Check for updates

Water Supply

IWA PUBLISHING

© 2022 The Authors

Water Supply Vol 22 No 4, 3784 doi: 10.2166/ws.2022.019

An evolution of statistical pipe failure models for drinking water networks: a targeted review

N. A. Barton [a], S. H. Hallett [a,*], S. R. Jude [a] and T. H. Tran [b]

[a] School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom
[b] Centre for Competitive Creative Design (C4D), School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom
*Corresponding author. E-mail: s.hallett@cranfield.ac.uk

NAB, 0000-0002-1822-3590; SHH, 0000-0002-8776-7049; SRJ, 0000-0001-6737-7476; THT, 0000-0002-3989-4502

ABSTRACT

The use of statistical models to predict pipe failures has become an important tool for proactive management of drinking water networks. This targeted review provides an overview of the evolution of existing statistical models, grouped into three categories: deterministic, probabilistic and machine learning. The main advantage of deterministic models is simplicity and relatively minimal data requirements. Deterministic models predicting failure rates for the network or large groups of pipes perform well. These models are also useful for shorter prediction intervals that describe the influences of seasonality. Probabilistic models can accommodate randomness and are useful for predicting time-to-failure, interarrival times and the probability of failure. Probability models are useful for individual pipe models. Generally, machine learning approaches describe large complex data more accurately and can improve predictions for individual pipe failure models yet is complex and requires expert knowledge. Non-parametric models are better suited to the non-linear relationships between pipe failure variables. Census data and socio-economic data require further research. Choosing the most appropriate statistical model requires careful consideration of the type of variables, prediction interval, spatial level, response type and level of inference required.

Key words: data analytics, drinking water, machine learning, pipe failure, statistical modelling

HIGHLIGHTS

- Discusses key statistical models, including regression, probability, and machine learning.
- Reviews fundamental characteristics, limitations, and progress.
- Compares between the main outcomes and discusses future research.
- Synthesises the findings and includes an aid for different decision-making contexts.

1. INTRODUCTION

Developing reliable infrastructure decision-making tools is essential for managing large deteriorating Water Distribution Networks (WDNs), which pose economic, societal, and environmental threats if they fail (St. Clair & Sinha 2012). Water companies are interested in proactive management to determine which assets are likely to fail in the short (monthly, interannual), medium (annual) or long term (asset management period) (Watson *et al.* 2004). One area of innovation and opportunity is predictive pipe failure modelling, an area that focuses on predicting the time and location of pipe failures (also referred to as breaks), and a variety of such statistical models exist in the literature. Early pipe failure models were developed with the emergence of modern asset management and a drive to understand the cost of pipe failures. Over the past few decades, many models have been developed, benefitting from advances in computing and geographical information systems and databases able to collect and manipulate the substantive volumes and variety of data involved (Deadman 2010). Water companies now use pipe failure models in seeking new ways to understand and manage assets. Many such approaches focus on medium (annual) to long term (>annual) planning and concern the vital area of asset management, assessing the condition of a WDN and exploring strategies for pipe replacement, rehabilitation, and maintenance. However, few approaches focus on the short term, which would provide a framework to help operational management, an area of continued discussion. Model development is based on an initial assessment of the desired output, the specific needs of each water company and the

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

performance required across varying geographic locations. Other key variables concern the age of the WDN, the mix of material types, and the range of variables available within the data. It is essential to understand the different modelling approaches, each with its own merits and reasons for achieving different outcomes.

Pipe failure models may be broadly classified as heuristic, physical or statistical. Several comprehensive reviews summarise pipe failure models extensively (Kleiner & Rajani 2001; St. Clair & Sinha 2012; Nishiyama & Filion 2013; Scheidegger *et al.* 2015; Wilson *et al.* 2017). These literature reviews are systematic and focus on a particular period (the last decade) or type of model (deterministic and probabilistic), or specific aspects such as large-diameter pipes. These review papers do not capture the machine learning models widely used today. The idea of this exposition is to look purposefully at the progression of pipe failure models encompassing a range of statistical models, including deterministic, probabilistic and machine learning used in the literature. We use a targeted or focussed review process, selecting high-quality articles over time that help identify the trend and current state of pipe failure models, rather than an exhaustive list of literature or literature from a specific period or type. This approach is a means of discussing a narrative of change and progression in statistical pipe failure models. Key historical models are discussed, as pipe failure models evolved from simple deterministic models to current machine learning, describing some of the fundamental issues and discussing limitations and progress. The need to detail machine learning is essential given the increasing trend towards these models in the past decade. Therefore, an emphasis is placed on various methods being leveraged for their capability to describe and encapsulate complex relationships and return more accurate predictions. Regulatory incentives driving innovation (Ofwat 2020) are likely to spur further desire for machine learning and data analytics utilised to support data management of these substantive data repositories (Ponce Romero *et al.* 2017). Figure 1 presents a taxonomy of the different model types discussed in this review.

2. HEURISTIC MODELS

Water utilities have historically adopted heuristic models to calculate the failure rates of pipes to prioritise replacement scheduling. Heuristic models are often not included in reviews; however, they represent a common approach for assessing infrastructure where factors and mechanisms are not well understood. Instead, heuristic models use descriptive analytics such as assumptions, expert opinion, or subjective weightings on leakage rates, material type, and age to determine the necessity of a pipe's replacement. Heuristic models are accessible to water companies and do not require extensive prior knowledge of variable selection and data pre-processing. Furthermore, limited or missing variables will not affect overall accuracy. However, heuristic models are rarely optimal since they rely on subjective opinions and fail to capture the potential risks or accurately describe future failure rates (St. Clair & Sinha 2012; Fitchett *et al.* 2020; Snider & McBean 2020).

