Water Supply

 Water Supply Vol 22 No 3, 3114 doi: 10.2166/ws.2021.418

Monitoring the health status of water mains using a scorecard modelling approach

Yuzhi Huang[a], Raufdeen Rameezdeen[a], Christopher W. K. Chow[a,b,*], Nima Gorjian[a,c], Yangyue Li[a], Zijun Liu[a] and Peiqing Ju[a]

[a] UniSA STEM (Science, Technology, Engineering and Mathematics), University of South Australia, Scarce Resources and Circular Economy (ScaRCE), Mawson Lakes, South Australia 5095, Australia
[b] UniSA STEM (Science, Technology, Engineering and Mathematics), University of South Australia, Future Industry Institute, Mawson Lakes, South Australia 5095, Australia
[c] South Australian Water Corporation, Adelaide, South Australia 5000, Australia
*Corresponding author. E-mail: christopher.chow@unisa.edu.au

CWKC, 0000-0001-5829-8944

ABSTRACT

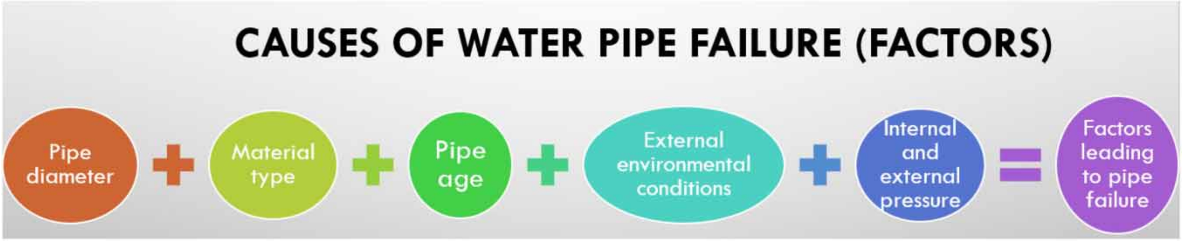
There has been considerable research into prediction of water mains failure, however, those models are very complex and fail to convey the message of the health status of an asset to the relevant stakeholders. The study developed a 'pipe health scorecard' based on historical failure data which could be used for operation, maintenance, refurbishment, or replacement decisions by a water utility. This scorecard model was developed by using 160,413 pipe-condition data sets from the South Australian Water Corporation over ten years. Measures such as the Kolmogorov–Smirnov (KS) statistic, Area Under the ROC Curve (AUC), and Population Stability Index (PSI) showed the model is strong enough to predict the health status of water mains. The study found the factors influencing water mains failure to be in the order of importance: length, material, age, location (road vs verge), diameter, and operating parameters. The development of a simple but reliable model for the assessment of the health status of water mains will have major benefits to the water utility with the ability to identify and potentially replace water pipes prior to failure. Additional benefits of flexible scheduling of maintenance and replacement programs would contribute to cost savings.

Key words: asset management, data and information, pipe failure, scorecard, water utilities

HIGHLIGHTS

- This study explored the use of a scorecard to prioritise pipe-replacement decisions.
- A scorecard model was developed using a large real pipe-condition data set extracted from a water utility's computerised asset management system.
- Factors influencing water mains failure have been identified in the order of importance: length, material, age, location (road vs verge), diameter, and operating parameters.

GRAPHICAL ABSTRACT



INTRODUCTION

The causes of pipe failure are complex and water utilities should not just base their judgement purely on the age of pipes to set up the replacement program. The average asset life of a regular water pipe is about 100 years. However, some pipes may be able to achieve a longer usable life, while others may need to be replaced sooner due to their condition. According to data from the South Australian Water Corporation (the water utility from where the data for this study was obtained), approximately 8% of water pipes have lasted more than 100 years, and around half have a life of more than 50 years (Hekmati *et al.* 2020). The aging of pipes is inevitable, but age is not always the only indicator of pipe failure. Many studies have found in addition to age, pipe diameter, material type, external environmental conditions, and internal and external pressure loads are related to water pipe failure (Christodoulou *et al.* 2010; Fares & Zayed 2010; Jafar *et al.* 2010; Kleiner & Rajani 2012; Laucelli *et al.* 2014; Aydogdu & Firat 2015; Kabir *et al.* 2016a; Farmani *et al.* 2017; Ward *et al.* 2017; Fu *et al.* 2019; Motiee & Ghasemnejad 2019; Sattar *et al.* 2019). Researchers have analysed the elements of water pipe failure and built failure prediction models suitable for different regions based on historical data (Rezaei *et al.* 2015; Dawood *et al.* 2020a). The water pipe fault prediction models can effectively predict the time when water pipes may fail in the future and alert water utilities to replace them in advance, to avoid a series of impacts.

