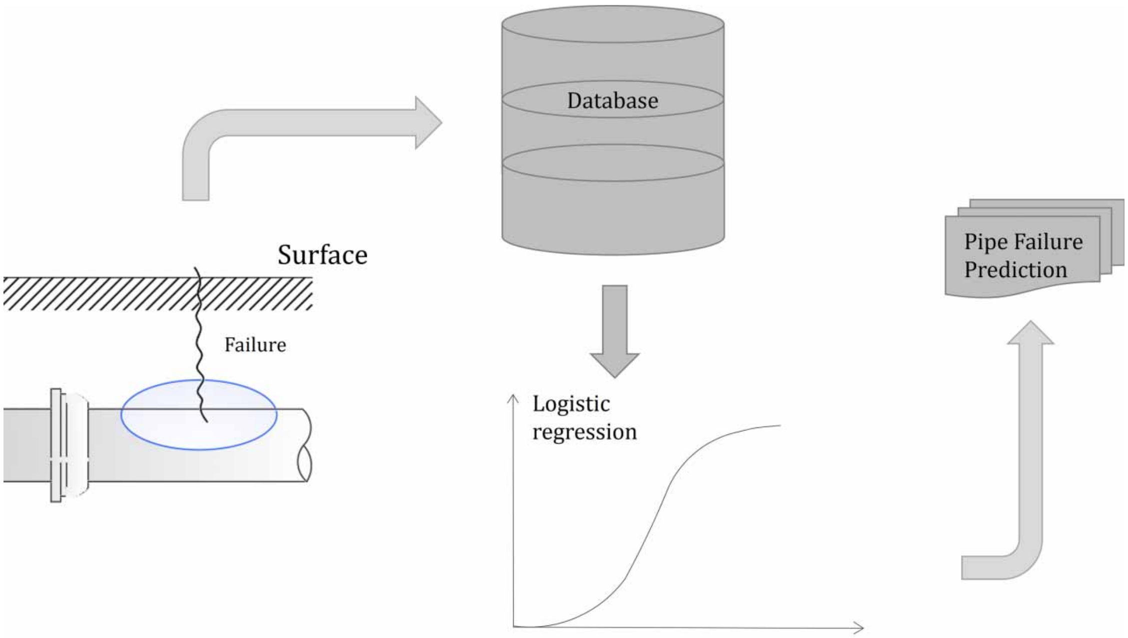
This study presents a pipe failure analysis that aims to identify key factors affecting pipe failure in a pilot study area (PSA). For this purpose, pipe failures that occurred in the PSA were recorded for 2 years. Recorded pipe failures were integrated into the existing geographical information systems of the PSA and then visualized. Then, the logistic regression model that calculates failure probability with unbalanced and small- and medium-sized data was performed to analyze pipe failures. Pipe material, age, diameter, burst pressure, and average water pressure are the factors used in the study. The developed model achieved a success rate of 70.1%, suggesting that it can predict a considerable portion of pipe failures with a relatively satisfactory performance. Findings show that the failure probability of pipes with an age greater than 20 years is 1.675 times higher than that of pipes with an age less than 20 years. Moreover, polyvinyl chloride pipes are 5.265 times and high-density polyethylene pipes are 9.027 times more likely to fail than ductile pipes. Further research should delve into the impacts of the number of service connections, pipe lengths, and traffic load on pipe failures.

Key words: logistic regression, pipe failure, real losses, water distribution systems, water losses

HIGHLIGHTS

- Pipe failure analysis is a key to implementing asset management strategies.
- Logistic regression allows analysis with unbalanced and small- and medium-sized data, which is the case for many water utilities.
- The developed model achieved a success rate of 70.1%, indicating that a significant proportion of pipe failures can be predicted with relatively satisfactory performance.

GRAPHICAL ABSTRACT



1. INTRODUCTION

As a finite resource, water is under pressure due to climate change, population growth, and industrialization. Each year, an estimated amount of 48 billion m^3 of water is lost in water distribution systems (WDSs) (Kingdom *et al.* 2006). Water losses comprise real and apparent losses. Real losses are associated with leakages on mains, distribution, and service connection pipes, and with leakage and overflows at storage tanks, while apparent losses result from meter inaccuracies, data handling errors, and illegal water consumption (Lambert *et al.* 1999). Real losses are the total volume of water that is physically lost. Water utilities are responsible for providing safe and adequate amounts of water to consumers through WDSs, which are complex systems that should be managed proactively. A reliable urban water infrastructure system is a must for continuous water service to consumers. Failure of WDSs poses many environmental, economic, and public health-related risks (Clair & Sinha 2012). Pipes, the major actor of WDSs, deteriorate over time depending on the pipe material, diameter, soil properties, etc.