3. PHYSICAL MODELS

Physical models (mechanistic models) ordinarily aim to describe the mechanisms that contribute to pipe failure by analysing the pipe's stresses and strengths and have been reviewed in detail by Rajani & Kleiner (2001). Relating these mechanisms involves analysing detailed data on pipe structural properties (quality of installation, material type, age, diameter, and manufacturing details), internal and external loads, proximal factors (environmental conditions), and material deterioration (Rajani & Kleiner 2001). Physical models do not require large quantities of historical data for calibration or operation (Wilson *et al.* 2017); alternatively, pipe investigation is necessary to determine the pipe's state of deterioration and site investigations can be expensive and time-consuming. Once created, the model represents the specific pipe and does not extrapolate to other areas of the WDN, limiting the use of the model. For these reasons, physical models are ordinarily limited to large supply mains (Wilson *et al.* 2017).

4. STATISTICAL MODELS

Statistical models aim to predict pipe failures from historical data containing several correlated variables, which may be categorised broadly as pipe-intrinsic, environmental and operational (Barton *et al.* 2019, 2020). Finding an appropriate model assumes that the response is not random but is functionally related to the variables. Variables take the form $x_1, x_2, \ldots, x_n$, and form an n-dimensional input dataset typically denoted as X. In pipe datasets, ordinarily X is represented by quantitative data, being either inherently categorical (soil or pipe material) or continuous (e.g., pH). These variables remain mainly static whilst other factors include a dynamic time-dependent aspect (such as with weather and pipe pressure (Wols *et al.* 2019)). The response variable is similarly defined by an output space of Y described from $y_1, y_2, \ldots, y_m$, either numerical, in which

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

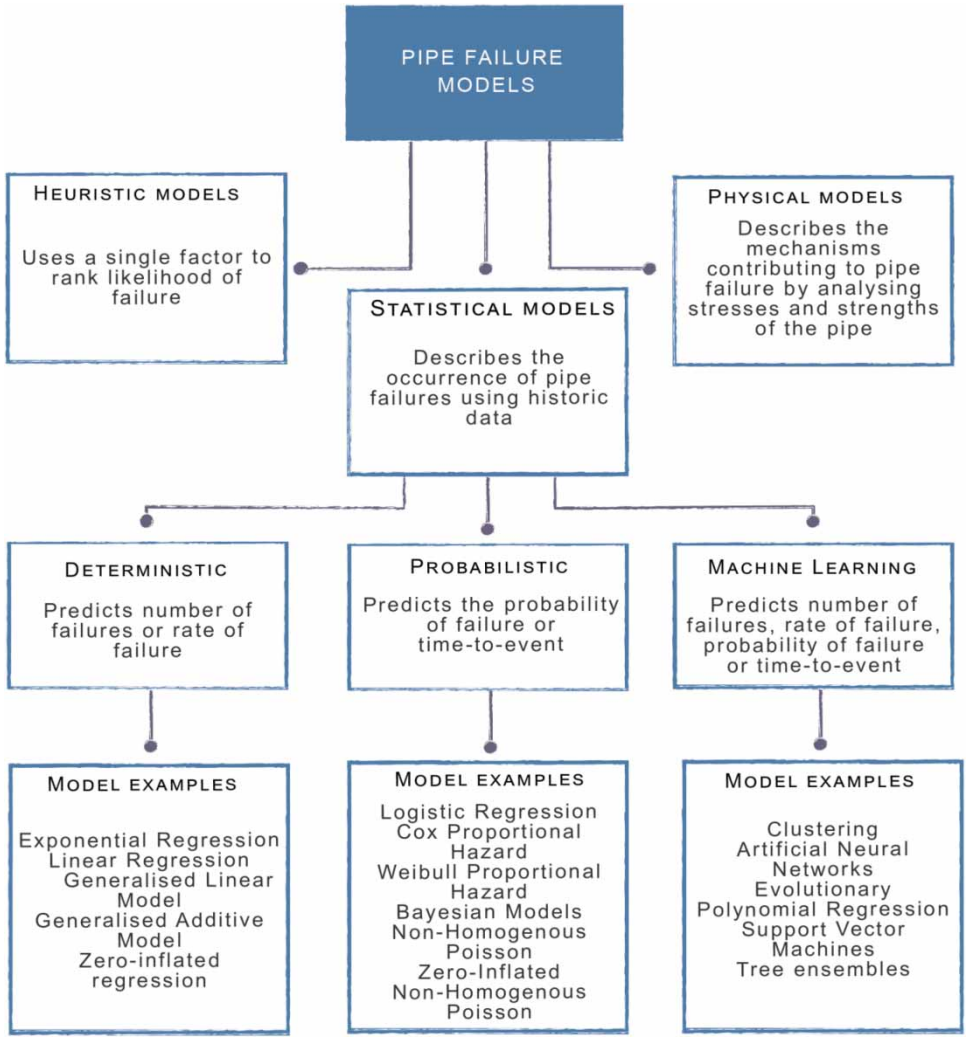

Figure 1 | Taxonomy of pipe failure models (adopted from Snider & McBean (2020)).

case it is regression and defined by the function $\varphi : x \rightarrow y$, where $y = R$, and here φ represents the prediction, categorical and therefore classification where $y = c_1, c_2, \ldots, c_m$ a set of classes, or probability where $y = P$ the probability of occurrence (Louppe 2014). Statistical models are favoured over physical models because they are cost- and time-effective, do not require information on the physical pipe mechanisms (such as deterioration) and are easily extrapolated to other pipes with similar characteristics across an entire WDN (Kleiner & Rajani 2001). Statistical models are classified in numerous ways, but commonly into deterministic, probabilistic and machine learning (Nishiyama & Filion 2014).

