Reliability Engineering and System Safety 186 (2019) 24–36

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system

Kayu Tang, David J. Parsons*, Simon Jude

School of Water, Energy and Environment, Cranfield University, College Road, Cranfield MK43 0AL, UK

ARTICLE INFO

Keywords:
Bayesian networks
Asset management
Reliability
Infrastructure
Statistics
Water

ABSTRACT

The reliability of the water distribution system is critical to maintaining a secure supply for the population, industry and agriculture, so there is a need for proactive maintenance to help reduce water loss and down times. Bayesian networks are one approach to modelling the complexity of water mains, to assist water utility companies in planning maintenance. This paper compares and analyses how accurately the Bayesian network structure can be derived given a large and highly variable dataset. Method one involved using automated learning algorithms to build the Bayesian network, while method two involved a guided method using a combination of historic failure data, prior knowledge and pre-modelling data exploration of the water mains. By understanding common failure types (circumferential, longitudinal, pinhole and joint), the guided learning Bayesian Network was able to capture the interactions of the surrounding soil environment with the physical properties of pipes. The Bayesian network built using data exploration and literature was able to achieve an overall accuracy of 81.2% when predicting the specific type of water mains failure compared to the 84.4% for the automated method. The slightly greater accuracy from the automated method was traded for a sparser Bayes net where the interpretation of the interactions between the variables was clearer and more meaningful.

1. Introduction

The robustness, reliability and resiliency of networked infrastructure are vital to the economy, security and wellbeing of a country. The water distribution system is a networked infrastructure, where the pipes constitute the arteries of the network that helps supply water for multiple purposes. In England and Wales in 2017–18 the average number of bursts across all supply companies was about 150 per 1000 km of mains [1], creating a substantial maintenance workload and cost for the companies. Investing substantial amounts of money into repairing and rehabilitating aging underground water supply assets is an ongoing process; however, this does not immediately solve or mitigate disruptions from unplanned maintenance [2]. Down times from unplanned maintenance can severely inconvenience consumers, while unrepaired leaks cause substantial losses, adding to the volume and cost of water that must be abstracted and treated. Thus there are both financial and social motivations to improve the reliability of the distribution system. In the UK, the industry regulator (Ofwat) sets standards for leakage, with the power to impose penalties for non-compliance.

A solution would be for water companies to undertake a more proactive form of maintenance, which would not only include the monitoring of underground assets, but also correction of failure root causes rather than the immediate symptoms [3]. Proactive maintenance would allow the water distribution system to be more robust when faced with different hazards and extreme weather [4]; however, proactive maintenance requires a sound and extensive risk and reliability assessment. This is because to find root causes in failures there is a need to not only identify the risk factors in the water distribution system, but also understand the dependencies between them [5].

The complexity of the water distribution system calls for advanced methods to model it. There has been previous research on using Bayesian networks to model the water distribution network [6–9], but these have not considered newer materials, such as plastic pipes. The aim of this research was to compare two approaches to building Bayesian networks to model the likelihood of failures for a wide range of pipe materials: an automated data-driven method and a guided method using prior knowledge and conventional data exploration. To be useful and trusted in practice, the model should give a structure that is comprehensible to the users and give results that agree with the data.

2. Water distribution system

In this study, pipe failures were modelled as circumferential,

* Corresponding author.
E-mail address: d.parsons@cranfield.ac.uk (D.J. Parsons).

https://doi.org/10.1016/j.ress.2019.02.001
Received 10 August 2017; Received in revised form 18 January 2019; Accepted 1 February 2019
Available online 06 February 2019

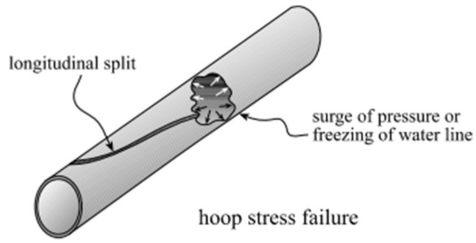


Fig. 1. Longitudinal split and blowout hold of pipe (from [11]).

longitudinal, pinhole and joint failures. By representing how these common pipe failures occurred, the Bayesian network should able to link the surrounding soil environment interactions with the different physical properties of pipe.

Metallic pipe failures from crack formations can occur circumferentially or longitudinally and can be instigated by corrosion. Corrosion such as graphitisation is a form of selective corrosion/leaching of the pipe material that can occur internally and externally, with a common occurrence seen in cast iron and ductile iron pipes [10]. A longitudinal split or blowout hole on metallic water pipe (Fig. 1) is mainly due to the combination of reduction in wall thickness from corrosion and internal water pressure, where the weakened pipe section can no longer withstand the forces from the water pressure [11].

Circumferential fracture of the pipe involves the movement of the surrounding soil near the failure area; common associations with this failure are poor bedding and frost penetration of the soil (Fig. 2). There is increasing evidence of corrosion occurring in pipes that have failed due to circumferential fracture [12]. The occurrence of corrosion could have been due to leaks created by cracks developed on the pipe. Water leaking into the surrounding environment could gradually erode the soil supporting the pipe, resulting in a circumferential fracture.

Different materials have different mechanical properties (common properties considered are strength, ductility, hardness, impact resistance, and fracture toughness), which directly affects the structural integrity of the pipe. Specific materials are used for large diameter pipes (e.g. transmission pipes) and small diameter pipes (e.g. water supply pipes) as the water volume, water pressure, water velocity, and mechanical loading differs greatly for different purposes. For example, since 1982 ductile iron has replaced grey cast iron as the material of choice for transmission pipes, as the material exhibits the same desired mechanical properties such as strength, which can manage efficiently the rigors of internal and external loadings (e.g. high water pressure and harsh terrains) and an expected long service life, but is more ductile and less prone to graphitisation [11].

Even though there are more effective protective layers now available for metallic pipes, the ease of installation and the corrosion resistance of plastic pipes have resulted in increased use of medium and high density polyethylene and polyvinyl chloride (MDPE, HDPE and PVC) for new pipes [10]. PVC pipes were first produced in the 1950s and gained popularity as material for water supply pipes in 1960 [13]. However, before installation PVC pipes used in the water distribution system should be handled with care, as performance can be severely reduced if they are exposed to direct sunlight before installation. PVC is a better choice for small diameter pipes compared to metallic pipes as

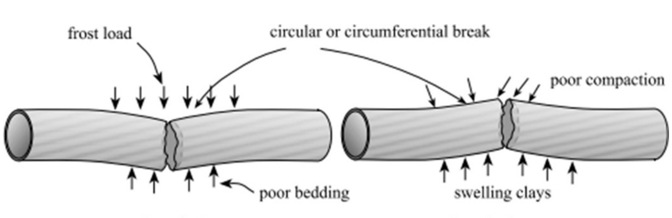


Fig. 2. Circumferential fracture of pipes (from [11]).

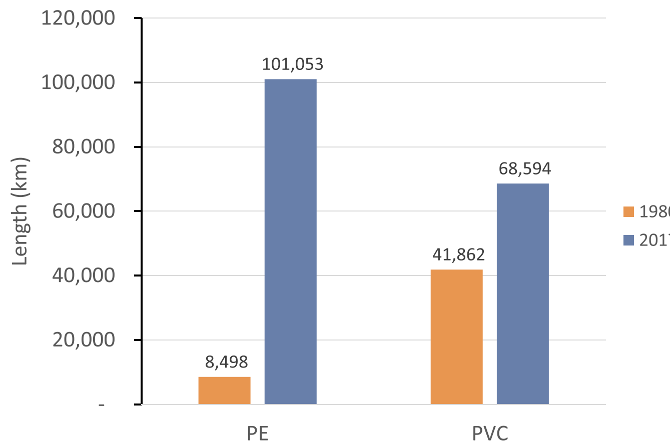


Fig. 3. PE and PVC total pipe length installed in UK in 1980 and 2017.

the mechanical characteristics allow for a smaller wall thickness whilst under the same nominal water pressure [14]. PVC also has a more favourable modulus of elasticity (ability for material to return to its initial state after deformation); therefore, PVC pipes have a lower risk of bursting compared to metallic pipes when under similar bending motion. From 1980 to 2017 PE pipe had an 11-fold increase from 8498 km to 101,053 km and PVC pipe from 41,862 km to 68,594 km (see Fig. 3, based on the UK Water Industry Research database described in Section 4).

3. Bayesian networks

Bayesian networks are part of a branch of statistical tools called advanced graphical models that can describe probabilistic relationships between variables [15]. A Bayesian network consists of two parts: a qualitative part in the form of a directed graph, and a quantitative part, in the form of conditional probability tables [16]. A directed graph consists of directed edges and nodes, where the variables in the model are represented by the nodes and the directed edges between the nodes indicate informational or causal dependencies among the variables [17]. Within this description, it will be assumed that all the variables have a finite number of possible values, although generalisations to continuous variables are possible. In a Bayesian network an important restriction is that the directed graph must be acyclic, that is the edges must not create loops or cycles within the network [18].