Real losses in WDS can be controlled by implementing active and passive leakage detection techniques (Karadirek & Aydin 2022). Repairing leaks and bursts that are reported by the local community and/or become visible on the surface is called the passive leakage detection technique, whereas active leakage control refers to a planned program for the awareness, localization, and repairment of leaks and bursts (Shammas & Al-Dhowalia 1993). Water utilities all around the world put a great effort into implementing an optimal plan for leakage control in WDSs (Robles-Velasco *et al.* 2020). Developing a decent plan is not only a must for water loss control strategies but is also helpful for asset management to determine short- and long-term failure probabilities in WDSs (Watson *et al.* 2004). The detection of pipe failures in WDSs is not an easy task, as water distribution pipes are buried (Robles-Velasco *et al.* 2020). Many studies have been carried out to predict pipe failures in WDSs (Mailhot *et al.* 2000; Tabesh *et al.* 2009; Alvisi & Franchini 2010; Kakoudakis *et al.* 2017; Wilson *et al.* 2017). Pipe failures can be predicted by implementing heuristic, physical, and/or statistical models that are summarized in previous studies (Scheidegger *et al.* 2015; Wilson *et al.* 2017; Barton *et al.* 2022). Heuristic models are sometimes incapable of reflecting all potential risks, as these models are based on subjective judgment (Fitchett *et al.* 2020; Barton *et al.* 2022). Physical models, which give a result of a certain pipe and are not extrapolated to other parts of WDSs, are associated with an understanding of pipe structural properties, environmental conditions, and deterioration of materials (Kleiner & Rajani 2001; Wilson *et al.* 2017; Barton *et al.* 2022). On the other hand, statistical models require historical data for the prediction of pipe failures in WDSs (Barton *et al.* 2019, 2020). Statistical models, one of the most preferred modeling techniques for pipe failure analysis, are based on applying statistical techniques using historical pipe failure data (Snider & McBean

2020). Statistical models can be classified as deterministic, probabilistic, and machine learning models (Snider & McBean 2020; Barton *et al.* 2022). Deterministic models are based on the calculation of input parameter coefficients by implementing regression techniques, whereas probabilistic models are used to predict the probability of pipe failure by using historical data (Snider & McBean 2020). Contrary to deterministic and probability models, machine learning models rely on the identification of the relationship between input and output data by adopting a data-driven approach (Snider & McBean 2020). Numerous pipe failure prediction models have been developed and these studies have been reviewed in previous studies (Wilson *et al.* 2017; Snider & McBean 2020; Barton *et al.* 2022). Barton *et al.* (2022) presented the advantages and disadvantages of statistical models. Yamijala *et al.* (2009) carried out a study using a 5-year failure history for the analysis of pipe breaks by implementing different statistical regression models and found that the logistic regression model could be useful for identifying pipe failure risks in WDSs. Kleiner & Rajani (2009) calculated the failure probability using a 44-year failure history by implementing a non-homogenous Poisson process. Jafar *et al.* (2010) used artificial neural networks for modeling pipe failures using a 14-year pipe failure history in a city in France. Debon *et al.* (2010) presented the results of a study using a 6-year failure history for the prediction of pipe failures in a medium-sized Spanish city and concluded that the logistic regression model showed a good prediction performance. Alvisi & Franchini (2010) carried out a study using a 6-year failure history. In this study, two probabilistic models, namely the Weibull proportional hazard model and the Weibull exponential model, were implemented to model pipe breakage. Bayesian belief networks were used by Francis *et al.* (2014) to predict pipe failure probability. The study revealed that the development of Bayesian belief networks for the prediction of pipe failure probability is complex but potentially very rewarding. Tang *et al.* (2019) predicted pipe failure probability using a 37-year failure by implementing a guided learning Bayesian network. Motiee & Ghasemnejad (2019) implemented linear regression, exponential regression, Poisson generalized linear regression, and logistic generalized regression models to predict the pipe failure rate in Tehran's WDS and concluded that the logistic regression model for the study area showed better prediction performance. The logistic regression and support vector classification were utilized to predict pipe failure by Robles-Velasco *et al.* (2020). For this purpose, 7-year pipe failure data from a Spanish city was used, and the study revealed that the performance of logistic regression was better than that of support vector classification. Moreover, the study suggested developing a global tool that integrates both failure probability and its resulting impacts (Robles-Velasco *et al.* 2020). Liu *et al.* (2022) predicted pipe failures by implementing two machine learning algorithms, namely random forest and logistic regression. The study revealed that the random forest algorithm showed better performance than the logistic regression. Most of the studies on pipe failure prediction have been carried out using pipe failure data over a relatively longer time. This study aims to provide an analysis of pipe failures in WDSs using the logistic regression model with a limited and unbalanced dataset. For this purpose, a study area was chosen, and pipe failures that occurred in the study area were recorded for 2 years. The studies presented in the literature have been mainly carried out with pipe failure data over a relatively longer time. Although the majority of pipes in the pilot study area (PSA) are relatively new, they are still failure-prone.

2. MATERIAL AND METHODS

2.1. Study area

Antalya is a city with a warm climate, which is located along the Mediterranean coast in the south of Türkiye. Groundwater is the only source of water supplied to the city. The city's WDS has been divided into nine pressure zones due to varying elevations. The WDS of the city serves from the sea level to 240 m above the sea level. The Konyaalti region is one of the pressure zones that is operated independently from the rest of the city and supplies water from the sea level up to 30 m above the sea level. Konyaalti's WDS has been selected as the PSA for this study, which is depicted in Figure 1.

Many studies have been carried out in the PSA for hydraulic and water quality modeling and water loss control (Karadirek *et al.* 2012, 2016). The PSA was divided into district-metered areas (DMAs) for better management of water losses, and the level of water losses was around 43.5% of the system input volume. The study revealed a significant amount of reduction in water losses (Karadirek *et al.* 2012). Since then, the population that is served by the WDS of the PSA increased, and thus, some of the DMAs have been merged, new groundwater wells have been included in the system, and the total length of the distribution pipes has increased. Although a lot of effort has been made in the last decade, the level of water losses in Antalya in 2020 was still around 43.84% of the system input volume (ASAT 2021). The PSA has an efficient supervisory control and data acquisition system (SCADA) that helps with the monitoring and evaluation of flow rates and water pressure levels. Active and passive leakage control methods have been applied in the PSA to control water losses. Active leakage