4.1. Deterministic models

4.1.1. Single-variate regression

Linear models are simple deterministic models that have been used widely in statistics for many decades. The single-variate linear model represents a one-dimensional space with a linear function given as $y_i = \beta_0 + \sum_i \beta_i x_i + \varepsilon$, where β_0 is the intercept, β_i is the weight of x_i and ε the error term. Its simplicity delivers reliability and interpretability, which is advantageous to water companies (sometimes referred to as a white-box approach). Models using this approach have been used to return the number of failures or by dividing the number of failures by pipe length provide a natural estimator of failure rate. Early aggregate models predicted the number of failures as a function of time, using linear relationships. Kettler & Goulter (1985) developed a time-linear regression model using pipe age and diameter, given by $N = k_0 A$, where N is the number of failures, k_0 is the regression parameter, and A is age of pipe at first failure. The Kettler & Goulter (1985) model was applied to a small

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

WDN in Winnipeg, Manitoba, using a ten-year failure record, and revealed a strong negative correlation between failure rates and diameter size. Despite the term of linear, linear models can also model and represent non-linear relationships. Shamir & Howard (1979) developed one of the first models on a small WDN of pipes, a single-variate time-exponential model using pipe age to predict the number of failures per year per 1,000 ft of pipe, such that: $N(t) = N(t_0)\exp^{A(t-t_0)}$, where t is time, $N(t)$ is the rate of failure at time t, t_0 is the reference year, and A is a failure rate growth coefficient. These early studies were limited by the single-variate approach, which proved of limited practicality for asset management due to its scale of operation, and for determining relationships with other important variables. Shamir & Howard (1979) also used a pipe group approach, assuming similar failure rates for all pipes within the group, which is not often the case.

4.1.2. Multivariate regression

Studies began to focus on predicting the number of failures as a function of multiple factors (or covariates) (Andreou *et al.* 1987a). Kleiner & Rajani (2000) were the first to include dynamic time-related weather variables (soil moisture, temperature, and frost) using multi-variate exponential models, which improved predictive capability and described the influence of the weather on failure rates. The main limitation in the study was the removal of static variables, which are essential since static variables such as soil are present in all WDNs and are known to influence pipe breakage (Rajani *et al.* 1996). Kleiner & Rajani (2002) again used the multi-variate exponential model and the same weather variables on groups of pipes, this time noting that more variables led to improved accuracy in the predictions. Additionally, operational factors such as cathodic protection were used, as was mains replacement rate, not used before. These variables were able to accommodate data bias in the WDN dataset caused by aggressive pipe replacement and the removal of several high failing pipes and essential information. The use of both environmental and operational variables meant the model could account more accurately for ageing over time, yet both models were simplistic in their approach and showed only moderate accuracy overall due to many conditional assumptions.

To improve multivariate analysis, different models were explored and further operational factors such as pipe pressure introduced. Wang *et al.* (2009) predicted annual failure rates for individual pipes using pipe-intrinsic factors for five pipe materials using multiple regression. The study was completed on a 432 km WDN in St Foy, Quebec City, based on data collected between 1987 and 2001. The approach showed a good level of accuracy through best-subset regression, with pipe length being an important factor: R^2 of 68.9, 65.0, 71.5, 78.4, and 81.3 for cast iron, ductile iron (without lining), ductile iron (with lining), PVC, and Hyprescon pipes, respectively. The major limitation is the focus on how annual failure rates affected variables, suggesting that longer pipes had lower failure frequency and that annual failure rates increased with age but did not predict when the subsequent failure would occur on a pipe. In a comparison of different models, Asnaashari *et al.* (2009) modelled failures for groups of pipes on a small-sized WDN (554 pipes) for ten years of pipe failure records, comparing multiple regression to Poisson regression for four materials separately (AC, CI, DI and PE). The authors reported a higher accuracy in the Poisson models compared with the multiple regression models; R^2 between 0.71 and 0.95, and between 0.52 and 0.88, respectively. The Poisson model was said to perform well because it can handle non-linear relationships more accurately. However, a deviation and Pearson chi-square test calculated overdispersion, showing low unity between the metrics, signifying overdispersion due to zero-inflated data. The study did not include dynamic time-related variables, which was another limitation, but focussed on intrinsic pipe factors of wall thickness, pipe diameter, pipe length, cover depth, location, failure history and pipe pressure.

4.1.3. Generalised models

Moving towards addressing the pipe-grouping issue, failure models started to predict individual pipes. The Generalised Linear Model (GLM) was used because of its capacity to model complex relationships through a differentiable monotonic link function (exponential, Poisson or logistic are favoured), allowing for non-normal distributions and taking the general form $y_i = \beta_o + b_i x_i$ (Yamijala *et al.* 2009). Yamijala *et al.* (2009) compared multiple models on a medium-sized WDN in Texas (<1,600 km) between 2000 and 2005, including exponential regression, Poisson GLM, and logistic regression, including important factors such as pressure, soil parameters and temperature. The model predicted failures for individual pipes on a six-month interval to capture temperature variation. The authors reported no good results due to data imbalance between the number of failures and non-failures observed in the data, potentially imbalanced further by the short prediction period. Nevertheless, the authors suggest that logistic GLM provided valid test model estimates of pipe reliability at this spatiotemporal scale since predicting the probability of a discrete outcome is often helpful for water companies (Yamijala *et al.* 2009).

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

The linear, exponential and Poisson models were performed on a reduced dataset, consisting only of pipes with failures to avoid the zero-inflation problem. The linear and exponential models were fitted with fewer variables than the Poisson and logistic model, which also used an iterative process to drop variables for a parsimonious model. In this instance, the Pearson's correlation coefficient was low for the Poisson model, which suggested overdispersion was not a problem in the data, unlike that found by Asnaashari *et al.* (2009), potentially due to using only pipes with failures. In a similar study by Motiee & Ghasemnejad (2019), the authors compared four models on individual pipes for a medium-sized WDN (583 km) in Tehran over four years of failure records (2004–2007), including linear regression, exponential regression, Poisson GLM, and logistic GLM. The authors found similar results, with the logistic regression model observed to be the most useful, but the study failed to include a range of important variables such as soil, temperature or pipe pressure and did not handle the zero-inflated data.