It is often assumed that the directed edges in Bayesian network represent causal relationships; however, probability theory is not intrinsically able to express causality, so edge directions are not necessarily indicative of causal effects [19]. However, it can also be argued that there is a case for the usage of the term 'causal' for edges in manually constructed Bayesian networks, as they are usually designed to represent the prior understanding of the causal structure. These manually constructed Bayesian networks are usually fairly sparse and their interpretation is clear and meaningful: "It seems that if conditional independence judgements are by products of stored causal relationships, then tapping and representing those relationships directly would be a more natural and reliable way of expressing what we know believe about this world. This indeed is the philosophy behind causal Bayesian networks" [20].

When two nodes are connected by a directed edge, the node at the tail is called the 'parent' and the node at the head is called the 'child', so the arrow points from the parent to the child (in Fig. 4, A and B are parents of C). If a node does not have any parents (nodes A and B in Fig. 4), the node will contain a marginal probability table that gives the probability of each of the possible states. Each node with parents contains a conditional probability table, giving the conditional probability of each of its states for every combination of states of its parent(s).

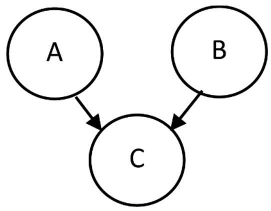


Fig. 4. Directed acyclic graph of events A, B and C.

These tables then allow the joint probability distribution of all the nodes to be calculated. Bayesian networks are attractive for probabilistic reasoning because their structure allows them to be decomposed for efficient calculation of the joint distribution in typically sparse networks in practice [21], although it is NP-hard in general [22]. Bayesian inference can be used during the construction of the model to 'learn' the conditional probabilities by entering evidence into multiple nodes, or when the model is being used to update the distributions of other nodes when evidence is entered at one or more nodes.

The flexible nature of Bayesian networks, due to their non-parametric nature and ability to deal with relationships between variables, may be especially well suited for environmental applications and risk assessments such as pipe failure in water distribution systems. Bayesian networks can also carry out probabilistic inference easily and efficiently for each specific failure outcome by considering the variables involved, rather than updating the whole model [21]. Another significant benefit of Bayesian networks is that they allow for the conditional dependencies or causal interactions between variables to be visualised. This provides an intuitive way of observing the relationships allowing stakeholders to make informed decisions in response to different hazard scenarios.

Identifying the structure of the Bayesian network is an extremely important first step in this method, comparable to model selection in a conventional statistical model. The second step, estimation of the probability tables, is equivalent to parameter estimation in statistical model. The initial structure learning process can be performed by using structure-learning algorithms that use optimisation methods to attempt to identify the relationships from the data to maximise likelihood or minimise measures such as the Bayesian Information Criterion, or can be guided manually using domain knowledge. The estimation of the probability tables is usually performed by statistical inference and optimisation, although conditional probabilities are sometimes estimated directly from data when the network is being constructed manually.

4. Methods

The water distribution system in the United Kingdom was selected as a case study for modelling in the Bayesian network. The pipe failure data and national pipe database were kindly provided by UK Water Industry Research (UKWIR) and the soil corrosivity and shrinkage database by Cranfield University's Land Information System (LandIS). The UKWIR pipe database includes the following pipe properties: whether a pipe was lined, pipe material, pipe age, pipe length and pipe diameter. Two methods were tested to derive the Bayesian network structure with the given data: an automated learning method and a guided method utilising literature and expert knowledge.

4.1. Data preparation

The UKWIR pipe failure and national pipe databases were cleaned before they were used in the Bayesian network learning process using statistical software R [23]. Data cleaning ranged from searching for non-coherent data such as failure dates occurring before installation dates to identifying non-unique pipe IDs and entries with missing data (e.g. missing pipe diameters, missing pipe materials etc.). The coordinates of the pipe failures found in the failure data were analysed against their start and end coordinates in the national pipe data to ensure that the data matched. There was an effort to keep as much of the data as possible, as excessive data removal of data would lead to a loss of information. Proxies for the missing data were used in some cases and where possible were provided by UKWIR. For example, if the installation date for grey spun cast iron pipe was missing, the year 1950 was used, as that was the year when this type of manufacturing technique was last used in a water mains capacity. When grouping information on whether pipes were lined or not, slip lining and close fit methods were reclassified as new pipes in the UKWIR data with a new installation year, so they were removed from the general pipe lining population.

Although some of the variables, such as pipe diameter, length and age were theoretically continuous, their distributions could not be described by standard parametric distributions, and the parameter learning algorithm required discrete distributions, so the values were grouped into discrete intervals. Water pipes with similar functions tend to share characteristics. For example, communication pipes, which deliver water to houses from water mains, are generally short and small in diameter, whereas, trunk mains are generally long and large in diameter. As a result, many of the values fell naturally into groups related to the functions of the pipes, within which the variability in the data was small and often localised around one value. Pipe diameter (mm) took discrete values, but these were too numerous to use and contained clusters of similar values (for example older pipes converted from British customary units and newer ones in metric units) for functionally equivalent pipes, which were grouped into [0–50), [50–200), [200–400), [400–600) and [600 +]. Pipe length (m) was grouped into [0–25), [25–100), [100–200), [200–500) and [500 +].

Pipe age (years) was also divided into discrete ranges of [0], [1–20), [20–40), [40–60), [60–80) and [80 +). These not only reflected the periods when different materials were introduced, but also acted as a performance measure proxy similar to Kabir et al. [7], where each increasing discretised level represented a higher level of potential failure (from very low probability to very high probability) in relation to the aging of the material. The pipe materials were grouped into asbestos cement, ductile iron, grey cast iron, steel, glass reinforced plastic, polyethylene, and polyvinyl chloride. Decommissioned pipe data with no prior failures was combined with failure data to see how well the Bayesian network would perform after preventive maintenance actions.

The final dataset was then partitioned into three sets: training (60%), testing (20%) and validation (20%). More failure data was placed into the training set, as it was expected that the model would learn better about the pipe failure characteristics. Each set was then used in specific phases of the Bayesian network building, this process was done to test for generalisation error in supervised machine learning. The data was partitioned using the function *createDataPartition* located in the R package *caret* [24]. The soil data was mapped onto the UKWIR failure and national pipe database with Geographic Information System (GIS) program ArcMap [25] using a spatial join.

4.2. Automated structure learning of Bayesian network

The automated learning method used a hill-climbing algorithm from the bnlearn [19,26] R package. The hill-climbing method is a score-based technique that starts with an empty network structure of all variables, then proceeds by adding, removing and reversing edges between nodes to maximise the goodness of fit of the model. The score for the goodness of fit in bnlearn utilised the log-*likelihood loss*, which is the negated expected log likelihood; hence, the lower the score the better the fit [27,28]. The structure of the Bayesian network was final when the score could no longer be improved.

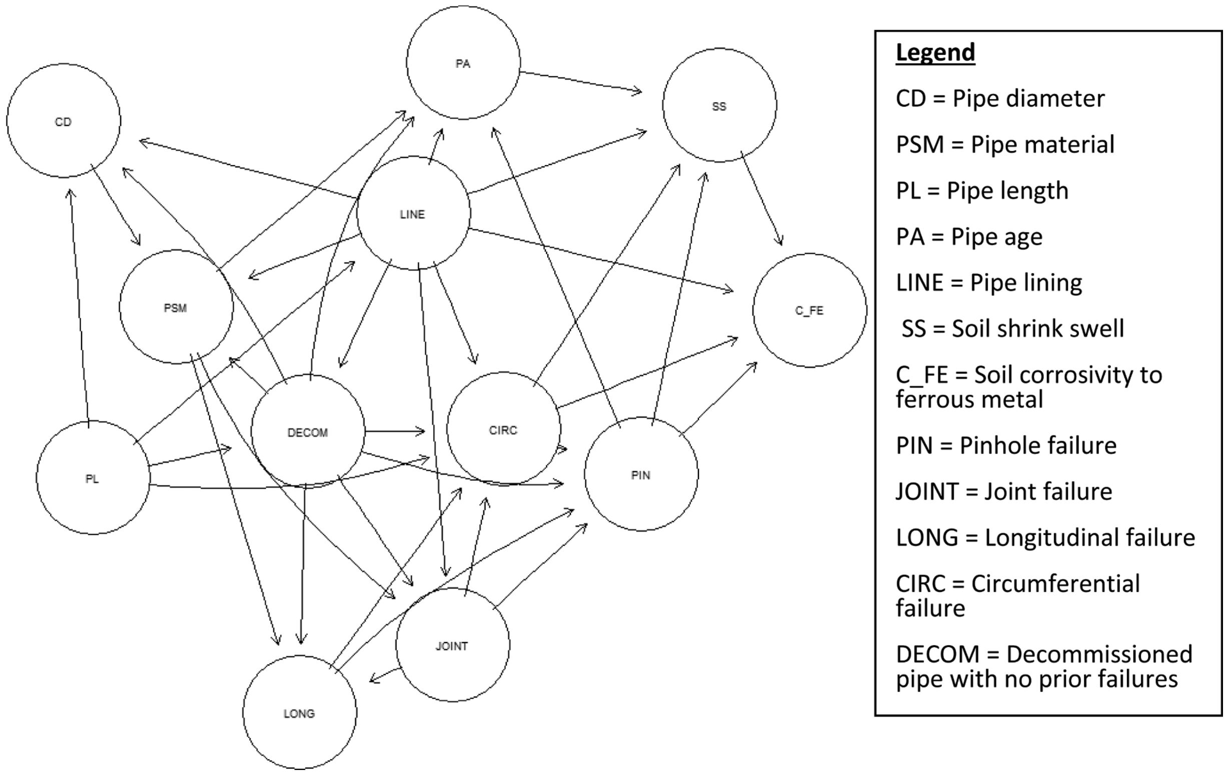


Fig. 5. Bayesian network constructed using Hill-climbing algorithm from R bnlearn package.