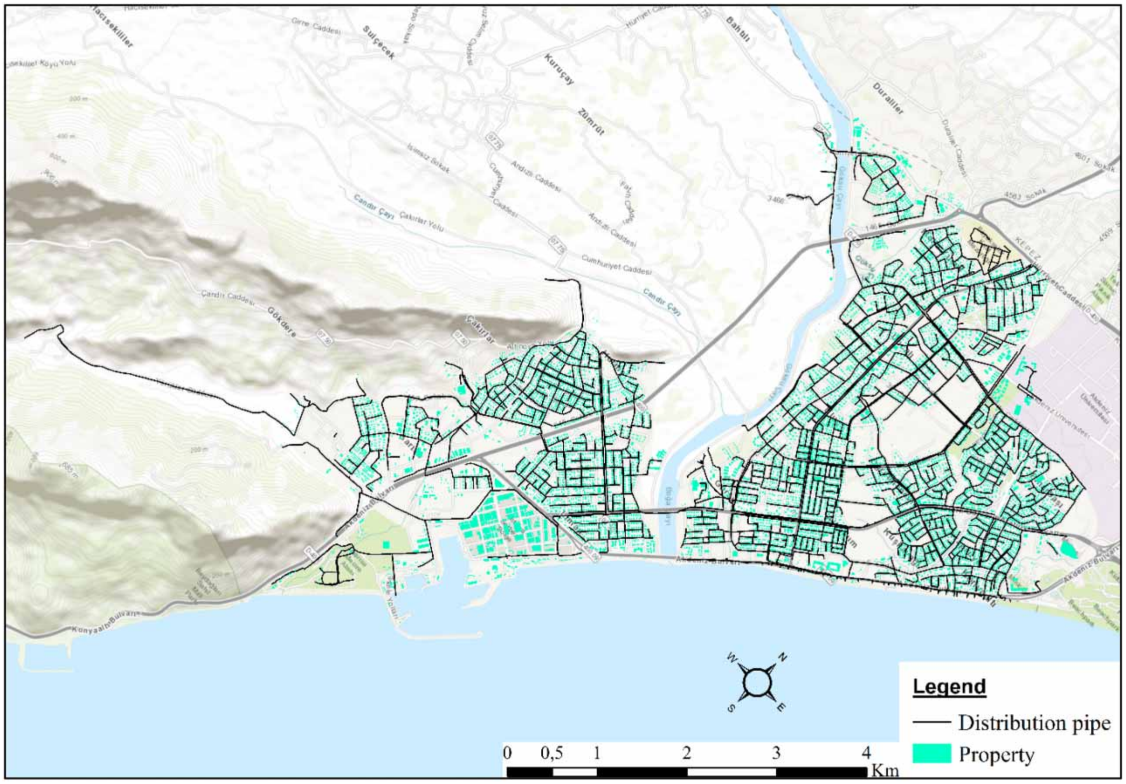


Figure 1 | Map showing the WDS of the PSA.

control strategies have been implemented with the help of the SCADA system in addition to leak detection with leak loggers and ground microphones.

Raw water extracted from *Bogacay* wells (eight groundwater wells) and *Candir* wells (five groundwater wells) is pumped to the PSA. There is only one balancing reservoir with a capacity of 15,000 m^3 that is used to balance consumption fluctuations and to store an adequate amount of water to meet the requirements in case of an emergency. The PSA serves around 80,000 people, and the total length of distribution pipes is around 265 km with different pipe materials and an average age of 18 years. The distribution pipes of the PSA consist of polyvinyl chloride (PVC), high-density polyethylene (HDPE), ductile, molecularly oriented PVC (O-PVC), steel, and asbestos cement pipes (ACP).

2.2. Statistical analysis

The main steps of the applied methodology of the pipe failure analysis are given in Figure 2. Logistic regression, which calculates failure probability *P*, has been selected to analyze pipe failures in the PSA. The pipe failure in this study stands for the maintenance activities including leaks and bursts. Leaks detected by proactive leakage management programs conducted in the study area have not been considered to be failures. Logistic regression allows analysis with unbalanced and small- and medium-sized data, which is the case for many water utilities.

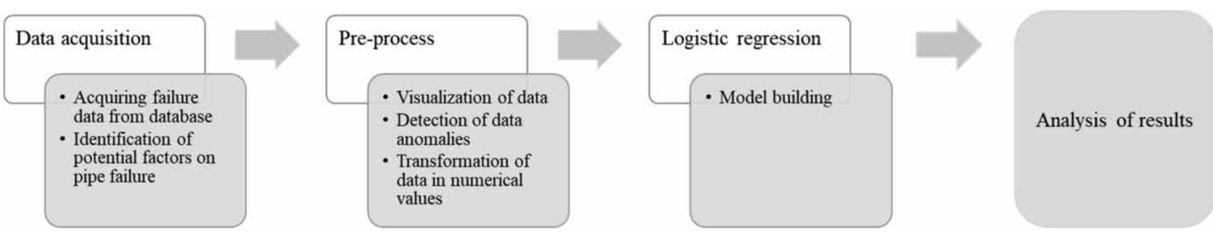


Figure 2 | The main steps of the pipe failure analysis applied in the study.

Logistic regression gives an output of 0 and 1 that can be explained as pipe failure probability (Robles-Velasco *et al.* 2020). The main goal of the logistic regression model is to express the relationship between dependent and independent variables (Hosmer *et al.* 2013). The general term of logistic regression is as follows (Motiee & Ghasemnejad 2019):

$$P = \frac{e^{(\alpha+\beta o+\beta 1x1+\ \beta 2x2\cdots\beta nxn)}}{1+e^{(\alpha+\beta o+\beta 1x1+\ \beta 2x2\cdots\beta nxn)}}$$

where α is the constant parameter of regression, β is the regression coefficients of variables, and x stands for the independent variable. The Logit function of the model can be expressed as follows:

$$\text{Logit}\ [P] = \log\frac{P}{1-P} = \alpha + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where P is the pipe failure probability, $1 - P$ stands for no failure, α is the constant parameter of regression, and β is the calculated regression parameter.