4.1.4. Zero-inflated models

Other studies have attempted to use models to accommodate the zero-inflated data inherent in most pipe failure datasets. The Zero Inflation Poisson (ZIP) model handles zero-inflation by allowing different probabilities through two components: the first generates zero counts, and the second generates counts with probability, some of which may be zero. In a study by Konstantinou & Stoianov (2020), several models were compared and evaluated on a network of 374 km during 2003–2016, using 550 pipe failures. A ZIP GLM and negative binomial GLM were amongst those used but were the only models that could manage the zero-inflated dataset for individual pipes. Both models poorly fitted the data and underestimated failures in the network: R^2 of 0.091 and 0.082, respectively. The authors recognised a lack of complete (no environmental variables) and reliable data, which is likely to have reduced the accuracy of the model, yet the machine learning models were capable of accurate predictions. Therefore, the zero-inflated and negative binomial models did not perform well on this dataset, but no explanation was given. The models were applied to pipe segments in two scenarios, first for each pipe and second for pipe lengths partitioned into 200 m lengths. The model performed better on individual pipe lengths because of the direct classification of the pipe type in the first scenario. The variables provided were sufficient and were standardised with a mean of 0 and standard deviation of 1, but with few pipe failures in the dataset, groups of pipes may have yielded better results.

4.1.5. Comments

The simplistic mathematical framework of deterministic models has made them popular with water companies, especially given the ease of interpreting their results. Single-variate models such as the time-linear and time-exponential models are easy to develop, and early models that predicted the whole network condition were informative. However, these approaches were of limited use since they do not provide enough information to describe pipe failures in detail, and the parametric design fails to describe the non-parametric nature of pipe failures. The introduction of multivariate models improves predictions by providing further information. Multivariate regression and Poisson regression models were used to some success when predicting and deriving further understanding of pipe failures. The introduction of dynamic variables showcased the influence of seasonal weather variation but can only be used with relatively large groups of pipes or the entire network and is not suitable for individual pipes. Further introduction of operational variables was necessary, yet studies observed that data was not always available. Predicting the number of failures or failure rate using pipe length as a normalisation factor was well suited to these high-scale models since failure frequencies are discrete and non-negative (Asnaashari *et al.* 2009).

A practical step in pipe failure models was the use of multivariate approaches predicting individual pipes due to the apparent advantages to water companies. However, this proved mathematically problematic due to the zero-inflation problem caused by an insufficient number of failures for each pipe. The GLM model extended more basic multivariate regression by allowing greater flexibility via a link function, with the Poisson distribution proving more successful than others, such as the exponential or linear distributions. One assumption of the Poisson model is that the response variance is equal to the mean, yet this is seldom the case in many pipe failure data, especially for zero-inflated data, resulting in overdispersion (Asnaashari *et al.* 2009). When this assumption is violated, the methodology does not meet the Poisson distribution assumptions (Asnaashari *et al.* 2009; Kleiner & Rajani 2012) and overestimates pipe failures with residuals showing significant error and bias (Konstantinou & Stoianov 2020), and other studies suggested no useful results (Yamijala *et al.* 2009). By definition, it can be suggested that predicting failure rates is problematic for individual pipes and would be best suited to models using large pipe-groups or network-wide models. In an attempt to resolve the overdispersion problem, zero-inflated models were introduced.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

The ZIP GLM and ZIP negative binomial GLM models are reported in the literature and treated failures and non-failures separately in the model. However, ZIP models show mixed outcomes, suggesting they do not necessarily improve predictive accuracy. Grouping pipes is essential for the Poisson model to meet the appropriate assumptions. Careful consideration should be given to selecting grouping schemes since they can affect the accuracy of the model and must consider the spatial scale for management decisions (Kleiner & Rajani 2001). Alternatively, individual pipe models should make use of probabilistic approaches. Table 1 provides a summary of the main deterministic models discussed.

4.2. Probabilistic models

4.2.1. Survival analysis

The main disadvantages of deterministic models are their inability to accommodate left-truncated data and their limited response, which led to the popularity of survival analysis. Survival analysis is capable of handling left-truncated data addressed analytically through the adaption of the likelihood function (see Carrión *et al.* (2010) for further detail), and can be used to predict service life, the time between failures, or the development of consecutive failures over time (Wilson *et al.* 2017). This is achieved through predicting the elapsed time between an initiating event and a terminal event, or the time between events (requiring a particular condition, see Røstum (2000, p. 43)), revealing a pipe failure history (Alvisi & Franchini 2010). The survival function and hazard function are fundamental aspects of survival analysis. The survival function is the probability that a pipe survives longer than a time, such that $S(t) = P(T > t)$, where $S(t)$ is the pipe at time t, and T represents the time until the occurrence of an event. The hazard function $h_i(t)$ is the probability of a pipe i failing within that