While there are a number of past studies dealing with pipe failure prediction using many advanced methodologies, most of them fail to convey the message of the health status of a pipe to the relevant stakeholders, i.e., the maintenance engineer, asset manager etc., due to their sophistication. In addition, it is not cost-effective to use sensors to monitor underground pipes in real-time due to their extent and geographical spread considering a water utility such as the South Australian Water Corporation (SA Water) that covers an entire state. Therefore, an easy-to-understand yet comprehensive tool that could be used for routine applications in the water sector is needed. Practitioners need a reliable, useful and cost-effective but non-sophisticated tool for their day-to-day decision making. The purpose of this study is to develop a 'pipe health scorecard' based on historical failure data which could be used for maintenance, refurbishment, or replacement decisions by a water utility. The parameters which affect the failure of water pipes will be examined and used to develop a risk rating model (scorecard) as a decision support tool (Bequé *et al.* 2017). The model could provide a score for pipes on the basis of their characteristics such as age, diameter, materials etc.

The scorecard method is mainly used by banks to assess the risk of lending to a new customer. Banks usually maintain a large database of information of their current and past customers. Using those historical data, a model is built to identify who are 'good' customers and who are 'bad' using a number of parameters (for example age, income, education level, job status, marital status, number of dependents, and so on). When a new customer requests a bank loan, the bank will score the risk of that application using the scorecard model. It will then decide whether to approve the loan or not based on the probability of that customer being in the 'good' or 'bad' group. This good/bad status is the binary target variable Y, which the bank will relate to all information available at the scoring time about the new loan request (Yap *et al.* 2011).

This technique has been refined and used by banks for quite a long period. The goal of the scorecard method is to quantify this relationship as precisely as possible to assist credit decisions and the monitoring and management of borrowers (Wang *et al.* 2021). Banks score borrowers at loan application, as well as at regular intervals during the loan period to assess the risk to the bank. This study uses that method to score the health status of underground pipes. Based on a very large historical database, we can identify parameters that decide that pipes are of good health or that they are not. Asset managers can score a pipe and decide its health based on the scorecard model. Hence, this tool is quite different and novel from traditional tools used by asset managers. The major advantage is its ability to communicate the outcome without much sophistication. Therefore, the potential for this model to be used as a routine health-monitoring system in a water utility is high. This paper describes the steps to identify and evaluate the critical factors of pipe failure using a case study to demonstrate the use of the scorecard model.

MATERIALS AND METHODS

Study design

This study was conducted in South Australia using SA Water's event and condition monitoring data. SA Water is responsible for the supply of pipe-borne water for the entire state with a population of 1.7 million and almost 80% of it is concentrated in the state capital, Adelaide. This study used secondary data analysis, and extracted data from SA Water's computerised maintenance management system (i.e. Maximo) that contained pipe maintenance, repair, and replacement data from 2011 onwards. There were 160,413 sets of pipe data extracted for model development. A pipe in this study is defined as a homogeneous section of the network between junctions or valves depending on where it is in the network. The target variable was pipe condition as recorded in the Maximo database representing a binary: failure (1) or non-failure (0). Factors related to the pipe condition were determined based on reported literature, namely, age, material, diameter, length, and operating pressure, which have been identified as key parameters (Berardi *et al.* 2008). Researchers have reported that pipes installed under sidewalks have a lower risk of damage than pipes installed under roads, which have additional external pressure imposed by vehicular traffic (Debón *et al.* 2010). Hence, the location was also included as an additional variable affecting pipe failure while the depths of the pipes were considered similar (unless the topography or a special circumstance dictated otherwise, which was very rare) and was not included as a variable. A summary of the variables included in this study is shown below:

Input Variable (data description with condition of pipe as target variable)

- Age: Age of the pipe at the time of event or inspection (yrs, 0–25; 26–50; above 50)
- Material: 26 different types of pipe material and based on the scorecard results they were categorised into three groups as follows.
  Type 1 – ABS, SS, GRP, DIFB, MS, PC, CU, PVC, PVCU, PVCO, PE100
  Type 2 – DICL, PVCU, PESO, MSCL
  Type 3 – WDCS, GWI, MSCS, AC, CI, CICL, VICT, CICS, RC, ACOS
- Diameter: Pipe diameter (mm, 0–90; 91–150; 151–300; more than 300)
- Length: Pipe length (m, 0–10; 11–20; 21–60; more than 60)
- Operating pressure: Internal water pressure of the pipe (psi, 0–25; 26–45; 46–70; more than 70)
- Location: Location of the pipe (road; verge)

Model development and validation

The scorecard model is a risk rating model which is used predominantly in the banking sector where it can learn by utilising a customer's historical data together with peer group data to predict the probability of that customer displaying a defined behaviour in the future. Hence, it is used to determine 'good' and 'bad' borrowers at the time of loan approval (Abdou & Pointon 2011; Louzada *et al.* 2016). The model was built using the R programming language with the scorecard package from GitHub (Xie 2020). Subsequent to pre-treatment of data, two groups (files) in a random arrangement as 70% and 30% were created. The 70% of data was used as the training set and the remaining 30% of data for validation. This procedure was aimed at evaluating the performance of the model.