The number of failures and categorical data of distribution pipes of the PSA were collected and transformed into numerical values. A pre-processing was carried out, as input greatly affects performance analysis. Records on pipe failures often include the failures of service connections. The failures on service connection pipes were excluded, as this study aims to analyze pipe failures of WDSs. Furthermore, samples representing anomalies were not included in the study. As previously mentioned, the pipe failure database of water utilities is usually limited. In this study, pipe failures that occurred in the PSA were recorded in 2020 and 2021. The total number of pipes with failures and without failures for the 2-year monitoring period are 811 and 1,466, respectively. The factors used in this study are presented in Table 1.

In this study, independent risk factors affecting pipe failure are pipe diameter, burst pressure, average water pressure, pipe material, and age. Pipe failure is a binary dependent variable based on the presence or absence of failure. Descriptive statistics are presented in statistical analysis as frequency, percentage, mean, standard deviation, median, minimum, maximum, 25th percentile ($Q1$), and 75th percentile ($Q3$). In categorical data analysis, Fisher's exact test was used if the percentage of cells with an expected value less than five was greater than 20%, and the Pearson χ^2 test was used if the expected value was less than five. The normality assumption was checked with the Kolmogorov–Smirnov test. The variables of pipes with failure and without failure do not fit the normal distribution. Therefore, the Mann–Whitney U test (a non-parametric statistical test used to determine if there is a significant difference between independent groups when assumptions of normal distribution are not met) was used in this study. In the analysis of the difference between the numerical data of the two groups, the Mann–Whitney U test was used as the data did not fit the normal distribution. Variables with a p-value less than 0.25 (using the χ^2 test with categorical variables and the Mann–Whitney U test with continuous variables) were included in the binary logistic regression analysis with the enter method to identify risk factors for pipe failure. A cutoff value of 0.25 is supported by literature (Bendel & Afifi 1977; Mickey & Greenland 1989). The p-value less than 0.05 was considered statistically significant.

Table 1 | Description of data used in the study

Factor	Definition	Unit	Minimum	Maximum
Mt	Pipe material: PVC O-PVC HDPE Steel Ductile ACP	–	–	–
Dia	Pipe diameter	mm	50	800
Age	Years since installation	Years	0.25	25
APr	Average water pressure	bar	3.49	5.40
Atu	Burst pressure (manufacturer's pressure rating)	bar	9.81	15.70

3. RESULTS

3.1. Analysis of distribution pipes

The distribution pipes of the PSA consist of PVC, HDPE, ductile, O-PVC, steel, and ACP. The percentage of each pipe material to the total pipe length is given in Figure 3(a). The PVC pipes are the most installed in the PSA with a ratio of 49.7% of the total pipe length, whereas the ratio of ACP pipes is only 2.4% of the total pipe length. ACP pipes have been started to be replaced in case of failures. The distribution systems are complex systems that get older over time like living organisms. The distribution pipes of the PSA have an average age of 18 years. The oldest pipe material of the PSA is ACP, whereas the O-PVC pipes are the youngest distribution pipes. Most PVC pipes (66.2%) are older than 20 years, and 58.3% of HDPE pipes are between 10 and 15 years of age. The percentage of each pipe's age to the total length of the pipes is given in Figure 3(b).