4.3. Data and knowledge guided structure learning of Bayesian network

4.3.1. Literature review

An extensive literature review was conducted to understand not only the risk factors involved with underground water distribution system asset failure, but also to identify the dependencies between them. The database from LandIS allows for surrounding soil characteristics to be taken into account in the form of corrosivity and shrink swell, thus along with the failure pipe data a Bayesian network can be used to model the different pipe failure types.

4.3.2. Data exploration

In the guided method, the data collected and knowledge used to build the structure of the Bayesian network worked in parallel. An approach similar to the one introduced by Babovic [29] was utilised to incorporate domain knowledge, where the Bayesian Network was constrained before combining with raw data. The domain knowledge consisted of information from experts and prior information available from historical data. The data exploration and pipe failure knowledge allowed for a more intuitively reasonable conceptual Bayesian network to be developed, by constraining the interactions between variables and imposing well-researched and known causal interactions that lead to pipe failures.

The statistical analysis of the data explored the relationships between the variables and the causal links found from the literature. The focus of the data and knowledge combination method was to obtain a structure of the Bayesian network that held true from an engineering point of view (literature knowledge) and real-life failure occurrences (data collected).

The exploration stage was carried out with visualisation of the data using boxplots and traditional multiple variable comparative statistical techniques. One-way analysis of variance (ANOVA) was used to test whether potential causal variables had significant effects on the pipe age at failure. If so, Tukey's honest significant difference test was used to identify which pairs of values of the variables were associated with significant differences. While it is uncommon to combine traditional statistical analysis with Bayesian network modelling, in this exploratory phase the decision was made to use the familiar classical approach rather than a Bayesian alternative.

4.4. Parameter learning

The structure of the Bayesian network was imported into GeNie, a general purpose Bayesian network commercial software [30] for the parameter learning stage. Likelihood maximisation with randomized initial values for the parameters was used, so that the process could be repeated from different starting points to avoid local minima. The process was completed when the expectation maximisation algorithm converted; that is, when the negative log likelihood had been minimised.

4.5. Validation

After learning the parameter values, the goodness-of-fit of the Bayesian network was validated using the receiver operating characteristic curve and the confusion matrix against the validation dataset. The receiver operating characteristic curve was used to express the quality and accuracy of the Bayesian network independent of the classification decision. It was created by plotting the true positive rate against the false positive rate for all threshold settings between 0 and 1. The true-positive rate is also known as *sensitivity* and the false-positive rate is also known as the *false alarm rate* and is the converse of the *specificity* [31]. The area under a receiver operating characteristic curve quantifies the overall ability of the test to discriminate between cases where a failure has occurred or not. Values close to 1 represent good performance, while 0.5 is no better than random.

Table 1
Failure factors identified in different literature.

	Focus		Physical										
	Water quality failure	Physical failure	Pipe joint	Pipe age	Pipe diameter	Pipe wall thickness	Pipe length	Pipe material	Valves	Pumps	Internal corrosion	External corrosion	Pipe breaks
Brandon et al. (1984)	✓	✓	✓					✓	✓	✓	✓	✓	
Savic & Banyard (2011)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Walski et al. (2001)	✓	✓			✓		✓		✓	✓	✓		
Babovic et al. (2002)		✓	✓	✓	✓		✓	✓			✓	✓	✓
Christodoulou et al. (2008)		✓			✓		✓	✓					✓
Francisque et al. (2009)	✓	✓		✓				✓					✓
Howard et al. (2004)	✓			✓	✓		✓	✓					
Islam et al. (2013)	✓			✓									
Kabir et al. (2014)	✓	✓	✓	✓	✓	✓		✓			✓	✓	
Lindley (2001)	✓												
Makar et al. (2001)		✓						✓			✓		✓
Makropoulos & Butler (2004, 2005, 2006); Makropoulos et al. (2003)	✓			✓	✓			✓					
Mamlook et al. (2003)	✓			✓				✓					
Rajeev et al. (2013)		✓			✓			✓			✓	✓	
Rogers (2011)		✓		✓	✓			✓					
Sadiq et al. (2010)	✓			✓									
Savić, A & Banyard, J (2011)	✓	✓	✓		✓	✓		✓	✓	✓			
Shi et al. (2013)		✓		✓	✓			✓					
Vairavamoorthy et al. (2004, 2007)	✓			✓									
Wang et al. (2009)		✓		✓	✓		✓	✓					

	Physical	Water quality								Operational				
	Temperature	Water age	Turbidity	Water colour	THHM	Free residual chlorine	Water pH	Water tempera-ture	Contaminants	Water pressure	Demand	Water velocity	Water hammering	Slamming valves
Brandon et al. (1984)	✓		✓	✓		✓		✓	✓	✓	✓	✓	✓	
Savic & Banyard (2011)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Walski et al. (2001)		✓								✓	✓	✓	✓	✓
Babovic et al. (2002)														
Christodoulou et al. (2008)														
Francisque et al. (2009)		✓	✓		✓		✓	✓						
Howard et al. (2004)									✓	✓				
Islam et al. (2013)		✓	✓	✓	✓	✓	✓	✓				✓		
Kabir et al. (2014)		✓	✓	✓		✓				✓	✓	✓		
Lindley (2001)									✓	✓				
Makar et al. (2001)														
Makropoulos & Butler (2004, 2005, 2006); Makropoulos et al. (2003)										✓				
Mamlook et al. (2003)											✓			
Rajeev et al. (2013)										✓				
Rogers (2011)														

(*continued on next page*)

Table 1 (*continued*)

	Physical	**Water quality**								**Operational**				
	Temperature	Water age	Turbidity	Water colour	THHM	Free residual chlorine	Water pH	Water tempera-ture	Contaminants	Water pressure	Demand	Water velocity	Water hammering	Slamming valves
Sadiq et al. (2010)		✓	✓	✓	✓	✓	✓	✓						
Savić, A & Banyard, J (2011)										✓	✓			
Shi et al. (2013)	✓													
Vairavamoorthy et al. (2004, 2007)									✓					
Wang et al. (2009)														

	Surrounding conditions												
	Soil corrositivty	Soil pH	Redox potential	Sulphide content	Soil resistivity	Moisture content	Poorly aerated soil	Installation	Faulty bedding	Inadequate anchoring	Freezing Index	Third party activity	Loading
Brandon et al. (1984)	✓				✓			✓	✓	✓		✓	✓
Savic & Banyard (2011)	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Walski et al. (2001)													
Babovic et al. (2002)									✓	✓		✓	✓
Christodoulou et al. (2008)													
Francisque et al. (2009)													
Howard et al. (2004)													
Islam et al. (2013)													
Kabir et al. (2014)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lindley (2001)													
Makar et al. (2001)													✓
Makropoulos & Butler (2004, 2005, 2006); Makropoulos et al. (2003)		✓											
Mamlook et al. (2003)													
Rajeev et al. (2013)								✓			✓	✓	✓
Rogers (2011)													
Sadiq et al. (2010)													
Savić, A & Banyard, J (2011)												✓	
Shi et al. (2013)													
Vairavamoorthy et al. (2004, 2007)													
Wang et al. (2009)													

5. Results

5.1. Automated structure learning of Bayesian network

The Bayesian network built using the hill-climbing algorithm (Fig. 5) had a complex structure with a large number of connections between the nodes. These included edges directed from the failure consequence nodes to risk influencing nodes (e.g. from circumferential failure to soil shrink swell), contrary to the naïve causal interpretation. There were also edges between design variables (e.g. pipe length linking diameter), between environmental variables (soil corrosivity and shrink swell), between consequence variables (e.g. from longitudinal failure to circumferential failure) and between variables that would be expected to be independent (e.g. lined pipe and soil shrink swell). These apparently unrealistic relationships arising from unguided fitting of a model to a large set of highly variable data, meant that the model would not be credible to industry practictioners.

5.2. Data and knowledge guided structure learning of Bayesian network

5.2.1. Literature review

A summary of contributing factors to water distribution system failure reported in different studies is presented in Table 1 [6,7,13,32–48]. From the literature, corrosion along with material variability were found to heavily influence failure rates [49–51]. The weather is also an important influencing factor on soil shrinkage as it affects the moisture in the soil, when there is wet weather the soil will expand and shrink when there is dry weather. Under soil expansion, vertical loads, typically crushing loads, are exerted onto the pipe, while soil shrinkage induces shear stress [52]. Therefore, there was a need to combine soil corrosivity and shrink swell data with the pipe failure data in the Bayesian network.