Table 1 | Summary of the main deterministic pipe failure models

Author	Model[a]	Network size	Failure history	Pipe materials[b]	Variables	Spatial level	Response
Shamir & Howard (1979)	Time-exponential model	–	–	–	Pipe-intrinsic	Network	Number of failures
Kettler & Goulter (1985)	Time-linear regression model	–	1959–1985	AC, CI	Pipe-intrinsic	Network	Failure rate
Kleiner & Rajani (2002)	Multivariate exponential model	–	1973–1998	CI, DI	Environmental Pipe-intrinsic	Pipe groups	Failure rate
Wang *et al.* (2009)	Five multiple regression models	432 km	1987 and 2001	CI, DI, HY, PVC	Pipe-intrinsic	Pipe groups	Failure rate
Asnaashari *et al.* (2009)	Multiple regression & Poisson regression	554 pipes	10 years	AC, CI, DI, PE	Pipe-intrinsic	Pipe groups	Number of failures
Yamijala *et al.* (2009)	Exponential regression & Poisson GLM	<1,600 km (85,000 pipes)	2000–2005	AC, CI, DI, PVC, ST	Pipe-intrinsic Environmental	Individual pipes	Failure rate
Kleiner & Rajani (2010)	Zero-inflated Poisson	146.6 km (1,091 pipes)	1961–2006	AC, CI	Pipe-intrinsic Environmental	Pipe groups	Number of failures
Motiee & Ghasemnejad (2019)	Linear regression, exponential regression & Poisson GLM	583 km	July 2004–December 2007	AC, CI, DI, PE	Pipe-intrinsic	Individual pipes	Failure rate
Konstantinou & Stoianov (2020)	ZIP GLM and ZIP negative binomial GLM	374 km	2003–2016	AC, CI, DI	Pipe-intrinsic Environmental Operational	Individual pipes/pipe lengths of 200 m	Failure rate

–, Unavailable information.

[a]GLM=Generalised Linear Model; ZIP=Zero-Inflated Poisson.

[b]CI, cast iron; DI, ductile iron; ST, steel; AC, asbestos cement; PVC, polyvinyl chloride; PE, polyethylene; HY, Hyprescon; RC, reinforced concrete.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

time, formally expressed as $h_i(t) = f(t)/1 - F(t) = f(t)/S(t)$, where $f(t)$ is the probability density of function T and $F(t)$ is a cumulative distribution function (Rodríguez 2007). The hazard function is represented as either a constant, increasing, decreasing or bathtub-shaped (Røstum 2000), where a hazard of greater than 1 signifies an increased hazard and a decreased life expectancy of the pipe. The survival rate and failure of a pipe are illustrated in Figure 2, taking an exponential shape.

The Proportional Hazard (PH) is a standard model introduced by Cox (1972) and focuses directly on the hazard function. Early pipe failure models using the Cox PH were developed by Andreou *et al.* (1987a, 1987b), including a semi-parametric model developed to determine slow and fast failure phases. The proportional hazard and survival function are fundamental elements that define the Cox PH, which takes the general form: $h_i(t/X) = h_0(t) \times \exp(\beta_1 x_1 + \ldots + \beta_i x_i)$, with the assumption that the baseline hazard h_0 varies over time t, equivalent to the total hazard rate when the covariates equal 0 (Kabir *et al.* 2016; Wilson *et al.* 2017). The model predicted the probability of failure for individual pipes applied to two water utility networks in the USA. The authors reported that the failure rate increased with every subsequent failure up to the third failure, after that remaining constant. This was one of the first studies understanding the relationship between failures over time. The authors made other practical observations: smaller-diameter pipes fail more often, pipe pressure has a low correlation with pipe failures and pipes that fail early in their life-cycle outperform those that fail later. The main limitation with the study was the use of varying pipe lengths, with some pipes over 1 km, a length where conditions affecting pipe failure can vary greatly. A further limitation was the use of a single pipe material, which does not represent the whole network, and the lengths of pipe vary greatly in the study, which can affect the break rate. The main advantage of the Cox PH is that the hazard function assumes no form. Therefore, it is semi-parametric, which makes evaluating the effect of covariates on the hazard easier. The authors developed their model specifically for this reason.

When the objective is to predict future failures over a time horizon and of a particular form, a parametric model is considered more appropriate, such as the Weibull PH (Røstum 2000). The Weibull PH has also been widely used for pipe failure models (Mailhot *et al.* 2000; Røstum 2000; Alvisi & Franchini 2010) and is a simpler parametric model that assumes the natural log of time-to-failure is linearly related to the covariates (Le Gat & Eisenbeis 2000). Le Gat & Eisenbeis (2000) used a Weibull PH to understand the inter-arrival time of pipe failures, using short maintenance records to forecast failures for different pipe materials. In this study, the authors used a Monte Carlo simulation to generate random data in the model, which helped improve the predictions. Applied to two datasets in Charente-Maritime and Lausanne, the model worked well in the first but failed to predict the second dataset well due to inherent increased pipe degradation, inadequate data, and missing environmental variables. Mailhot *et al.* (2000) used a failure order model applied to a 21-year (1976–1996) dataset in Chicoutimi near Quebec City, with a WDN of approximately 352 km (2,096 pipes). The authors used four combinations of the Weibull and exponential distributions, comparing different installation periods on the shape of the hazard function, specifically considering short failure records and left-truncation of the data. The results showed that the installation period strongly influenced pipe failures and it was concluded that an explicit unrecorded period of data must be included when failure history is absent. However, the model was limited by considering subsequent failures to follow a single-variable

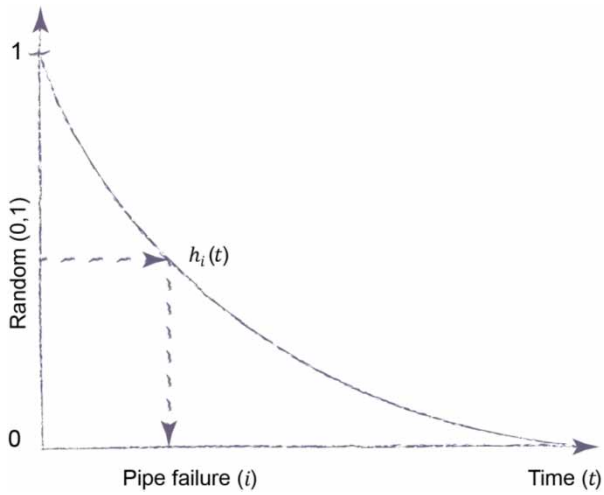

Figure 2 | Illustrative example of a survival curve (adapted from Le Gat & Eisenbeis (2000)).