The scorecard model is a binary classification model, which means that the model can only understand that the pipe had a failure or not (1 or 0). The pipes without any failure in the system are defined as '0' and the pipes with failure are defined as '1'. The data set of the model was analysed by the scorecard method using following steps. (1) The data set was entered and processed by the model, and it was then separated into different groups ready for further use. (2) The variables of each data set were analysed, and the non-numerical variables were not considered in this step. The rest of the variables were analysed for Weight of Evidence (WOE) purposes and classified into different categories. WOE indicates the predictive power of an independent variable in relation to the dependent variable. It is a measure of the separation between 'failed' and 'non-failed' pipes in relation to the independent variables of the model as follows.

$$\mathrm{WOE} = \ln \frac{\%\ \mathit{of\ non\text{-}failures}}{\%\ \mathit{of\ failures}}$$

Finally, the scorecard of this model and its performance parameters were acquired. The model development process is shown in Figure 1. It should be noted that the performance is assessed at the 'pipe' level.

The remaining 30% of data set aside was used for model validation purposes. The scores of different pipes were calculated by substituting the scorecard into the test set. This score was used to determine whether the selected pipes had a failure record before. The predicted results were compared with the actual cases to validate the accuracy of the model. Three different types of accuracy were checked including overall accuracy, and the accuracy of predicting failed pipes and non-failed pipes. Comparison of these outputs would be a proof of the precision of this model.

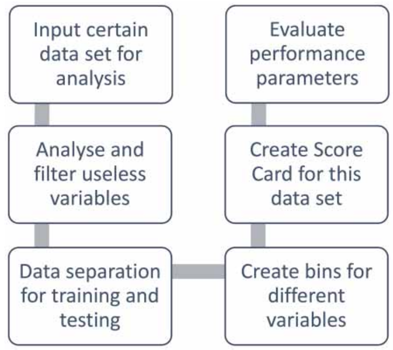


Figure 1 | Stages of the scorecard analysis.

RESULTS AND DISCUSSION

Model description

The descriptive statistics of the input variables used in the analysis are shown in Table 1. Of the 160,413 pipes, 24,242 (15.1%) failed at least once and 136,171 (84.9%) never failed in the record.

Strength of the model

The strength of the model is measured using several methods. The Kolmogorov–Smirnov statistic (KS) is a standard measure of model performance in scorecard analysis. The Kolmogorov–Smirnov statistic measures how far apart the distribution functions of the scores of the failed and non-failed pipes are. It calculates the difference, distance, or separation between the failed and non-failed pipes. The higher the value, the better the prediction of the model. As shown below, a KS value of 0.565 is considered to be adequate to discriminate between failed and non-failed pipes based on the independent variables included in the model (Zeng 2017). The other measure used in this study to evaluate the performance of the scorecard model is the Area Under the ROC Curve (AUC). AUC can be calculated geometrically, by the probability of a correct ranking of a non-failed and failed pair of pipes, and by the Wilcoxon Rank-Sum statistic (Zeng & Zeng 2019). The bigger the AUC value, the better the performance of the model, and hence, an AUC of 0.842 provides another proof that the model is robust enough for prediction. Finally, the Population Stability Index (PSI) is applied in order to investigate the differences in the distribution of the two categorised variables. It compares the predicted probability of the scoring data set to the training data set that was used to develop the model. The higher the value of the coefficient, the greater the statistical distance between the distributions. The following are recommended based on the PSI (Nehrebecka 2018):

- 0–0.1: no significant changes, no action required;
- 0.1–0.25: some changes found. It is recommended to check; and
- >0.25: significant changes found, it is recommended to re-establish the scorecard model.

With a PSI value of 0.0001, the model does not need any modifications.

Information value and predictive power of variables

The scorecard and the performance parameters of the model: Information Value (IV) is an indicator of correlation between a binary variable y and a nominal variable x and it shows the predictive power of the variables.