Due to their absence in the data, certain important variables identified in the literature were not able to be included in the Bayesian network; hence, during the analysis proxies were used in their place. The pipe diameter was used as surrogate for the wall thickness, as there was a positive correlation between pipe diameter and wall thickness. Short pipe lengths could indicate possible fittings or junctions in roads.

The lining inside the pipe plays a role in reducing or mitigating internal corrosion, where common lining methods included cement mortar, epoxy and rapid cure, slip lining and close fit. Therefore a variable stating whether a pipe was lined was included in the Bayesian network.

5.2.2. Data exploration

All of the variables identified from the literature review as potentially important that were available in the data (directly or via proxies) were explored. The effects of the variables on the pipe age at failure were firstly visualised in boxplot diagrams; observations from the boxplots ranged from obvious differences between states (Fig. 6) on mean pipe age at failure to more subtle differences (Fig. 7). Fig. 6 shows that each material probably had a different mean pipe age at failure, where it is most apparently shown between grey cast iron and glass reinforced plastic pipes. On the other hand, some variables exhibit boxplots where states have similar median, interquartile range and data range (e.g. Fig. 7 between the states 100 m–200 m and 200 m–500 m).

The one-way ANOVA (Table 2) found low p values ($p < 0.0001$) for all the variables considered, indicating that they all had significant effects, even when boxplots showed small effects, such as Fig. 7. However, the residual sum of squares was still large for all variables. This is commonly found in large, highly variable data sets, where small differences may be statistically significant, but not practically important.

For each variable, a Tukey honest significant difference test was used to test whether the difference in response between each pair of values of the variable was significant. For the pipe materials (Table 3), which had the lowest residual sum of squares, the results indicated that the difference in mean pipe age at failure was both large (typically greater than 10 years) and highly significant between every pair of materials, as the adjusted p-value was zero or close to zero, apart from

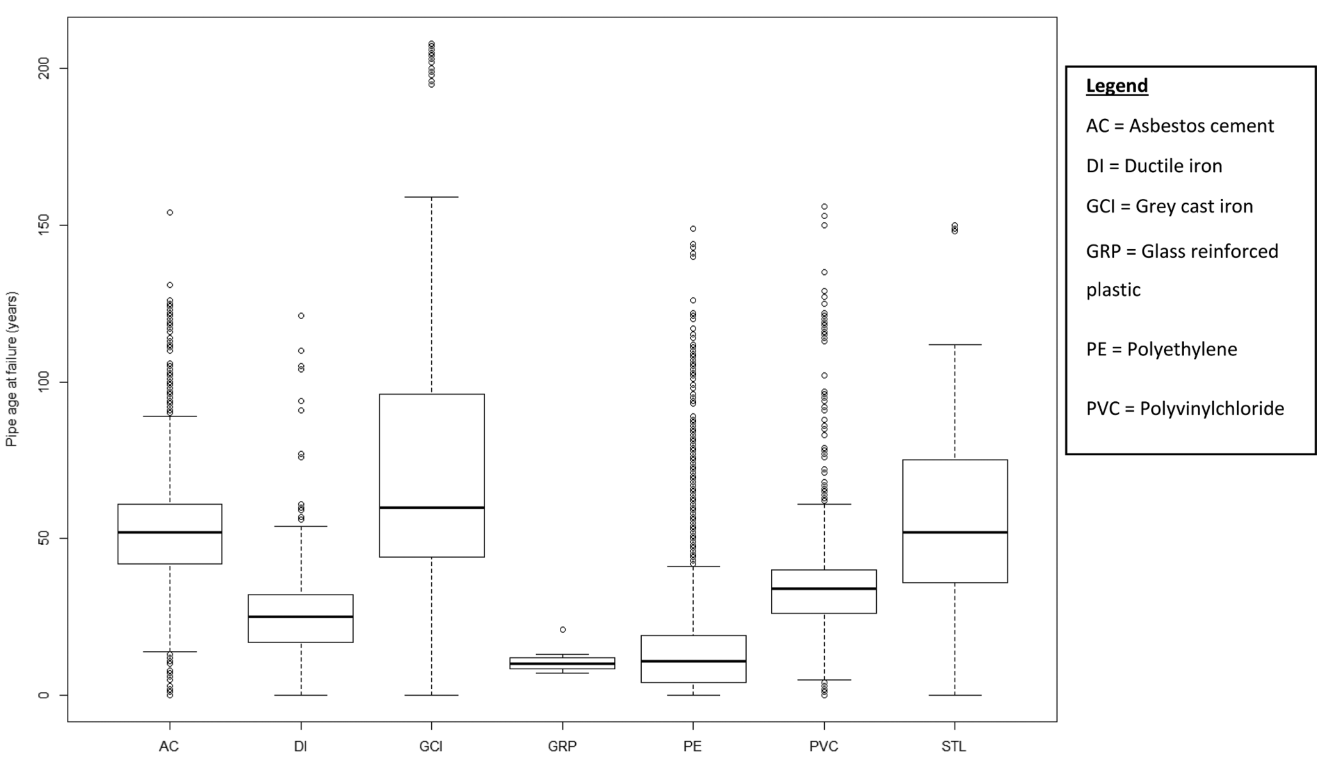


Fig. 6. Boxplot of pipe age at failure against pipe material.

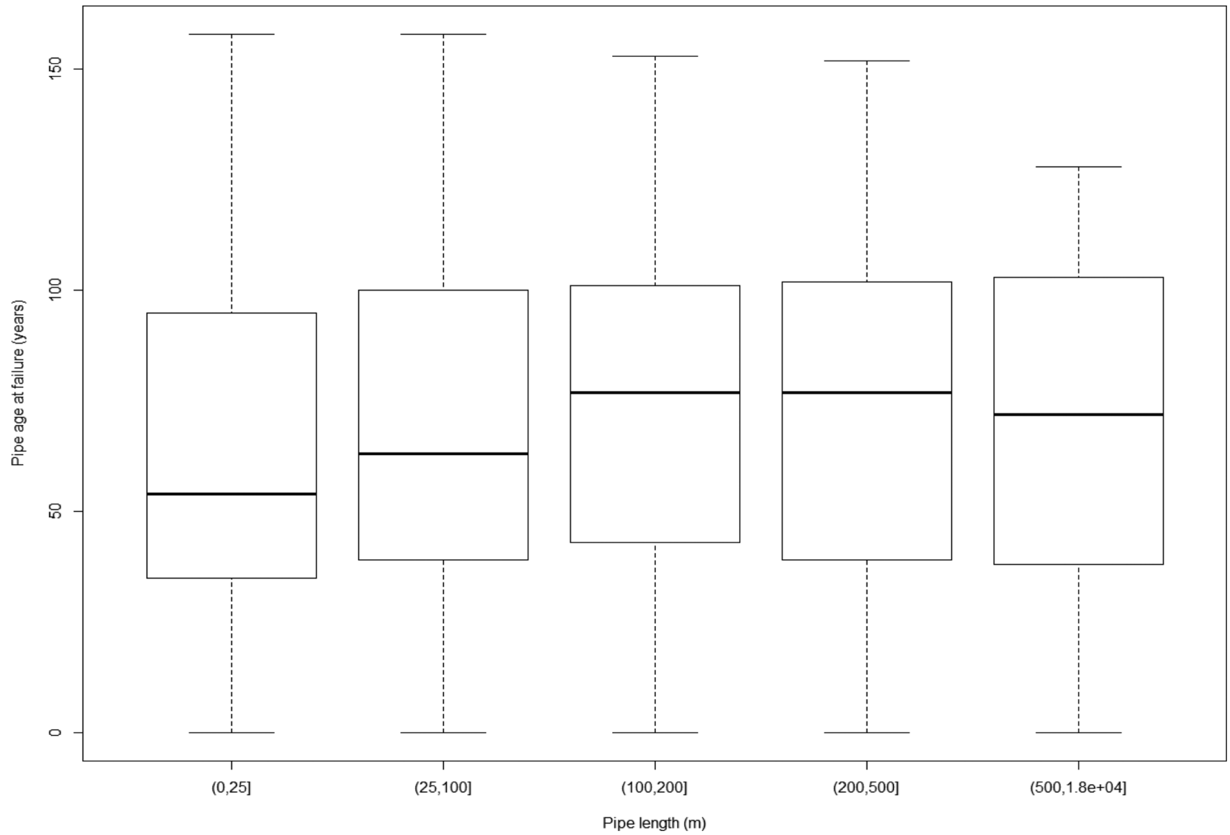


Fig. 7. Boxplot of pipe age at failure against pipe length.

between polyethylene and glass reinforced plastic (p = 0.88). The pipe failure data contained only a very small number of records for glass reinforced plastic pipes, so there is more uncertainty associated with results. For pipe length (Table 4), the differences were generally smaller (less than 4 years), but most were significant. It is unclear whether this difference is of practical importance, but pipe length was being used as a proxy for the number of junctions, so it was retained in the model.

Based on the analysis, there was no reason to exclude any of the variables considered from the model. The final model structure is shown in Fig. 8 and Fig. 9.