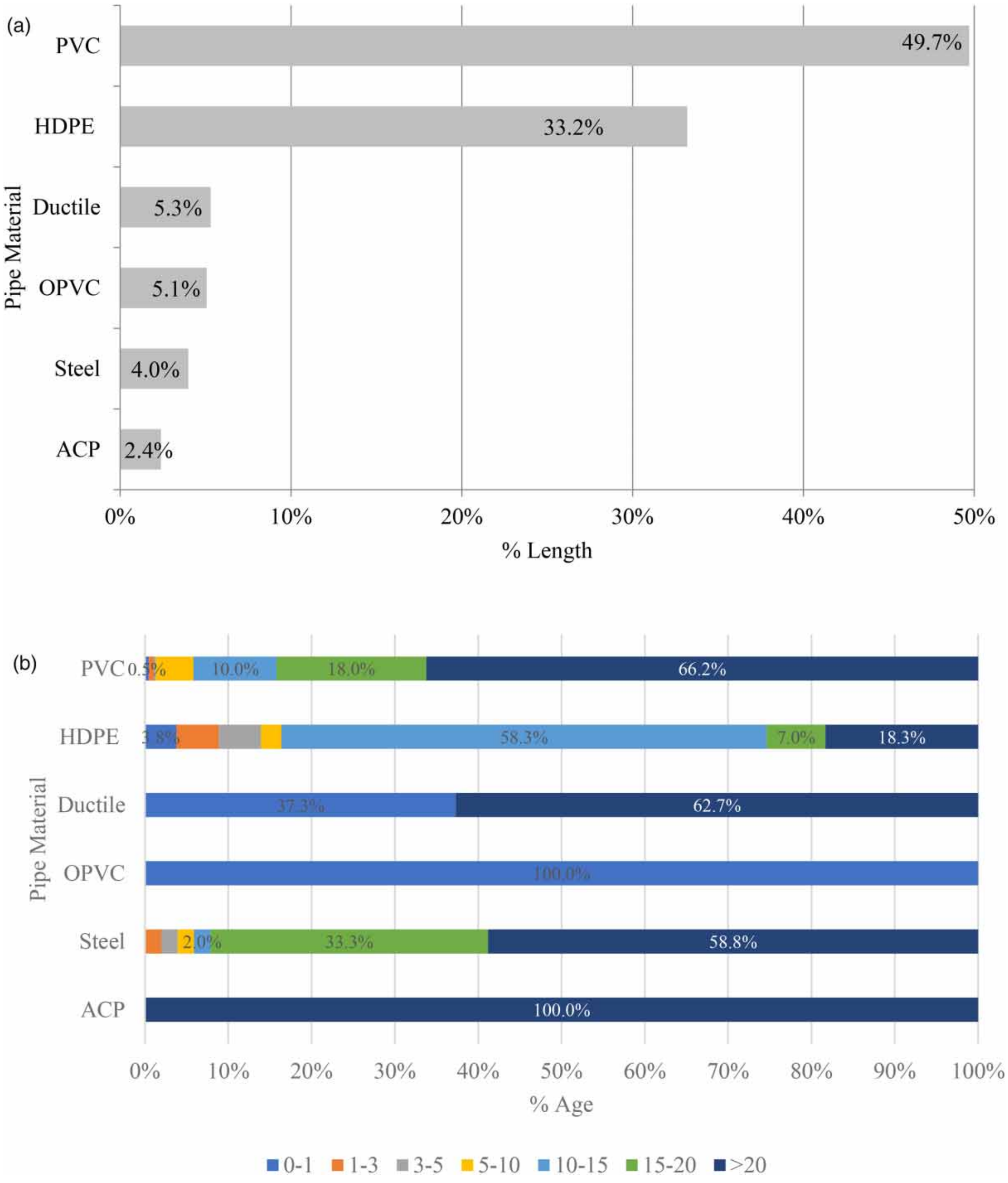


Figure 3 | (a) Percentage of distribution pipes based on the pipe material and (b) distribution of pipe age.

3.2. Analysis of pipe failures

The total number of pipe failures (811) during the study period is relatively high. The number of recorded pipe failures during the study period is around 1.53 failures per km/year. The locations of pipe failures during the study period are visualized and given in Figure 4. The locations of pipe failures that occurred in the PSA during the study period have a positive correlation with population density. The pipe failure frequency increases with population density.

Older pipes are generally expected to have a higher failure frequency. Pipes that are not older than 1 year have the lowest pipe failure frequency, whereas pipes with an age between 5 and 10 years have the highest failure frequency in the PSA (Figure 5(a)). In addition to pipe age, material, diameter, burst pressure, and average water pressure are the parameters that should be considered in pipe failure analysis. The PVC and HDPE pipes, which are the most installed in the PSA, have the most failures. The steel pipes have the lowest pipe failures, with 0.05 failures per km/year, while the number of failures of PVC and HDPE pipes is 1.81 and 1.80 failures per km/year, respectively (Figure 5(b)).

As it is stated, pipe failures take place due to many variables that should be addressed. Age, material, diameter, burst pressure of pipes, and average water pressure in the pipes should be analyzed altogether. Logistic regression, which allows analysis with unbalanced and small- and medium-sized data, was applied to calculate failure probability and explain the relationship between pipe failure and the factors mentioned above. A p-value less than 0.05 is considered significant in all statistical analyses. In the first phase, univariate statistical analyses were carried out to determine which factors should be included in the logistic regression model as only variables with $p < 0.25$ were included in the model. The relationships between pipe diameter, average water pressure, and pipe failure were investigated using the Mann–Whitney U test. According to the analysis results, it can be concluded that pipe diameter and average water pressure have statistical differences between the pipe failures group in Table 2 (respectively, $p = 0.001 < 0.05$ and $p = 0.001 < 0.05$).

The relationships between burst pressure, pipe material, pipe age, and pipe failure were examined by implementing Pearson's χ^2 and Fisher's exact tests and are summarized in Table 3.

The analysis shows that there is a relationship between burst pressure and pipe failure as the p-value is 0.028, which is lower than 0.05. As expected, the pipe material has also been found as an important factor. To investigate the independent risk