- Pipe age (yrs): 0.490
- Pipe material: 0.514
- Pipe diameter (mm): 0.085
- Pipe length (m): 1.536
- Operating pressure (psi): 0.039
- Location : 0.130

Note: KS=0.565; AUC=0.842; PSI=0.0001

Table 1 | Descriptive statistics for the variables in the analysis of the base model

Variable	Non-failed		Failed		
	N	%	*N*	%	c^2
Pipe age (y)					6,776.88**
≤25	38,334	28.2	1,239	5.1	
26–50	50,186	36.9	9,495	39.2	
>50	47,651	35.0	13,508	55.7	
Pipe material					7,008.57**
Synthetic plastic polymer (M1)	26,621	19.5	1,041	4.3	
DI and MS (M2)	24,720	18.2	1,434	5.9	
AC and CI (M3)	84,830	62.3	21,767	89.8	
Pipe diameter (mm)					1,461.14**
≤90	12,520	9.2	3,563	14.7	
91–150	89,815	66.0	16,702	68.9	
150–300	19,866	14.6	2,861	11.8	
>300	13,970	10.3	1,116	4.6	
Pipe length (m)					22,079.15**
≤10	56,554	41.5	734	3.0	
11–20	13,577	10.0	750	3.1	
21–60	22,963	16.9	3,376	13.9	
>60	43,077	31.6	19,382	80.0	
Operating pressure (psi)					661.43**
≤25	7,601	5.6	572	2.4	
26–45	29,081	21.4	4,619	19.1	
46–70	60,342	44.3	10,817	44.6	
>70	39,147	28.7	8,234	34.0	
Location					1,488.74**
Verge	11,554	8.5	348	1.4	
Road	124,617	91.5	23,894	98.6	

Note: $^{**}p<0.01$.

Information value is one of the most useful techniques to select important variables in a predictive model. It helps to rank variables on the basis of their importance. The IV is calculated using the following formula:

$$IV = \sum (\%\ of\ non\text{-}failures - \%\ of\ failures) \times WOE$$

IV ranks variables based on the importance and amount of information they carry. It should be used only during the creation of the model, as it analyses each feature individually (Barddal *et al.* 2020). The following are the acceptable ranges of a variable's predictive power (Yap *et al.* 2011):

- <0.02 Useless for prediction
- 0.02–0.1 Weak predictor
- 0.1–0.3 Medium predictor
- 0.3–0.5 Strong predictor
- >0.5 Very strong predictor

In general, variables having IV value less than 0.02 should be removed from the model (Yap *et al.* 2011). The smallest IV in the model is 0.039 for operating pressure. Hence, none of the variables were removed from the model. Results show that length, material, and age demonstrated strong predictive powers. The location had medium predictive power while the predictive powers of diameter and operating pressure were weak.

Model validation

The validation of the model was carried out with the 30% of data set aside. Once the scorecard was acquired, it was substituted with the test set for validation. If the score of a pipe is higher than the basepoint (513), it means that the status predicted for the pipe concerned is that it has never failed before. If the score is less than the basepoint, it indicates a failure. As which pipes have failed in the 30% test set is known, the model results can be compared with the real status of pipes to determine the accuracy of the model. Table 2 shows the outcome of the comparison. In Table 2, there are three different accuracies: prediction of non-failures; prediction of failures; and the overall accuracy. These three stand for the accuracy of predicting only the non-failed pipes, failed pipes, and a combination of them, respectively. The overall accuracy was 76.83% while the accuracy rates for non-failures and failures were 76.88% and 76.57%, respectively. The model provides close accuracy levels for prediction of failed and non-failed pipes, which is well accepted as a robust model.

The scorecard points as given in Table 3 show the impact of bins within each parameter; the higher the scorecard points, the lower the possibility for the pipe to fail. In other words, when the length, age and operating pressure of pipe increase, the possibility of failure increases. By contrast, when the diameter increases, the possibility of failure decreases. Regarding materials, pipes made of synthetic plastic polymer showed the lowest failure possibility while those of AC and CI showed the highest out of the three categories. Results on the location of the pipes showed pipes buried under the road have a higher possibility of failure than those located under the verge. These results supported earlier studies which included similar parameters in their analysis of factors affecting water mains failure. They found that the factors influencing water mains failure include age, material, diameter and length (Fares & Zayed 2010; Christodoulou 2011; Aydogdu & Firat 2015). According to Ward *et al.* (2017), material and age are major factors affecting pipe performance and pipes of the same age and material would normally behave in a similar way.