Table 2
One-way ANOVA test between all variables and pipe age at failure.

	Degree of freedom	Sum of squares	Mean square error	F value	Pr(>F)
Shrink swell	5	651,519	130,304	134	< 2.2e-16
Residuals	98,756	95,912,626	971		
Corrosion	8	580,708	72,588	75	< 2.2e-16
Residuals	98,753	95,983,437	972		
Pipe diameter	4	336,163	84,041	86	< 2.2e-16
Residuals	98,757	96,227,982	974		
Pipe length	4	86,224	21,556	22	< 2.2e-16
Residuals	98,757	96,477,921	977		
Lining	1	317,570	317,570	326	< 2.2e-16
Residuals	98,760	96,246,575	975		
Failure type	3	2,648,085	882,695	928	< 2.2e-16
Residuals	98,758	93,916,060	951		
Pipe material	6	29,638,802	4,939,800	7289	< 2.2e-16
Residuals	98,755	66,925,343	678		

5.3. Parameter learning of Bayesian network

The goodness-of-fit of both Bayesian networks was tested by the receiver operating characteristic curve and calculating the proportion of correct responses (accuracy) when comparing fitted results with the validation dataset using the Bayesian network against actual results. The guided Bayesian network obtained an overall model accuracy of 81.4% on validation data and the automated Bayesian network obtained a slightly higher accuracy of 84.4% (Table 5). The greatest difference in prediction accuracy between the two methods was for pipes decommissioned without failures with the guided Bayesian network having a comparatively lower accuracy of 65.9% compared to the automated Bayesian network of 79.1%. This was not unexpected, as the network from the automated method contained more interconnections between nodes, and hence more fitted parameters, than the guided method. Although the guided learning Bayesian network was much sparser, it was only less accurate by 3%.

The area under the receiver operating characteristic curve also shows that the performance of the two methods was either similar (longitudinal and circumferential failures) or the automated method performed better (Table 6). Figs. 10 and 11 show an example for the longitudinal pipe failures: the area under the curve was 0.72 and 0.75 for the guided and automatic methods respectively, which indicated that both Bayesian networks were “good” at separating cases where longitudinal breaks had or had not occurred.

From the training dataset used in parameter learning of the Bayesian network, grey cast iron contained the majority of failures (67%). Grey cast iron has been installed in the UK water distribution system for the longest period and it is the most abundant pipe material with 173,203 km (as of June 2017 from UKWIR national mains data), so this may not be a surprising result. It also supports the finding that the majority of failures are circumferential breaks and pinhole failures, as they are both common failures in grey cast iron. It was found that a majority of failures occurred in low shrink swell and low corrosivity soils; this is probably due to a high proportion of water pipes being

Table 3
Tukey HSD test between states in pipe material with respect to pipe age at failure.

Paired states	Difference	Lower	Upper	*P* value adjusted
GRP-GCI	−57.38	−70.18	−44.59	0.00
PE-GCI	−52.02	−53.03	−51.01	0.00
GRP-AC	−40.81	−53.62	−28.00	0.00
PE-AC	−35.44	−36.62	−34.27	0.00
PVC-GCI	−34.62	−35.37	−33.88	0.00
DI-AC	−27.03	−28.49	−25.58	0.00
PVC-AC	−18.05	−19.01	−17.08	0.00
STL-GCI	−11.72	−15.48	−7.95	0.00
PE-DI	−8.41	−10.01	−6.80	0.00
STL-AC	4.86	1.05	8.68	0.00
PVC-DI	8.99	7.53	10.44	0.00
GCI-AC	16.58	15.83	17.32	0.00
PVC-PE	17.39	16.21	18.57	0.00
PVC-GRP	22.76	9.95	35.57	0.00
STL-PVC	22.91	19.09	26.72	0.00
STL-DI	31.90	27.93	35.86	0.00
STL-PE	40.30	36.43	44.18	0.00
GCI-DI	43.61	42.29	44.93	0.00
STL-GRP	45.67	32.34	59.00	0.00
GRP-DI	−13.77	−26.63	−0.92	0.03
PE-GRP	5.37	−7.46	18.19	0.88

Key: AC = asbestos cement; DI = ductile iron; GCI = grey cast iron; GRP = glass reinforced plastic; PE = polyethylene; PVC = polyvinylchloride; STL = steel.

Table 4
Tukey HSD's test between states in pipe length (m) with respect to pipe age at failure.

Paired states	Difference	Lower	Upper	*p* value adjusted
[500,18,000]–[0,25]	−3.47	−4.66	−2.28	0.00
[500,18,000] – [25,100]	−2.47	−3.44	−1.50	0.00
[500,18,000] – [100,200]	−3.24	−4.25	−2.23	0.00
[500,18,000] – [200,500]	−2.42	−3.47	−1.36	0.00
[100,200] – [25,100]	0.78	0.07	1.49	0.02
[25,100] – 0,25]	−1.00	−1.95	−0.06	0.03
[200,500] – [0,25]	−1.05	−2.08	−0.02	0.04
[200,500] – [100,200]	−0.83	−1.65	−0.01	0.05
[100,200] – [0,25]	−0.22	−1.21	0.76	0.97
[200,500] – [25,100]	−0.05	−0.82	0.72	1.00

installed in these soils, because they have the most suitable conditions for underground pipes.

Further analysis with the Bayesian network found that there was a higher probability of early life pinhole and joint failures for plastic pipes than grey cast iron pipes (Table 7). For PE pipes 35% and for PVC pipes 29% of these types of failures occurred in the initial installation year, compared to 10% for grey cast iron pipes. The total rate of PVC pipe failures consistently fell over time, whereas the failure rate of PE pipes remained relatively static from initial installation to year 20, before a fall of 8% from year 20 to year 40. On the other hand, grey cast iron failure rates were relatively low from initial installation to year 20, but increased rapidly from year 20 to year 40.

6. Discussion

6.1. Automated structure learning of Bayesian network

The data was cleaned thoroughly before modelling, but the automated algorithm is not immune to errors in the data regardless of how well the data is cleaned. Hence, it was possible that relationships could be wrongly determined or omitted due to unrepresentative or noisy data. The basic hill-climbing algorithm that was used requires the joint variables in the underlying network to follow a multinomial distribution, but in reality the data may not support this and with high-dimensional datasets this adds another layer of complexity for the automated learning.

This method resulted in a network that was complicated by relationships that existed in the data, but were not relevant to the aim of predicting failures from environmental and design variables. For example, there were relationships between design variables such as pipe lining and diameter (because small pipes are unlikely to be lined), and between soil characteristics (corrosivity and shrink-swell). Including these relationships will increase the complexity of the calculations during probabilistic inference.

6.2. Data and knowledge guided structure learning of Bayesian network

ANOVA makes assumptions of normality, homoscedasticity and absence of multi-collinearity, but the data did not meet all of requirements. For example, the distribution of age at failure was skewed and 'long-tailed' rather than normal, which is typical of failure time distributions, and the variance depended on the pipe material. The ANOVA had high residuals but low p-values, as is often the case with very large, highly variable datasets [53]. As a classical statistical method, its assumptions differ from those of Bayesian probability theory. However, the ANOVA results were used for only indicative purposes to examine relationships between variables identified from literature, these limitations were not critical.

The proposed Bayesian network model is imperfect, but could be a useful basis for utilities to build upon and to guide the collection of appropriate data for analysis in future. Bayesian networks are also 'portable' in the sense that, if the model was built correctly it is easy to provide prior distributions and include knowledge specific to another water distribution system without losing the historic information. The flexibility of Bayesian networks for both prognostic (forward) and diagnostic (backward) reasoning means that they can be integrated with cost models to aid decision making and also be extended onto multi-hazard frameworks when looking at external factors such as earthquakes, landslides and climate change etc.

6.3. Comparison of automated and manual network construction

In many industries comprehensive data collection has still not been achieved, which also applies to the water industry. The Bayesian network obtained using the hill-climbing algorithm (Fig. 5) had many flaws, and did not model the problem as it would be understood by the industry, but it did achieve higher failure prediction accuracy due to the vast number of connected nodes. Using the data and knowledge guided method the user has more control compared to automated learning algorithms. To obtain an accurate structure of the Bayesian network with either score-based or constraint-based algorithms, would need more comprehensive data about the water system or methods that could prevent inference of relationships between environmental or design variables. Without these there is a possibility of overfitting of the model when using maximum likelihood score-based methods and problems with the structure of the Bayesian network from the constraint-based approach due to errors arising early on in the search [54]. Applying the data and knowledge-guided method to learn the structure of Bayesian networks with this type of data will allow water utility companies to have a skeleton of a Bayesian network that will allow for more additional variables once the data collection becomes available.