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

exponential-type distribution. Alvisi & Franchini (2010) applied both models from Le Gat & Eisenbeis (2000) and Mailhot *et al.* (2000) to a WDN in Ferrara, Italy, with a failure record of seven years and ~3,500 failures. Comparing groups of pipes, the authors found that the models performed similarly in predicting failures, yet the model by Mailhot *et al.* (2000) was restrictive due to only considering the pipe installation period, where large networks have more data availability. Scheidegger *et al.* (2013) used a Weibull distribution to model the first failure and all subsequent failures using an exponential distribution to facilitate smaller datasets by simplifying the approach to accommodate fewer variables.

Debón *et al.* (2010) compared different models to understand which performed more reliably when predicting the time of failure. Three models were chosen: a Cox PH model, Weibull accelerated lifetime model and a GLM. The results led the authors to conclude there was minimal difference between the Cox and Weibull models, but comparing the Cox model to the GLM, the GLM showed a better fit when comparing the Receiver Operating Curve (ROC) (0.77 and 0.83, respectively), a measure of accuracy that would later be used in many more studies. Unusually, the authors noted that pipes installed under sidewalks were less prone to failure. Comparing the different survival analysis models and a Poisson model using a WDN consisting of 31,662 individual pipes (4,281 km) grouped by similar characteristics, Kimutai *et al.* (2015) reported a lower RMSE for the Weibull PH for two out of three materials, suggesting it performed better than the other models on this dataset. The Weibull model managed to capture the failure phases of metallic pipes more appropriately. Comparing the Weibull to a machine learning gradient-boosting model, Snider & McBean (2019) found that the machine learning outperformed a Weibull PH model on a WDN with a long failure history (1960–2006). The dataset had a failure period of 55 years, with 47% of the pipes having at least one failure record, which was reported as advantageous for the models. The machine learning predicted failures early because it was unable to account for censored data like the Weibull PH model.

4.2.2. Logistic regression

Logistic regression calculates the probability P of failure over the joint probability distribution $P(x, y)$ through the logit function (Louppe 2014), where a probability of 1 indicates a failure and a probability of 0 no failure, and all probabilities must be within this [0-1] interval (St. Clair & Sinha 2012). As previously discussed (see section 4.1.3), probabilistic models such as logistic regression have proved to be more useful when compared with Poisson, linear and exponential models, and returning the probability of a failure is enough to inform management decisions (Yamijala *et al.* 2009; Motiee & Ghasemnejad 2019), yet these studies are limited in their linear assumptions. Further studies used the logistic function in the Generalised Additive Model (GAM), rarely used or discussed in pipe failure literature, but valuable since it extends the GLM by including a smoothing function f_i that can measure arbitrarily non-parametric relationships, such that a simple form might be $y_i = \beta_0 + \beta_i x_i + f_i(x_i)$ (Barton *et al.* 2020). The semi-parametric models are advantageous over parametric models since they are adaptable to the complexity of pipe failure data. Chen *et al.* (2019) performed a comparison of models at two levels, the road level and the census level (2.3 km^2 grid areas), predicting the probability of failure using data recorded between 2010 and 2014. The study was one of the first to include a GAM with a logistic regression function and found that it performed best compared with several other deterministic and machine learning models at the road level, being able to identify areas of a high probability of failure more accurately. The authors acknowledged the limited accuracy overall due to a lack of asset-level information, the short prediction interval, the limited network information, and the limited failure records. Data imbalance was addressed, but the full details were not provided. Variable selection was performed for the models but found to have a limited effect on the overall accuracy. Compared with historical failure-rate-based ranking models, the GAM performed better and was preferred since it is easier to develop and computationally efficient when compared with machine learning models.

4.2.3. Bayesian models and expert knowledge

Due to a lack of suitable data, researchers have explored Bayesian models for their ability to incorporate prior beliefs or expert knowledge by constructing a probability distribution that describes the model uncertainty prior (prior distribution) to the data from the experiment to provide a formal estimate of failure rate (posterior distribution), such that $Posterior_i = Prior_i \times Likelihood_i / \sum_i Prior_i \times Likelihood_i$ (Watson *et al.* 2004; Kabir *et al.* 2016); a significant advantage given the complexity of WDNs and their local geographic nature (Deadman 2010). Watson *et al.* (2004) were one of the first to employ a hierarchical Bayesian model to simulate an object-orientated event, assuming the underlying failure rates for all pipes are drawn from the same prior distribution, called the hyperprior. Adjusting the hyperprior variance determines the individual failure information, where a small variance implies pipes have similar failure rates and a high variance that

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

pipes are different. The model was tested on two pipes over 50 years, updating the model every five years. Compared with a natural estimation (Poisson distribution), the Bayesian model with gamma prior distribution provided a better failure rate estimate. More recent developments combine survival analysis with the Bayesian model. Kabir *et al.* (2016) developed a Bayesian Weibull PH Model (BWPHM) using Bayesian model averaging for parameter estimation. The covariates included length, diameter, vintage, land use, freezing index, rain deficit, soil resistivity, and corrosivity. When comparing the BWPHM against a Cox PH and Poisson model for a medium-sized WDN (4,281 km) on individual pipes and data collected between 1956 and 2013, the results were favourable, with a Root Mean Squared Error (RMSE) of 8.65, 11.73, and 13.08 respectively for cast iron pipes. However, the Bayesian averaging benefits are not clear since the Weibull approach has been shown to outperform the Poisson and Cox PH models in previous studies (Kimutai *et al.* 2015). The model showed soil resistivity and soil corrosion to be the most important variables.