The relationship between scorecard values and pipe material as shown in Table 3 and Figure 2(a) clearly discriminate those of synthetic plastic polymer, DI, and MS versus AC and CI with the former having positive scorecard values, and relatively higher mean and median values compared with the latter. Past research has proven that synthetic plastic polymer pipes fail in an unexpected brittle way and the rate of failure is much less compared with other materials (Christodoulou 2011). These pipes have been used increasingly in recent times because of their durability and cost-effectiveness (Dawood *et al.* 2020b). According to Kabir *et al.* (2016b), the survival rate of DI pipes is higher than that of CI pipes. The deterioration mechanism of DI and MS pipes is mainly attributed to electrochemical corrosion, which generates a small hole initially and then with time the hole keeps on growing until the pipe bursts (Kabir *et al.* 2015; Kutyłowska 2015). CI pipe failure is mainly a circumferential crack separating the pipe into two pieces. AC pipes are subject to deterioration due to multiple chemical processes and their failure could be classified into five types: longitudinal, circumferential, holes, joints, and others (Hu & Hubble 2007; Davis *et al.* 2008).

The study also found a significant change of the scorecard value between pipe ages up to 25 years of age and beyond with values moving from positive to negative as shown in Table 3. Figure 2(b) also illustrates a clear distinction between these two groups with the mean and median values dropping significantly between younger and older pipes. Past research provides clear evidence for this observation as a large number of studies have made systematic observations of pipe deterioration with age (Harvey *et al.* 2014; Chik *et al.* 2017; Farmani *et al.* 2017; Dawood *et al.* 2020a). According to Christodoulou (2011), aged pipes must be replaced at approximately 30 years.

Table 2 | Accuracy levels of the scorecard model

	Non-failed pipes	Failed pipes	Total
Number of correct predictions	44,818	8,005	52,823
Number of wrong predictions	13,481	2,449	15,930
Accuracy level	76.88%	76.57%	76.83%

Table 3 | Scorecard and performance parameters of the model (basepoint = 513)

	Scorecard points
Pipe Age (yrs)	
≤25	83
26–50	−1
>50	−22
Pipe Material	
Synthetic plastic polymer	53
DI and MS	33
AC and CI	−10
Pipe Diameter (mm)	
≤90	−16
91–150	−4
151–300	4
>300	5
Pipe Length (m)	
≤10	184
11–20	83
21–60	14
>60	−65
Operating Pressure (psi)	
≤25	25
26–45	7
46–70	0
>70	−10
Location	
Verge	47
Road	−2

The current study found pipe length to be the most influential variable of the model. It showed pipe lengths less than 10 m recording a very high scorecard value compared with the other three categories. In addition, those of more than 60 m of length scored a negative value, which had a clear distinction from the other three categories. Figure 2(c) confirms this observation as the mean and median values follow the same pattern. According to Aydogdu & Firat (2015), pipe length is a significant factor in the prediction of water main failure. Part of Adelaide, specifically areas closer to the foothills of the Mount Lofty Ranges and Adelaide Hills, is notorious for its clay soil which impacts longer pipes more than shorter ones when they shrink and expand due to moisture movement (Gallage *et al.* 2012).

The results had a clear distinction between pipes buried under the road (negative scorecard value) versus those under the verge (positive scorecard value). The mean and median values also showed clear differences as given in Figure 2(d). Past researchers also found heavy traffic volumes on the road affecting the water mains underneath (Fares & Zayed 2010; Christodoulou 2011). The results also showed that when the diameter of a pipe increases, the failure probability reduces. Pipes up to 150 mm diameter had negative scorecard values while the mean and medians as shown in Figure 2(e) discriminated pipes up to 90 mm diameter from the rest of the categories. In the study of Aydogdu & Firat (2015), pipe diameter was found to affect the failure rate with small-diameter pipes (less than 250 mm) being more exposed to failure risk than large-diameter pipes. Therefore, small-diameter pipes should be replaced more frequently than medium- or large-diameter pipes (Fares & Zayed 2010; Christodoulou 2011). Though the study had a very weak predictive power for operating pressure, operation of pipes with less than 25 psi seems to be discriminated strongly from those above that value according to Figure 2(f). In