6.4. Parameter learning of Bayesian network

From the fitted Bayesian network PE and PVC pipes were identified as having lower failure rates compared to their metallic counterparts [55] and although this was true with the overall model, there was a surprising finding with regards to the failures found in plastic pipes

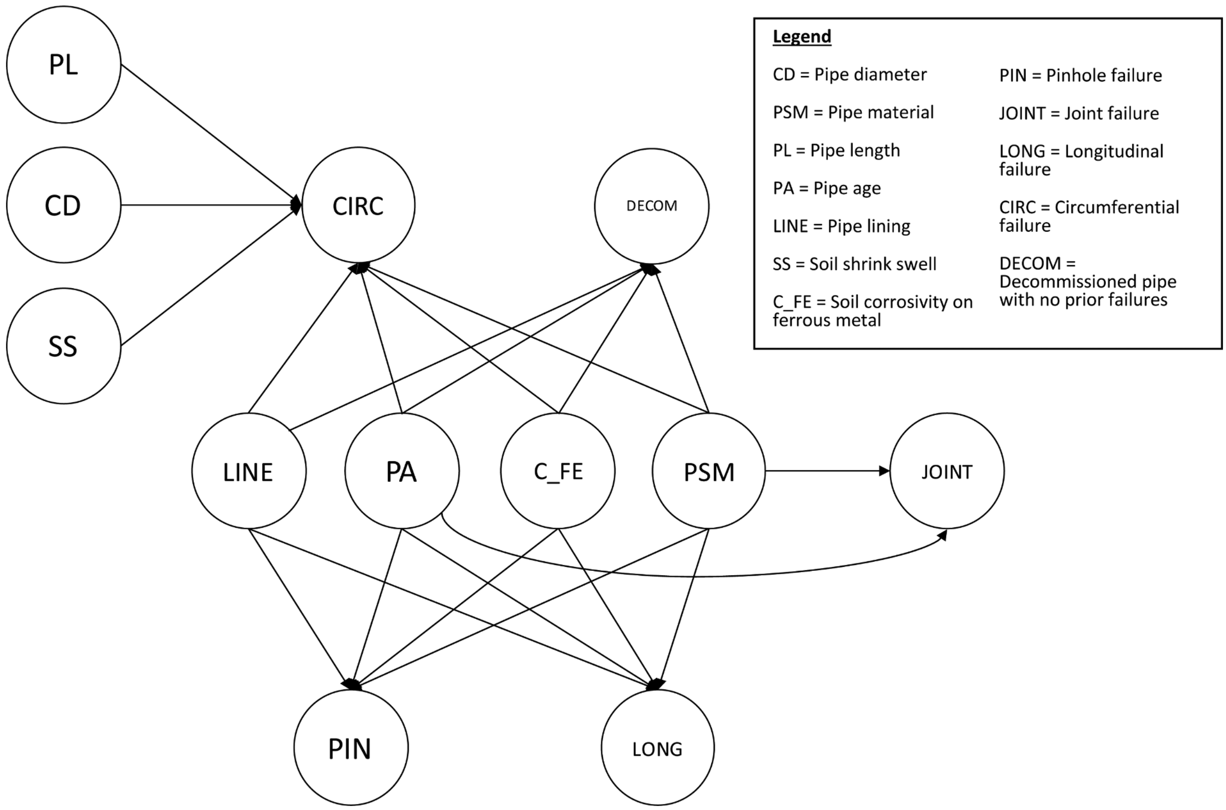


Fig. 8. Data and knowledge driven built Bayesian network.

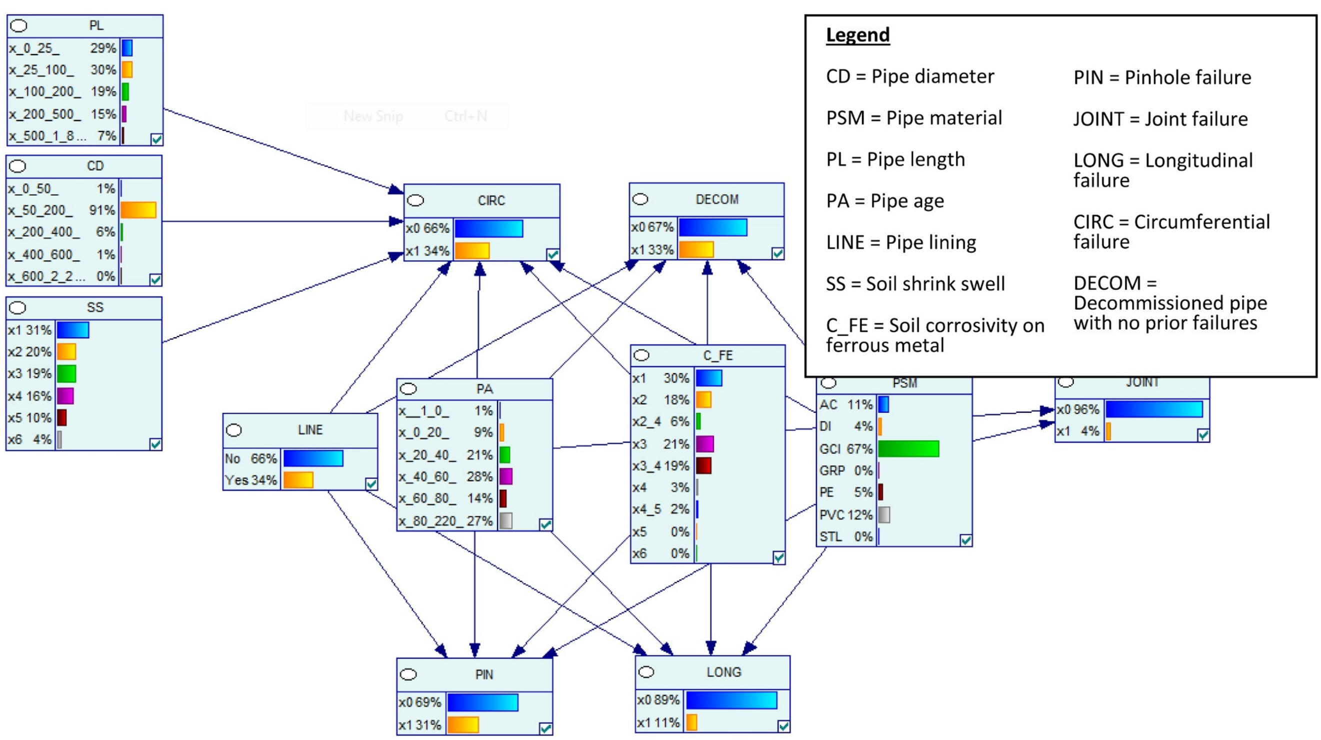


Fig. 9. Bayesian network initial states learned using expectation-maximisation algorithm.

presented in Table 7. PVC pipes in the UK were first used in the 1970s; however, there have been a steady decline in the use of PVC with utility companies favouring PE pipes. The data showed that there have been a higher than expected frequency of early life pinhole and joint failures of the plastic pipes with the probability of failure within the first year for PE pipes being 35% and PVC pipes 25%. Although theoretically there should not be any pinhole failures in plastic pipes (as this was a common trait for metallic pipes that has been under corrosion) this recorded failure could have been caused during installation or handling of the pipes before installation. Currently from the UKWIR failure data,

Table 5
Accuracy of the Bayesian network achieved during validation stage for both guided and automated learning.

Failure type	Accuracy Guided	Automated
Joint	95.6%	95.6%
Longitudinal	88.9%	89.3%
Circumferential	79.3%	78.7%
Pinhole	76.6%	79.3%
Decommissioned	65.7%	79.1%
Overall	81.2%	84.4%

Table 6
Area under the curve for each failure type for both guided and automated learning.

Failure type	Area under curve Guided	Automated
Joint	0.61	0.77
Longitudinal	0.72	0.75
Circumferential	0.79	0.78
Pinhole	0.71	0.79
Decommissioned	0.68	0.84

installation failure and handling error leading to failure cannot be proven; however, with increasingly strict industry standards for the installation [56–59] and handling of plastic pipes, the probability of early life failure should significantly decrease. Hence, as more data is trained onto the Bayesian network the probability of early life plastic pipe failures found by the model should decrease.