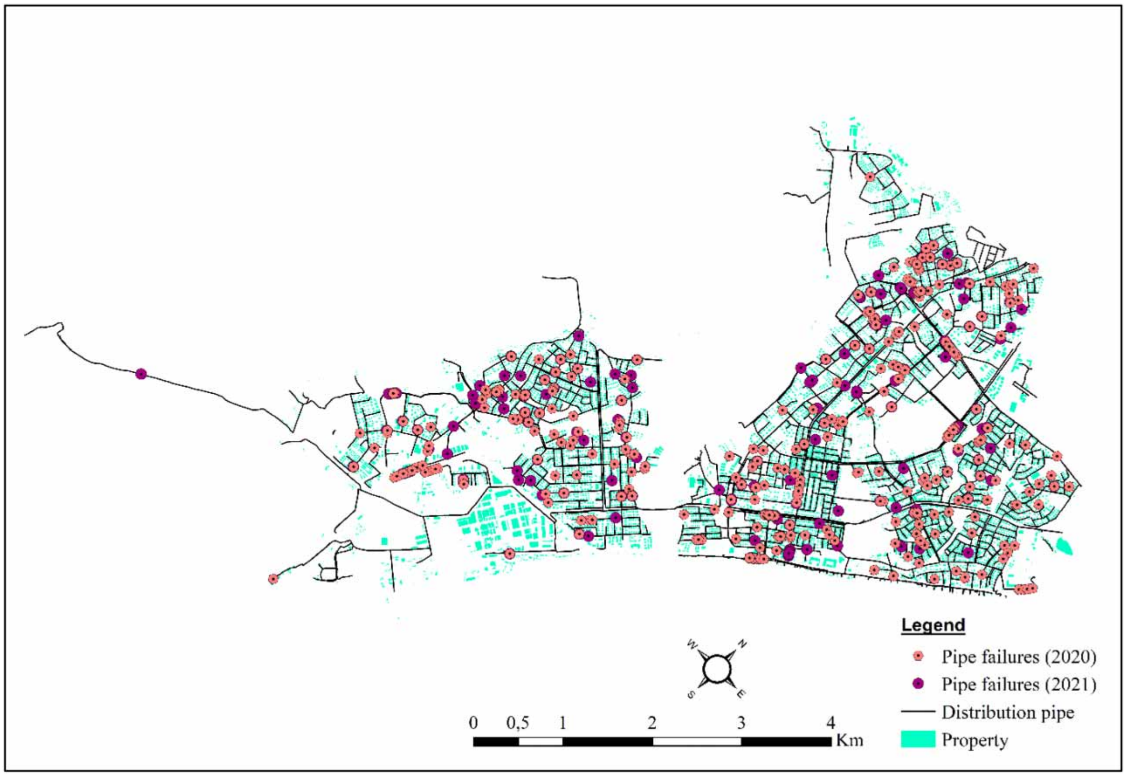


Figure 4 | Map showing the locations of pipe failures in the PSA.

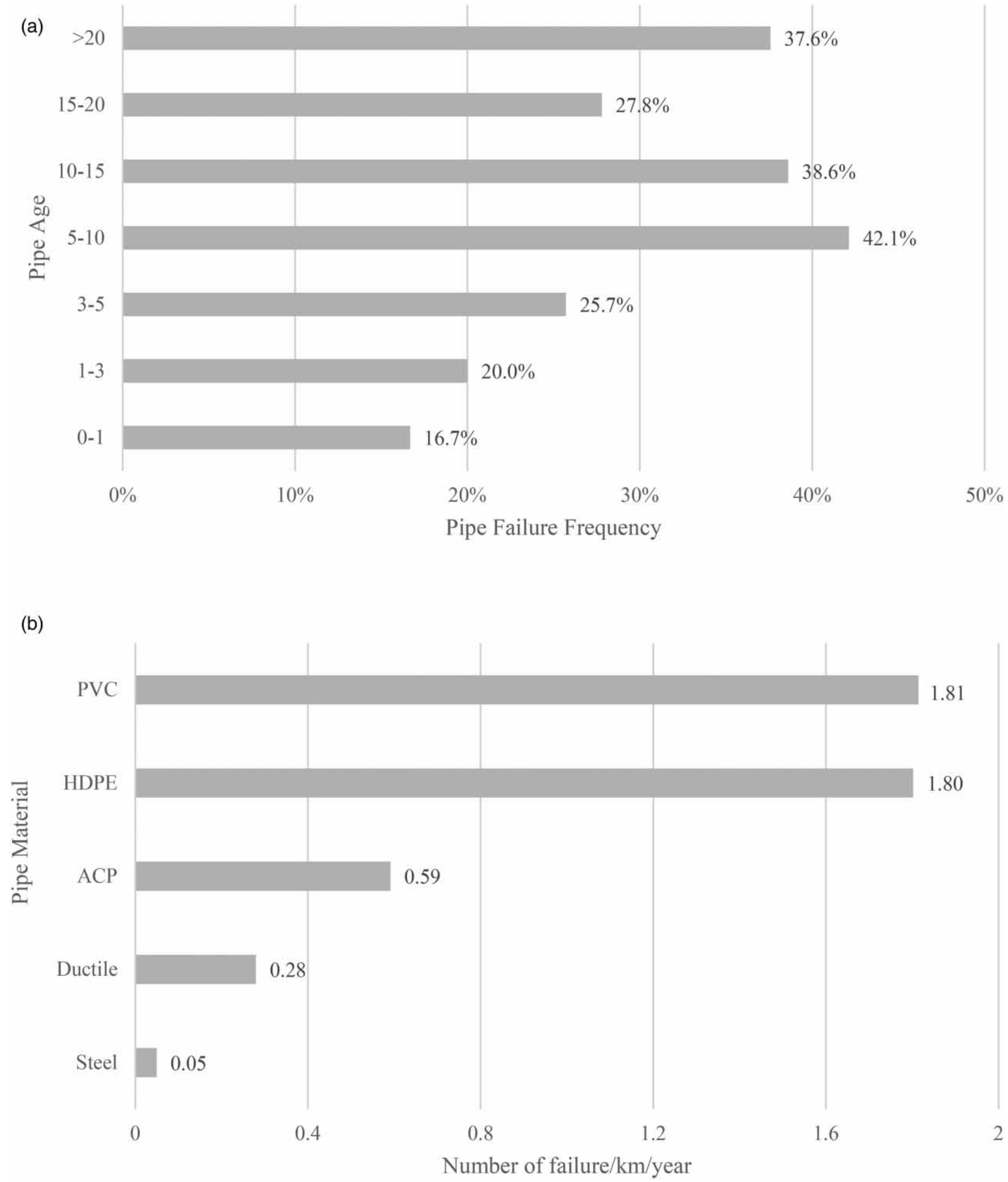


Figure 5 | (a) Pipe failure frequency based on pipe age and (b) number of pipe failures per km/year.