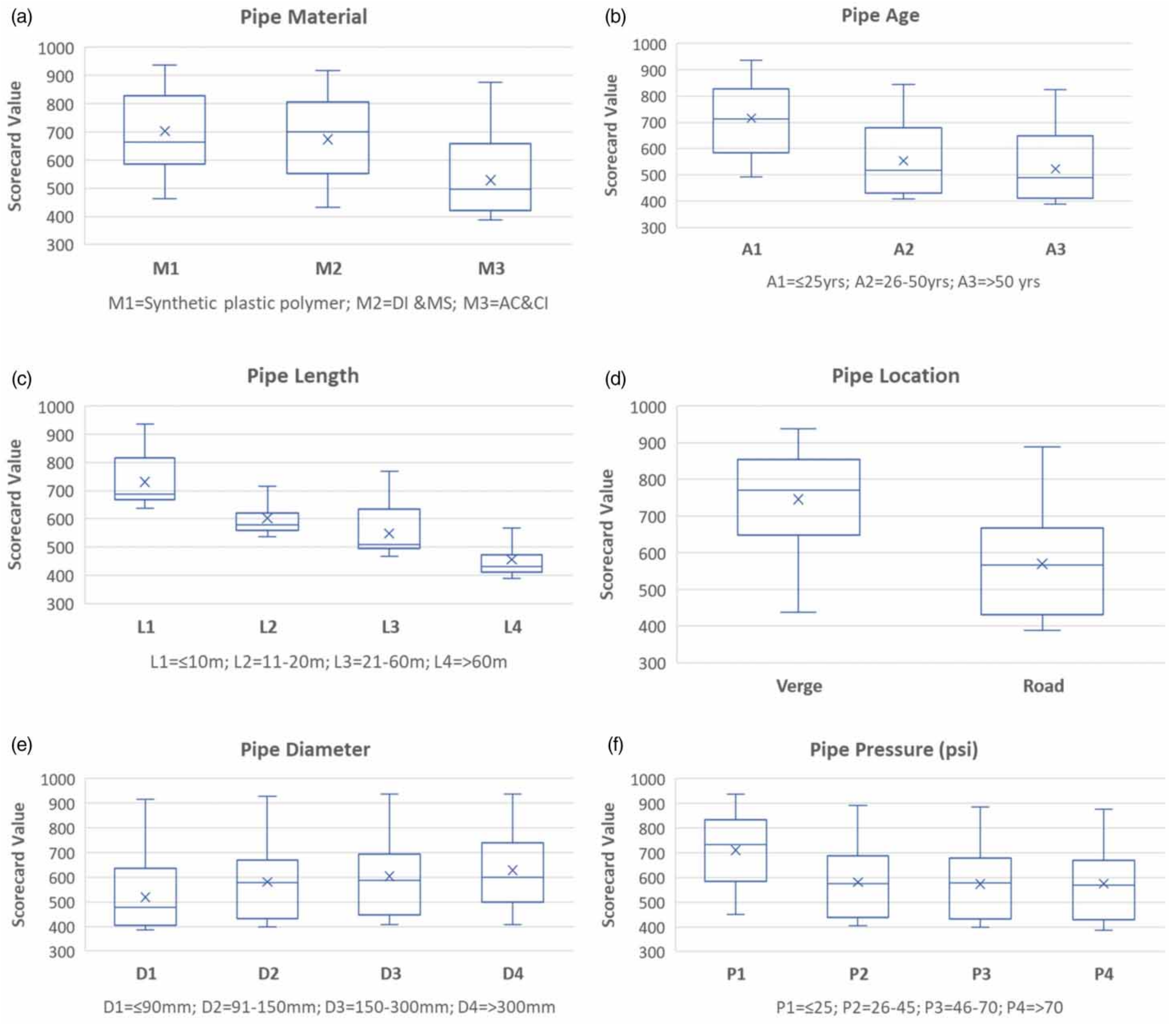


Figure 2 | Boxplots showing the relationship between input variables and scorecard values: (a) pipe material, (b) pipe age, (c) pipe length, (d) pipe location, (e) pipe diameter and (f) pipe pressure.

addition, those operating above 70 psi recorded a negative scorecard value. These results were confirmed by the studies of Fares & Zayed (2010) and Christodoulou (2011) where they found an increase in water pressure to increase the failure probability of water mains.

The results of this study provide water utilities with a practical and simple tool to check the health status of their current assets. The score of a pipe is calculated by the equation below:

$$Score = Points\ of\ (Material + Diameter + Length + Location + Pressure + Age) + Basepoint$$

If the score is higher than the basepoint (513), the probability of that pipe failing is less, and the higher the score, the smaller the chance of failure. By contrast, if the score is less than the basepoint, it indicates a possibility of failure in the future, and as above, the lower the score, the higher the chance of failure. Two example calculations, for pipes bearing Asset IDs 3509073 and 3786596, taken from the SA Water database, are shown in Table 4. When the points were substituted for each parameter of these pipes, the total scorecard values were higher than the basepoint scope of 513. This gives the asset manager an

Table 4 | Example calculations (basepoint=513)

Asset ID	Material	Diameter	Length	Location	Op. pressure	Age	Scorecard points
3509073	AC −10	100 mm −4	1.29 m 184	Road −2	69.92 0	54 yrs −22	659
3786596	CI −10	150 mm −4	28.12 m 14	Verge 47	53.65 0	52 yrs −22	538

indication that the pipes concerned are of relatively good health. In addition, it shows the pipe bearing Asset ID 3509073 is relatively healthier than 3786596.