Further developments included the use of Bayesian Networks (BN), capable of capturing conditionally dependent and independent relationships through a directed acyclic graph structure with edges and parent and child nodes and links representing the causal dependencies between them, therefore explicitly quantifying uncertainties. A full description of the model is provided by Tang *et al.* (2019). This complex approach can explicitly quantify uncertainties, identify complex relationships, and incorporate engineering knowledge and data. In a study by Francis *et al.* (2014), a BN was applied to a WDN in the United States using a short dataset collected between 2010 and 2011. The data included 3,686 failure records, which were poorly located on the street level. To overcome this, the pipes were grouped at the census level. The results showed sustained challenges for the model, which the authors suggested were due to the monthly prediction interval and zero-inflation problem, and the inability of the BN model to discriminate between signal and noise due to limited information in the distributions. The study was limited by the unavailability of important pipe data, and therefore census, demographic, weather, and soil variables were used and standardised to deal with scaling issues; however, not many studies have tried to incorporate such population characteristics. The authors noted that many BN structures assume all random variables to be discrete, and so variables were discretised, which can have a significant impact on model accuracy, and in this study may have biased the results because of the zero-inflation problem. The authors also suggested that no single technique is better for obtaining the joint distribution from expert knowledge. Tang *et al.* (2019) applied an automated BN based on historical data and a guided BN utilising historical data, literature, and expert knowledge. The methods were applied to individual pipes collected across many WDNs across the UK. The authors reported poor accuracy (Area Under Curve (AUC) of 0.786 and 0.702, respectively) because of poor data quality, which influenced the Bayesian approach despite expert knowledge. Although not reported in this study, expert knowledge can be problematic with issues arising from different expert knowledge providing different posterior distributions (Kabir *et al.* 2018). Scholten *et al.* (2013) suggested that experts cannot be expected to know the distributional form or the correlation of variables with failure, and so developed a means to elicit expert knowledge into a form of stated quantiles, using an approach that minimised cognitive bias. The method used several experts to define the likely probability or frequency of relationships. Using the Bayesian inference and Weibull model to predict service life, the results showed improved service life estimation under scarce data. The authors compared the Weibull distribution to lognormal and gamma and found the Weibull more effectively approximated the non-parametric curve. However, these distributions take standard recognisable forms, which is rarely the case, therefore some studies have used the Monte Carlo Markov Chain (MCMC) technique to solve this problem (Economou *et al.* 2008; Lin & Yuan 2019).

4.2.4. Non-homogeneous poisson process

The Non-Homogeneous Poisson Process (NHPP) is a probabilistic modelling approach that captures pipe deterioration, expressed as a failure rate using a monotonic function of time such as the log-linear or log-power relationship. A critical feature of the model is that it relaxes the Poisson process and can, therefore, allow the failure rate to vary over time, capturing the deterioration mechanism (Economou *et al.* 2012). Røstum (2000) applied the NHPP to a WDN of 808 km in Trondheim, Norway, favouring the NHPP compared with a Weibull PH that overestimated the number of predictions. Economou *et al.* (2008) developed an NHPP and a Zero-Inflated NHPP (ZINHPP) using Bayesian parameter estimation and random effects. The NHPP process was flexible and captured the deterioration of pipes well, yet the ZINHPP model did not outperform the NHPP since the data had adequate information on pipe failures. Economou *et al.* (2012) further updated their initial study to account for the over-parametrisation. Here the model was applied to a case study of two medium-sized WDNs (532 pipes and 1,349 pipes, respectively) with different observation periods of 11 years (1990–2001) and 41 years (1962–2003). The authors found that the ZINHPP model performed better than the NHPP when comparing the Deviance Information Criteria

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

(DIC) for one dataset (New Zealand data DIC 344 and 515, respectively), allocating more probability to no failure. Both models showed similar results for the second dataset, potentially because there were more failures (only 22% non-failures compared with the first dataset of 85% non-failures) in the data; therefore, the data was less imbalanced. Neither Røstum (2000) nor Economou *et al.* (2012) used environmental variables limiting the study, and the main limitation of the NHPP is that the influence of previous failures cannot be modelled.

Kleiner & Rajani (2010) applied a ZIP model for a Non-Homogenous Poisson Process (NHPP) to a WDN in western Canada, comprising 1,091 pipes (length of 146.6 km) with a failure history between the dates 1961 and 2006. The variables used included material, diameter, installation year, length and soil and weather. The model successfully predicted the number of failures in each group for the five-year training dataset (R^2=0.61) but failed to accurately predict failures for individual pipes (R^2=0.43), overestimating the number of failures, which may have arisen due to many pipes having zero failure, and a few pipes having many failures. The study was performed on only 150 mm diameter unlined cast iron pipes, which limits the study, and whilst weather variables were used in training, they cannot be used to forecast pipe failures (Kleiner *et al.* 2010). Further developments for time-dependent weather variables (temperature, water temperature and freezing index) were undertaken by Rajani *et al.* (2012) to understand the influence on pipe failures and the best time-step for predictive accuracy. Predicting the mean failure rate for three datasets in the USA and Canada, the findings showed all influenced failure rates, but the temperature was vital since it is usually available. Comparing a prediction interval of 5, 15, 30, 60 and 90 days, the results showed improved accuracy with more extended periods, ultimately deciding on 30 days to be the most appropriate. However, the study modelled six large groups of pipes and did not address individual mains within the group, and the authors noted poorer results with shorter intervals. Kleiner & Rajani (2012), in a continuation of their research (Kleiner & Rajani 2010), used an NHPP model to predict annual failures from a long failure history (1961–2006) and a medium WDN of 146.6 km. Comparing four alternative models (an ordered list heuristic model, naïve Bayes, logistic regression and NHPP) and using important weather variables, the authors noted that the large dataset introduced noise which led to difficulty in obtaining meaningful results, and weather variables showed limited influence due to the pipe depth (2.4 m). The authors also noted time lags between the occurrence of failures and their discovery, potentially having discernible impacts on short prediction intervals. Amongst the models, only the NHPP could predict the number of failures. Therefore, the models were compared by their ability to rank pipes expected to experience the highest number of failures. The results revealed that no model was superior in ranking pipes overall, but that the NHPP was better at considering the time-dependent covariates and was advantageous since it could predict the number of failures.