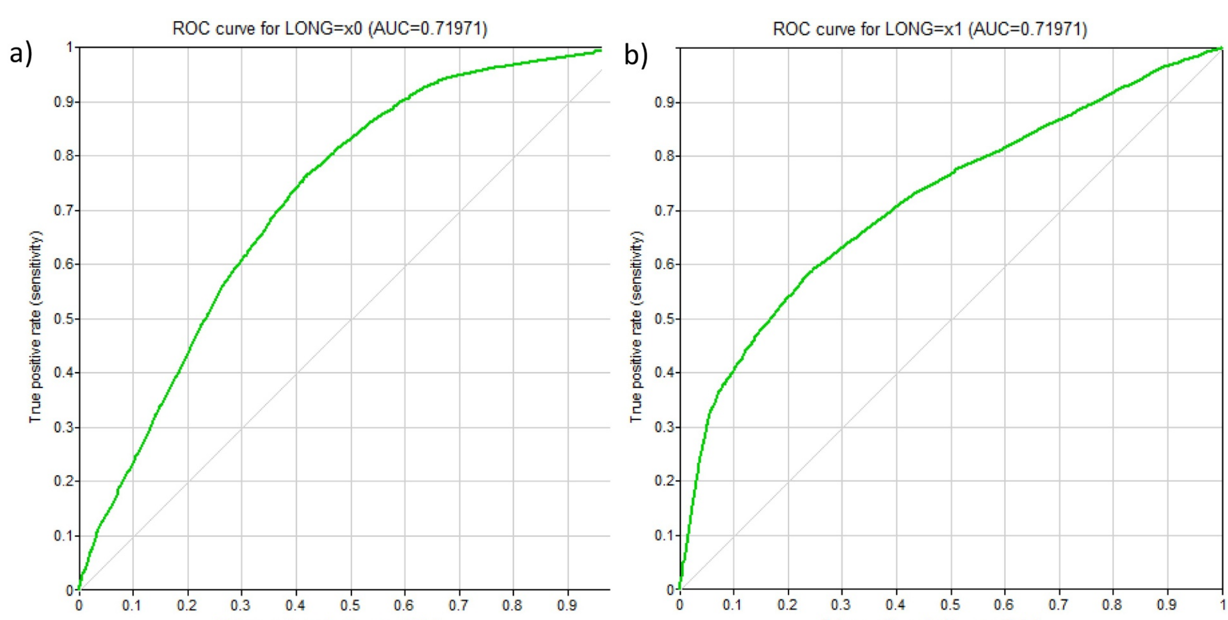
6.5. Limitations

In the validation results, longitudinal failure and circumferential failure had relatively high accuracy and "good" area under the receiver operating characteristic curve for the guided Bayesian network structure learning method. However, although the Bayesian network was able to achieve 95.6% accuracy in predicting joint failure, the area under the curve indicated the model was very "poor" at discriminating whether joint failures had occurred or not, which was supported by a sensitivity of 99.9% and specificity of 0.05%. This apparent contradiction was due to the very small proportion of positives (joint failures) in the data, which meant that a small error in the accuracy of predicting true negatives (non-failures) resulted in a large number of false positives relative to true positives. On the other hand, the automated method obtained not only the same high prediction accuracy for joint failure, but was also "good" at differentiating whether a joint failure has occurred or not, with a sensitivity of 99.9% and specificity of 3.56%. Although there was no difference in the sensitivity between the automated and guided method the specificity improved by over 700% under the automated method. With a dataset that consists of many non-failures and our interest lying in pipes that failed it can be argued that the specificity is more important than the sensitivity. It is possible to produce more coherent Bayes nets under the automated Bayesian network learning method by setting causality restrictions between variables, but whether this will improve accuracy requires further exploration.

With the Bayesian network built, probabilistic inference can be efficiently and easily carried out to find the posterior probability of either one of the failure types happening. However, the model does not include temporal information, which reduces the usefulness. Complex infrastructure systems, such as the water distribution system, are strongly affected by time-varying operating environments, such as the weather and water temperature. For example, underground water distribution pipes are subject to different shrink swells and levels of corrosivity in the soil depending on the weather over a period of time. Hence to overcome this limitation dynamic Bayesian networks can be adopted, which are Bayesian networks that can model temporal dependencies or time series structures. Therefore the Bayesian network built here can be further developed to include temporal nodes.

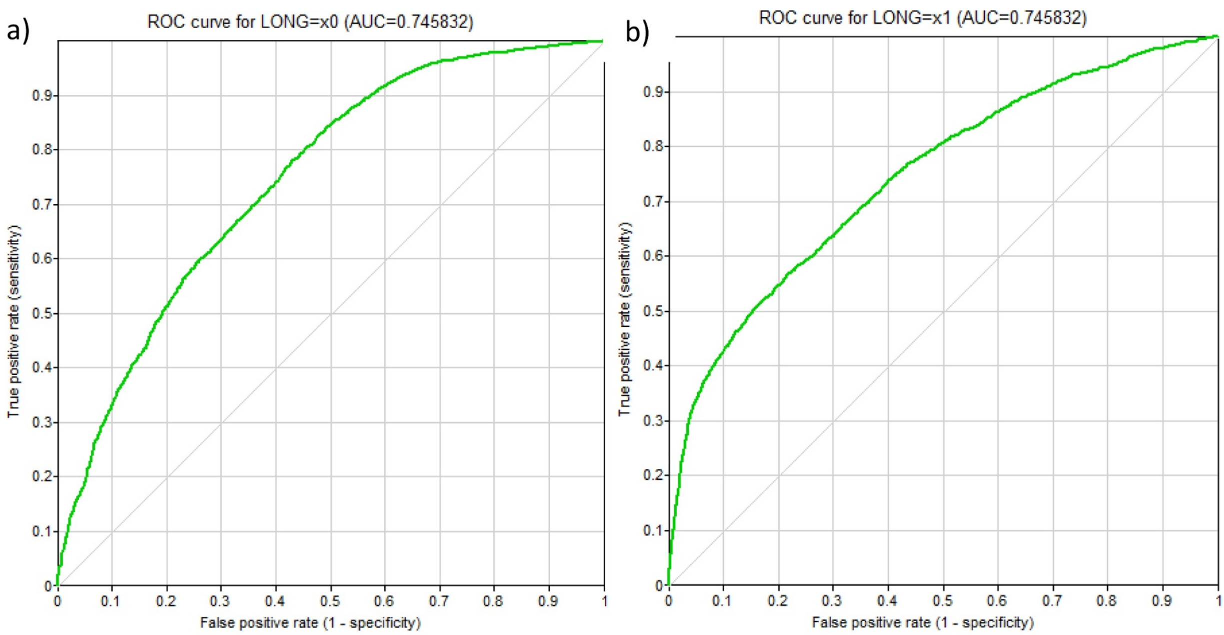
7. Conclusions

Bayesian network models of pipe failure in water distribution systems were constructed by automated learning and a manual method based on a literature review and exploratory data analysis. The



a) Longitudinal break not occurring, b) Longitudinal break occurring

Fig. 10. Guided learning receiver operating characteristic curve for longitudinal pipe failures.



a) Longitudinal break not occurring, b) Longitudinal break occurring

Fig. 11. Automated learning receiver operating characteristic curve for longitudinal pipe failures.

Table 7
Probability of joint or pinhole failure in PE, PVC and GCI pipes in years 0–40.

Pipe age	PE	PVC	GCI
0	35%	29%	10%
[0,20]	39%	23%	4%
[20–40]	31%	17%	45%

automated method resulted in an excessively complex network, with counter-intuitive relationships between many of the variables. Therefore, for this type of problem, involving extensive but highly variable data, collected over an extended period in a geographically distributed system, the manual approach is preferable. By utilising a detailed literature review along with a data exploration stage to verify the relationships, the analyst has full control of the building process and the knowledge the model encodes, as well as assessing the whether the quality of the data is compatible with the literature knowledge.

The automated Bayesian network learning method was able to obtain better accuracy than the guided Bayesian network learning method due to more having interconnections between nodes compared to the sparser network from the guided method. However, the improvement in the overall accuracy was only 3%

The Bayesian network found that a higher than anticipated proportion of failures occurred in the early life of plastic pipes, which suggests there may have been areas of concern when the novel material was first introduced. However, with better technology introduced for manufacturing and standards for handling and installation the volume of these early life failures should be reduced. Nevertheless, the early life failures provide an interesting area for investigation, especially the comparison between PE and PVC due to their differing failure trend over time.

Acknowledgements

This work was supported with funding by UK Water Industry Research Ltd and The Engineering and Physical Sciences Research Council [grant number EP/L505407/1].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2019.02.001.

References

[1] Ofwat. Service and delivery report. https://www.ofwat.gov.uk/publication/service-and-delivery-2017-18/; 2018 accessed January 8, 2019.
[2] Selvakumar A, Clark R, Sivaganesan M. Costs for water supply distribution system rehabilitation. Water Resour Plan Manag ASCE 2002;128:303–6.
[3] Røstum J. Statistical modelling of pipe failures in water networks. Norwegian University of Science and Technology; 2000.
[4] Fuchs-Hanusch D, Friedl F, Scheucher R, Kogseder B, Muschalla D. Effect of seasonal climatic variance on water main failure frequencies in moderate climate regions. Water Sci Technol Water Supply 2013;13:435–46. https://doi.org/10.2166/ws.2013.033.
[5] Huang C-N, Liou JJH, Chuang Y-C. A method for exploring the interdependencies and importance of critical infrastructures. Knowledge Based Syst 2014;55:66–74. https://doi.org/10.1016/j.knosys.2013.10.010.
[6] Babovic V, Drécourt J-P, Keijzer M, Friss Hansen P. A data mining approach to modelling of water supply assets. Urban Water 2002;4:401–14. https://doi.org/10.1016/S1462-0758(02)00034-1.
[7] Kabir G, Tesfamariam S, Francisque A, Sadiq R. Evaluating risk of water mains failure using a Bayesian belief network model. Eur J Oper Res 2014;240:220–34. https://doi.org/10.1016/j.ejor.2014.06.033.
[8] Francis RA, Guikema SD, Henneman L. Bayesian Belief Networks for predicting drinking water distribution system pipe breaks. Reliab Eng Syst Saf 2014;130:1–11. https://doi.org/10.1016/j.ress.2014.04.024.
[9] Kabir G, Tesfamariam S, Sadiq R. Predicting water main failures using Bayesian model averaging and survival modelling approach. Reliab Eng Syst Saf 2015;142:498–514. https://doi.org/10.1016/j.ress.2015.06.011.
[10] Atkinson K, Whiter JT, Smith PA, Mulheron M, Bouaziz MA, Guidara MA, et al. Reliability analysis and service life prediction of pipelines. Urban Water J 2013;3:1–16. https://doi.org/10.1080/1573062X.2012.739630.
[11] Rajani B, Kleiner Y. Comprehensive review of structural deterioration of water mains: physically based models. Urban Water 2001;3:151–64. https://doi.org/10.1016/S1462-0758(01)00033-4.
[12] Balkaya M, Moore ID, Saglamer A. Study of nonuniform bedding support because of erosion under cast iron water distribution pipes. Am Soc Civ Eng 2012:1247–56. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000689.
[13] Savic D, Banyard J. Water distribution systems. ICE Publishing, London, UK: Thomas Telford Publishing; 2011.
[14] Mora-Rodríguez J, Delgado-Galván X, Ramos HM, López-Jiménez PA. An overview of leaks and intrusion for different pipe materials and failures. Urban Water J 2013;11:1–10. https://doi.org/10.1080/1573062X.2012.739630.
[15] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, CA, USA: Morgan Kaufmann; 1988.