CONCLUSIONS

There has been considerable research into failure prediction modelling of water mains. However, the existing models are very complex and unhelpful when it comes to practice because they do not convey the necessary message to the asset manager. The development of a simple but accurate model which could be used for the assessment of the health of water mains will have major economic benefits to the water utility with the ability to identify and potentially replace water pipes prior to failure. Additional benefits in the form of flexibility of scheduling of maintenance and replacement of such critical pipes would have large savings for a water utility. This in turn will reduce inconvenience to the neighbourhood and reduction of traffic interruptions and delays as opposed to current situations brought about from unplanned failures of water mains. Ideally, a failure prediction model will factor in as many variables as possible to predict the health status of the existing network using failure records.

To achieve a simple yet accurate and comprehensive model of water mains' health assessment, this study used the scorecard modelling technique which is predominantly used by the banking sector. This study used 160,413 data sets of pipe conditions since 2011 extracted from the South Australian Water Corporation's computerised maintenance management system for the development of the model. Measures such as the Kolmogorov–Smirnov statistic (KS), Area Under the ROC Curve (AUC), and Population Stability Index (PSI) showed the model is strong enough for predicting the health status of water pipes. The following are the main findings from the analysis of the above data:

- Parameters such as pipe length, material and age had high predictive power in the model compared with other variables.
- Location of the pipe, buried under the road versus verge, had medium predictive power.
- Parameters such as diameter and operating pressure had weak predictive powers, even though they need not be removed from the model.
- Performance of the synthetic plastic polymer pipes was better than DI and MS and they in turn recorded better performance than AC and CI.
- When the age, length and the operating pressure of the pipe increases, its probability of failure increases.
- The failure probability of pipes buried under the road is much higher than those under the verge.
- When the diameter of the pipe increases, its failure probability reduces.

While the model has a 76% accuracy level of prediction, which is one of the major limitations, the 'Scorecard Value' could help the asset managers to establish the health status of a pipe and make decisions on replacement as well as impending repairs. Hence, it would help in the overall financial planning in a water utility. While there are many opportunities to increase the accuracy level of the model by having additional parameters and longer-term data, the advantage of a scorecard value is its simplicity. It conveys a clear and uncomplicated message to the decision maker unlike other complicated models (see Dawood *et al.* 2020a for a comprehensive review). Scorecard modelling is extensively used in the banking industry to weed out 'good' versus 'bad' investors at the time of loan application. However, this concept could be used by the water industry for asset management purposes. This paper contributes to the extant literature by highlighting the possibility of using scorecard modelling through empirical data obtained from a water utility in Australia.

ACKNOWLEDGEMENTS

The research conducted was based on the relevant data provided by the local water utility, SA Water. The research team would like to thank the SA Water team for providing data and valuable inputs, especially Mrs Amanda Mussared.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors and contributors.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Abdou, H. A. & Pointon, J. 2011 Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management* **18** (2–3), 59–88.

Aydogdu, M. & Firat, M. 2015 Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM methods. *Water Resources Management* **29** (5), 1575–1590.

Barddal, J. P., Loezer, L., Enembreck, F. & Lanzuolo, R. 2020 Lessons learned from data stream classification applied to credit scoring. *Expert Systems with Applications* **162**, 113899.

Bequé, A., Coussement, K., Gayler, R. & Lessmann, S. 2017 Approaches for credit scorecard calibration: an empirical analysis. *Knowledge-Based Systems* **134**, 213–227.

Berardi, L., Giustolisi, O., Kapelan, Z. & Savic, D. A. 2008 Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics* **10** (2), 113–126.

Chik, L., Albrecht, D. & Kodikara, J. 2017 Estimation of the short-term probability of failure in water mains. *Journal of Water Resources Planning and Management* **143** (2), 04016075.

Christodoulou, S. E. 2011 Water network assessment and reliability analysis by use of survival analysis. *Water Resources Management* **25** (4), 1229–1238.

Christodoulou, S., Agathokleous, A., Charalambous, B. & Adamou, A. 2010 Proactive risk-based integrity assessment of water distribution networks. *Water Resources Management* **24** (13), 3715–3730.

Davis, P., De Silva, D., Marlow, D., Moglia, M., Gould, S. & Burn, S. 2008 Failure prediction and optimal scheduling of replacements in asbestos cement water pipes. *Journal of Water Supply: Research and Technology – AQUA* **57** (4), 239–252.

Dawood, T., Elwakil, E., Novoa, H. M. & Gárate Delgado, J. F. 2020a Water pipe failure prediction and risk models: state-of-the-art review. *Canadian Journal of Civil Engineering* **47** (10), 1117–1127.