Table 2 | The relationship between pipe diameter, average water pressure, and pipe failure

Factor	Failure/no failure	*n*	Mean ± SD (minimum–maximum)	Median (Q1–Q3)	*p*
Pipe diameter	No failure	1,466	139.86 ± 110 (32–800)	110 (110–110)	0.001
	Failure	811	125.14 ± 110 (50–800)	110 (110–110)	
Average water pressure	No failure	1,466	4.61 ± 4.65 (3.49–5.4)	4.65 (4.58–4.92)	0.001
	Failure	811	4.61 ± 4.74 (3.49–5.4)	4.74 (4.44–4.96)	

SD: standard deviation; Q1: 25th percentile; Q3: 75th percentile.

Table 3 | The relationship between burst pressure, pipe material, pipe age, and pipe failure

Factor	Failure status No failure *n* (%)	 Failure *n* (%)	Total *n* (%)	*p*
Atu				
9.81	1,367 (93.25)	740 (91.20)	2,107 (92.53)	0.028[a]
12.26	5 (0.34)	0 (0)	5 (0.22)	
15.70	94 (6.41)	71 (8.80)	165 (7.25)	
Pipe material				
PVC	873 (59.55)	475 (58.57)	1,348 (59.20)	<0.0001[b]
HDPE	437 (29.81)	315 (38.84)	752 (33.03)	
Steel	47 (3.21)	5 (0.62)	52 (2.28)	
ACB	50 (3.41)	8 (0.99)	58 (2.55)	
Ductile	59 (4.02)	8 (0.99)	67 (2.94)	
Pipe age				
<20	775 (52.86)	395 (48.71)	1,170 (51.38)	0.057[b]
≥20	691 (47.14)	416 (51.29)	1,107 (48.62)	

[a]Fisher's exact tests were used.
[b]Pearson's χ^2.

factors affecting the pipe failures, the variables that were significant in the univariate analysis ($p < 0.25$) were included in the binary logistic regression model. To examine the multicollinearity problem, variance influence factor (VIF) values were examined, and it was seen that there was no multicollinearity problem between the variables included in the model (VIF < 10). Details of binary logistic regression analysis are summarized in Table 4.

The success rate of logistic regression models depends on the nature of the problem, data quality, and the specific goal of the model. The success rate of the developed model in this study, which is 70.1%, can be considered to be relatively good. Pipe diameter, burst pressure, and average water pressure were not among the risk factors affecting pipe failure ($p > 0.05$), whereas pipe material and age were found to be independent risk factors for pipe failures ($p < 0.05$). The final model that represents

Table 4 | The details of the binary logistic regression model

Predictor	*B*	Wald	Sig.	Exp(*B*)	95% CI for exp(*B*) Lower	 Upper
Dia	0.001	1.254	0.263	1.001	0.999	1.003
Atu (9.81)		0.244	0.885			
Atu (12.26)	0.093	0.244	0.621	1.097	0.76	1.584
Atu (15.70)	−20.22	0.000	0.999	0.000	0.000	
Apr	0.036	0.162	0.687	1.036	0.872	1.232
Pipe material		52.144	<0.0001			
Mt_(PVC)	1.661	13.6	<0.0001	5.265	2.178	12.729
Mt_(HDPE)	2.2	23.192	<0.0001	9.027	3.687	22.103
Mt (Steel)	−0.348	0.313	0.576	0.706	0.208	2.392
Mt_(ACB)	0.165	0.088	0.767	1.179	0.397	3.505
Age (<20)	−0.516	22.955	<0.0001	0.597	0.483	0.737
Constant	−2.47	12.292	<0.0001	0.085		

Model χ^2: 65.38; −2 Log Likelihood (−2LL): 1,871.3; $n = 2{,}277$; $p < 0.0001$; success rate = 70.1%; Cox&Snell $R^2 = 0.203$.

the odds of pipe failure is given below:

$$\text{Logit} = \text{Log(odds of pipe failure)} = -2.47 + (1.661{*}\text{Mt(PVC)}) + (2.2{*}\text{Mt(HDPE)}) - (0.516{*}\text{Age}(<20))$$

Findings show that the failure probability of pipes with an age greater than 20 years is 1.675 times higher than that of pipes with an age less than 20 years. Furthermore, PVC pipes are 5.265 times and HDPE pipes are 9.027 times more likely to fail than ductile pipes.

4. CONCLUSIONS

Pipe failure analysis provides valuable insights into understanding the vulnerabilities of infrastructure systems and plays a key role in maintaining resilience in water supply systems. In this study, the logistic regression model was applied to analyze failure probability. The study shows that the logistic regression model is efficient for such studies as logistic regression allows analysis with unbalanced and small- and medium-sized data, which is the case for many water utilities. Pipe age and material are the risk factors for pipe failure in the PSA. The success rate of the developed model was found to be 70.1%, indicating that a significant proportion of the pipe failures can be predicted with relatively good performance. The importance of pipe length in pipe failure analysis studies cannot be overstated. Unfortunately, due to a lack of reliable data, pipe length could not be included as a feature in this study. Although the majority of pipes in the PSA are relatively new, they are still prone to failure. The number of service pipe connections, traffic load, and pipe lengths should be taken into account in further studies.

ACKNOWLEDGEMENTS

This study was supported by the Antalya Water and Wastewater Administration (ASAT) and Akdeniz University, Antalya, Türkiye. The authors would like to thank Huseyin Ozgur Baydilli for his support during data collection. The authors would like to express their gratitude to the anonymous reviewers for their valuable comments.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alvisi, S. & Franchini, M. 2010 Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system. *Civil Engineering and Environmental Systems* **27** (1), 1–22.

ASAT. 2021 *Annual Report of Antalya Water Supply System*. (In Turkish). Available from: https://www.asat.gov.tr/images/editor/Laboratuvar/Su%20kayiplari%20yillik%20raporu%202020.pdf (accessed 30 March 2023).