[16] Jensen F V, Nielsen TD. Bayesian networks and decision graphs. New York, NY: Springer New York; 2007. https://doi.org/10.1007/978-0-387-68282-2.
[17] Heckerman D. A tutorial on learning Bayesian networks. Technical Report, Microsoft Research, MSR-TR-95-06 1995.
[18] Wiegerinck W, Kappen B, Burgers W. Bayesian networks for expert systems: theory and practical applications. In: Robert B, Groen FCA, editors. Interactive collaborative information systems (Studies in computational intelligence volume 281) Berlin, Heidelberg: Springer; 2010. https://doi.org/10.1007/978-3-642-11688-9_20.
[19] Nagarajan R, Scutari M, Lèbre S. Bayesian networks in R: with applications in systems biology. New York: Springer; 2013. https://doi.org/10.1007/978-1-4614-6446-4_1.
[20] Pearl J. Causality: models, reasoning, and inference. 2nd ed New York: Cambridge University Press; 2009.
[21] Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. J R Stat Soc Ser B 1988;50:157–224.
[22] Dagum P, Luby M. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artif Intell 1993;60:141–53. https://doi.org/10.1016/0004-3702(93)90036-B.
[23] R. Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017https://www.R-project.org/.
[24] Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26. https://doi.org/10.1053/j.sodo.2009.03.002.
[25] ESRI. ArcGIS Desktop: Release 10. Redlands. CA: Environmental Systems Research Institute; 2012.
[26] Scutari M. Learning Bayesian networks with the bnlearn R package. J Stat Softw 2010;35:1–22.
[27] Scutari M. Learning Bayesian Networks with the bnlearn R Package. J Stat Softw 2010;35(3):1–22http://www.jstatsoft.org/v35/i03/.
[28] Koller D, Friedman N. Probabilistic graphical Models: principles and techniques. Cambridge, MA, US; London, UK: The MIT Press; 2009.
[29] Babovic V. Introducing knowledge into learning based on genetic programming. J Hydroinformatics 2009;11:181. https://doi.org/10.2166/hydro.2009.041.
[30] BayesFusion. GeNIe Modeler—User Manual 2017.
[31] MedCalc Software bvba. MedCalc Statistical Software—User Manual 2016.
[32] Makar JM, Desnoyers R, McDonald SE. Failure modes and mechanisms in gray cast iron pipes. NRC Publ Arch 2000:1–11.
[33] Makropoulos CK, Butler D, Maksimovic C. Fuzzy logic spatial decision support system for urban water management. J Water Resour Plan Manag—ASCE 2003;129:69–77.
[34] Makropoulos CK, Butler D. Spatial decisions under uncertainty: fuzzy inference in urban water management. J Hydroinformatics 2004;6:3–18.
[35] Makropoulos CK, Butler D. A neurofuzzy spatial decision support system for pipe replacement prioritization. Urban Water 2005;2:141–50.
[36] Makropoulos CK, Butler D. Spatial ordered weighted averaging: incorporating spatially variable attitude towards risk in spatial multicriteria decision-making. Environ Model Softw 2006;21:69–84.
[37] Mamlook R, Al-Jayyousi O. Fuzzy sets analysis for leak diction in infrastructure systems: a proposed methodology. Clean Technol Environ Policy 2003;6:26–31.
[38] Sadiq R, Kleiner Y, Rajani B. Modelling the potential for water quality failures in distribution networks: framework (I). J Water Supply Res Technol 2010;59:255–73. https://doi.org/10.2166/aqua.2010.059.
[39] Sun HF, Shi BY, Bai YH, Wang DS. Bacterial community of biofilms developed under different water supply conditions in a distribution system. Sci Total Environ 2014;472:99–107.
[40] Vairavamoorthy K, Yan JM, Galgale H, Gorantiwar SD. IRS-WDS: a GIS-based risk analysis tool for water distribution systems. Environ Model Softw 2007;22:951–65.
[41] Wang Y, Zayed T, Moselhi O. Prediction models for annual break rates of water mains. J Perform Constr Facil—ASCE 2009;23:47–57.
[42] Brandon T.W. Water distribution systems. The Institution of Water Engineers and Scientists; 1984.
[43] Walski T.M., Chase V. Water distribution modelling. Haestad Methods Inc.; 2001.
[44] Christodoulou S, Deligianni A, Aslani P, Agathokleous A. Risk-based asset management of water piping networks using neurofuzzy systems. Comput Environ Urban Syst 2009;33:138–49. https://doi.org/10.1016/j.compenvurbsys.2008.12.001.
[45] Francisque A, Rodriguez MJ, Sadiq R, Miranda LF, Proulx F. Prioritizing monitoring locations in a water distribution network: a fuzzy risk approach. J Water Supply Res Technol—AQUA 2009;58:488–509. https://doi.org/10.2166/aqua.2009.011.
[46] Howard G, Godfrey S, Tibatemwa S, Niwagaba C. Water safety plans for piped urban supplies in developing countries: a case study from Kampala, Uganda. Urban Water 2004;2:161–70.
[47] Islam MS, Sadiq R, Rodriguez MJ, Najjaran H, Francisque A, Hoorfar M. Evaluating water quality failure potential in water distribution systems: a Fuzzy-TOPSIS-OWA-based methodology. Water Resour Manag 2013;27:2195–216. https://doi.org/10.1007/s11269-013-0283-6.
[48] Lindley, T. A framework to protect water distribution systems against potential intrusions. (Electronic thesis or dissertation); 2001. Retrieved from https://etd.ohiolink.edu/.
[49] Atkinson K, Whiter JT, Smith PA, Mulheron M. Failure of small diameter cast iron pipes. Urban Water 2002;4:263–71. https://doi.org/10.1016/S1462-0758(02)00004-3.
[50] Doyle G, Seica M V, Grabinsky MWF. The role of soil in the external corrosion of cast iron water mains in Toronto, Canada. Can Geotech J 2003;40:225–36.
[51] Pelletier G, Mailhot A, Villeneuve J-P. Modeling water pipe breaks—three case studies. J Water Resour Plan Manag 2003;129:115–23. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:2(115).
[52] Laucelli D, Rajani B, Kleiner Y, Giustolisi O. Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. J Hydroinformatics 2014;16:743–57. https://doi.org/10.2166/hydro.2013.082.
[53] Lin M, Lucas HC, Shmueli G. Research commentary —too big to fail: large samples and the *p* -value problem. Inf Syst Res 2013;24:906–17. https://doi.org/10.1287/isre.2013.0480.
[54] Spirtes P, Glymour C, Scheines R. Causation, Prediction, and search. Springer Verdag; 1993.
[55] Selvakumar A, Matthews J, Condit W. Decision support for renewal of wastewater collection and water distribution systems. USEPA; 2011. https://doi.org/10.13140/2.1.4439.2005. EPA/600/R-11/077.
[56] ASTM International. ASTM D2683-14 Standard Specification for Socket-Type Polyethylene Fittings for Outside Diameter-Controlled Polyethylene Pipe and Tubing 2014. https://doi.org/10.1520/D2683-14.
[57] ASTM International. ASTM D2657-07 Standard Practice for Heat Fusion Joining of Polyolefin Pipe and Fittings 2015. https://doi.org/10.1520/D2657-07R15.
[58] ASTM International. ASTM D3261-16 Standard Specification for Butt Heat Fusion Polyethylene (PE) Plastic Fittings for Polyethylene (PE) Plastic Pipe and Tubing 2016. https://doi.org/10.1520/D3261-16.
[59] ASTM International. ASTM F1055-16a Standard Specification for Electrofusion Type Polyethylene Fittings for Outside Diameter Controlled Polyethylene and Crosslinked Polyethylene (PEX) Pipe and Tubing 2016. https://doi.org/10.1520/F1055-16A.