Dawood, T., Elwakil, E., Novoa, H. M. & Delgado, J. F. G. 2020b Pressure data-driven model for failure prediction of PVC pipelines. *Engineering Failure Analysis* **116**, 104769.

Debón, A., Carrión, 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.

Fares, H. & Zayed, T. 2010 Hierarchical fuzzy expert system for risk of failure of water mains. *Journal of Pipeline Systems Engineering and Practice* **1** (1), 53–62.

Farmani, R., Kakoudakis, K., Behzadian, K. & Butler, D. 2017 Pipe failure prediction in water distribution systems considering static and dynamic factors. *Procedia Engineering* **186**, 117–126.

Fu, G., Yang, W., Li, C. & Shi, W. 2019 Reliability analysis of corrosion affected underground steel pipes considering multiple failure modes and their stochastic correlations. *Tunnelling and Underground Space Technology* **87**, 56–63.

Gallage, C. P. K., Chan, D. & Kodikara, J. 2012 Response of a plastic pipe buried in expansive clay. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering* **165** (1), 45–57.

Harvey, R., McBean, E. A. & Gharabaghi, B. 2014 Predicting the timing of water main failure using artificial neural networks. *Journal of Water Resources Planning and Management* **140** (4), 425–434.

Hekmati, N., Rahman, M. M., Gorjian, N., Rameezdeen, R. & Chow, C. W. K. 2020 Relationship between environmental factors and water pipe failure: an open access data study. *SN Applied Sciences* **2**, 1806.

Hu, Y. & Hubble, D. W. 2007 Factors contributing to the failure of asbestos cement water mains. *Canadian Journal of Civil Engineering* **34** (5), 608–621.

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.

Kabir, G., Tesfamariam, S. & Sadiq, R. 2015 Predicting water main failures using Bayesian model averaging and survival modelling approach. *Reliability Engineering & System Safety* **142**, 498–514.

Kabir, G., Tesfamariam, S., Loeppky, J. & Sadiq, R. 2016a Predicting water main failures: a Bayesian model updating approach. *Knowledge-Based Systems* **110**, 144–156.

Kabir, G., Tesfamariam, S. & Sadiq, R. 2016b Bayesian model averaging for the prediction of water main failure for small to large Canadian municipalities. *Canadian Journal of Civil Engineering* **43** (3), 233–240.

Kleiner, Y. & Rajani, B. 2012 Comparison of four models to rank failure likelihood of individual pipes. *Journal of Hydroinformatics* **14** (3), 659–681.

Kutyłowska, M. 2015 Neural network approach for failure rate prediction. *Engineering Failure Analysis* **47**, 41–48.

Laucelli, D., Rajani, B., Kleiner, Y. & Giustolisi, O. 2014 Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. *Journal of Hydroinformatics* **16** (4), 743–757.

Louzada, F., Ara, A. & Fernandes, G. B. 2016 Classification methods applied to credit scoring: systematic review and overall comparison. *Surveys in Operations Research and Management Science* **21** (2), 117–134.

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.

Nehrebecka, N. 2018 Predicting the default risk of companies. Comparison of credit scoring models: LOGIT vs Support Vector Machines. *Econometrics* **22** (2), 54–73.

Rezaei, R., Ryan, B. & Stoianov, I. 2015 Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. *Procedia Engineering* **119**, 253–262.

Sattar, A. M. A., Ertuğrul, Ö. F., Gharabaghi, B., McBean, E. A. & Cao, J. 2019 Extreme learning machine model for water network management. *Neural Computing and Applications* **31** (1), 157–169.

Wang, C., Fang, K., Zheng, C., Xu, H. & Li, Z. 2021 Credit scoring of micro and small entrepreneurial firms in China. *International Entrepreneurship and Management Journal* **17** (1), 29–43.

Ward, B., Selby, A., Gee, S. & Savic, D. 2017 Deterioration modelling of small-diameter water pipes under limited data availability. *Urban Water Journal* **14** (7), 743–749.

Xie, S. 2020 Scorecard. Github. Available from: https://github.com/ShichenXie/scorecard (last accessed 21 November 2020).

Yap, B. W., Ong, S. H. & Husain, N. H. M. 2011 Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications* **38** (10), 13274–13283.

Zeng, G. 2017 A comparison study of computational methods of Kolmogorov–Smirnov statistic in credit scoring. *Communications in Statistics – Simulation and Computation* **46** (10), 7744–7760.

Zeng, G. & Zeng, E. 2019 On the three-way equivalence of AUC in credit scoring with tied scores. *Communications in Statistics – Theory and Methods* **48** (7), 1635–1650.

First received 10 September 2021; accepted in revised form 19 November 2021. Available online 7 December 2021