Barton, N. A., Farewell, T. S., Hallett, S. H. & Acland, T. F. 2019 Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Research* **164**, 16.

Barton, N. A., Farewell, T. S. & Hallett, S. H. 2020 Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *Npj Clean Water* **3** (1), 12.

Barton, N. A., Hallett, S. H., Jude, S. R. & Tran, T. H. 2022 An evolution of statistical pipe failure models for drinking water networks: A targeted review. *Water Supply* **22** (4), 3784–3813.

Bendel, R. B. & Afifi, A. A. 1977 Comparison of stopping rules in forward stepwise regression. *Journal of the American Statistical Association* **72** (357), 46–53.

Clair, A. M. S. & Sinha, S. 2012 State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! *Urban Water Journal* **9** (2), 85–112.

Debon, A., Carrion, A., Cabrera, E. & Solano, H. 2010 Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering & System Safety* **95** (1), 43–48.

Fitchett, J. C., Karadimitriou, K., West, Z. & Hughes, D. M. 2020 Machine learning for pipe condition assessments. *Journal American Water Works Association* **112** (5), 50–55.

Francis, R. A., Guikema, S. D. & Henneman, L. 2014 Bayesian belief networks for predicting drinking water distribution system pipe breaks. *Reliability Engineering & System Safety* **130**, 1–11.

Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. 2013 *Applied Logistic Regression*. John Wiley & Sons, Inc., Hoboken, NJ.

Jafar, R., Shahrour, I. & Juran, I. 2010 Application of artificial neural networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling* **51** (9–10), 1170–1180.

Kakoudakis, K., Behzadian, K., Farmani, R. & Butler, D. 2017 Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. *Urban Water Journal* **14** (7), 737–742.

Karadirek, I. E., Aydin, M. E., 2022 Water losses management in urban water distribution systems. In: *Water and Wastewater Management: Global Problems and Measures* (Bahadir, M. & Haarstrick, A., eds). Springer International Publishing, Cham, pp. 53–65.

Karadirek, I. E., Kara, S., Yilmaz, G., Muhammetoglu, A. & Muhammetoglu, H. 2012 Implementation of hydraulic modelling for water-loss reduction through pressure management. *Water Resources Management* **26** (9), 2555–2568.

Karadirek, I. E., Kara, S., Muhammetoglu, A., Muhammetoglu, H. & Soyupak, S. 2016 Management of chlorine dosing rates in urban water distribution networks using online continuous monitoring and modeling. *Urban Water Journal* **13** (4), 345–359.

Kingdom, B., Liemberger, R. & Marin, P. 2006 *The Challenge of Reducing Non-Revenue Water (NRW) in Developing Countries How the Private Sector Can Help: A Look at Performance-Based Service Contracting*. The World Bank, Washington, DC, USA.

Kleiner, Y. & Rajani, B. 2001 Comprehensive review of structural deterioration of water mains: Statistical models. *Urban Water* **3** (3), 131–150.

Kleiner, Y. & Rajani, B. 2009 I-WARP: Individual water main renewal planner. In: *10th International Conference on Computing and Control for the Water Industry, CCWI 2009 Conference [Proceedings]*, p. 10.

Lambert, A. O., Brown, T. G., Takizawa, M. & Weimer, D. 1999 A review of performance indicators for real losses from water supply systems. *Journal of Water Supply: Research and Technology – AQUA* **48** (6), 227–237.

Liu, W., Wang, B. H. & Song, Z. Y. 2022 Failure prediction of municipal water pipes using machine learning algorithms. *Water Resources Management* **36** (4), 1271–1285.

Mailhot, A., Pelletier, G., Noel, J. F. & Villeneuve, J. P. 2000 Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. *Water Resources Research* **36** (10), 3053–3062.

Mickey, R. M. & Greenland, S. 1989 The impact of confounder selection criteria on effect estimation. *American Journal of Epidemiology* **129** (1), 125–137.

Motiee, H. & Ghasemnejad, S. 2019 Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. *Water Supply* **19** (3), 695–702.

Robles-Velasco, A., Cortes, P., Munuzuri, J. & Onieva, L. 2020 Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliability Engineering & System Safety* **196**, 12.

Scheidegger, A., Leitao, J. P. & Scholten, L. 2015 Statistical failure models for water distribution pipes – A review from a unified perspective. *Water Research* **83**, 237–247.

Shammas, N. K. & Al-Dhowalia, K. H. 1993 Effect of pressure on leakage rate in water distribution networks. *Journal of King Saud University – Engineering Sciences* **5** (2), 213–226.

Snider, B. & McBean, E. A. 2020 Watermain breaks and data: The intricate relationship between data availability and accuracy of predictions. *Urban Water Journal* **17** (2), 163–176.

Tabesh, M., Soltani, J., Farmani, R. & Savic, D. 2009 Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. *Journal of Hydroinformatics* **11** (1), 1–17.

Tang, K., Parsons, D. J. & Jude, S. 2019 Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliability Engineering & System Safety* **186**, 24–36.

Watson, T. G., Christian, C. D., Mason, A. J., Smith, M. H. & Meyer, R. 2004 Bayesian-based pipe failure model. *Journal of Hydroinformatics* **6** (4), 259–264.

Wilson, D., Filion, Y. & Moore, I. 2017 State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water Journal* **14** (2), 173–184.

Yamijala, S., Guikema, S. D. & Brumbelow, K. 2009 Statistical models for the analysis of water distribution system pipe break data. *Reliability Engineering & System Safety* **94** (2), 282–293.

First received 14 July 2023; accepted in revised form 9 December 2023. Available online 19 December 2023