4.2.5. Comments

Probabilistic models are flexible, can handle randomness, capture the complexity of multiple variables, and are capable of various response types. Probabilistic models are appropriate for modelling individual pipes, providing helpful information for rehabilitation and replacement. Survival analysis and, in particular, the PH model is insensitive to left-truncated data and can measure different failure phases of a pipe's life-cycle and times between failures. In this respect, survival analysis is unique since it can determine the best time to replace a pipe over a long prediction interval, considering the likely number of failures and the financial comparison between replacement or repair. Much of the literature has reported that the Weibull distribution suitably describes time-to-first failure, and the exponential distribution suitably describes subsequent failures. The main disadvantages of survival analysis include the complexity and difficulty handling short failure-recording periods, where explicit unrecorded periods of data should be included (Mailhot *et al.* 2000).

Logistic models offer a simple mathematical framework that water companies can readily use and provide a probability of failure between 0 and 1 or a classification approach of failure or no failure, helpful for supporting management decisions. Yamijala *et al.* (2009) and Motiee & Ghasemnejad (2019) have shown that logistic regression can work with short failure records, and is favourable compared with the linear, exponential and Poisson approaches when modelling individual pipes. However, all studies show limitations in data quality and focus on short-term intervals and short failure records, which may have resulted in the poor performance of other models. The main disadvantage is the assumption that the variables are related to pipe failures in a linear manner, which fails to describe complex non-parametric relationships. Therefore, Chen *et al.* (2019) explored the logistic approach using a GAM, which almost compared in performance to machine learning. However, other probabilistic models are more likely to outperform logistic regression, such as Bayesian models.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Bayesian models offer a dynamic approach to modelling failures, uniquely using a graph structure that can describe complex non-parametric relationships, are easy to interpret and can explicitly quantify uncertainty. Bayesian models are important given they can also include expert knowledge, advantageous given the lack of information in many datasets, but the joint distribution must be carefully chosen, and there is no preferred technique over others for obtaining this (Francis *et al.* 2014). Collecting expert opinions can be onerous (interviews can be used to elicit suitable responses (Scholten *et al.* (2013)) and introduces bias, since it is difficult for managers to establish the most important effects on pipe failures when they are often responsible for small areas of the entire network, instead of having a holistic view (Deadman 2010). Care must also be taken with insufficient failure data, since estimated parameters will be more sensitive to this prior information (Dridi *et al.* 2005). Bayesian models have been combined with survival analysis or non-homogeneous Poisson models as a means of parameter estimation, which has suggested improved accuracy. Like survival analysis, Bayesian models are computationally complex, which dissuades their use somewhat. The NHPP model is advantageous since it is flexible and allows failure rates to vary over time, capturing the deterioration mechanism in water pipes, which is useful when the actual time of a pipe failure is unavailable. Care must be taken when discretising data since this may cause bias in zero-inflated datasets (Francis *et al.* 2014). However, NHPP approaches account for the zero-inflation problem, yet little evidence suggests that zero-inflated models more accurately represent pipe failures. Table 2 shows a summary of the main probabilistic models discussed.

4.3. Machine learning

Recent developments in machine learning have greatly expanded the capabilities of statistical models and are now more commonly used than traditional models due to their improved predictive accuracy (Giraldo-González & Rodríguez 2020). The advantages of machine learning include removing unnecessary mathematical processing steps in some instances (decision trees (Winkler *et al.* 2018)), variable selection through shrinkage estimators, and tuning of interaction terms through cross-validation, which offers greater flexibility than simple regression models. For structured data, such as that seen in pipe failure data, supervised grey-box models are more appropriate (see Figure 3). Supervised models can be tuned for improved model accuracy and are interpretable through variable importance measures and partial plots (Wols *et al.* 2019), which is more appealing than black-box approaches to industry professionals (Barton *et al.* 2021). There are various studies in the literature that have applied machine learning and explored its effectiveness. Commonly used models in the literature include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Evolutionary Polynomial Regression (EPR) and more recently, tree models.

4.3.1. Clustering

Pipe failures can be modelled based on spatiotemporal relationships, assuming that more failures present within proximity to one another indicating distress in the network. Spatiotemporal relationships in failures were first noted by Clark *et al.* (1982) and Goulter & Kazemi (1988), reporting that consecutive failures were related to the first. These early studies provided essential insights not seen before. However, the studies did not constrain the relationships along with a pipe network, which distorted the results. De Oliveira *et al.* (2011) expanded this further by suggesting that spatiotemporal relationships should be network-restricted and used Dijkstra's algorithm (Dijkstra 1959) to ensure that the constraints of the pipe network have been considered. In the study, the authors used Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to find areas of the network underperforming by identifying hot spots with high failure rates. The authors conclude that poor repairs of failures might be responsible for producing new failures. Chen & Guikema (2020) used locally weighted DBSCAN and two Poisson-based models to determine the historical cluster of breaks and used the results as an exploratory variable in pipe failures. The results suggested that DBSCAN results in higher precision clusters and was then used in different regression models, including GLM, GAM, Random Forest (RF), and Gradient Boosting Trees (GBT), with each model including and excluding the cluster variable. The results revealed improved AUC scores for all models, including the cluster variable by 6.2%, 3.1%, 3.3% and 4.6%, respectively. The authors note that it is important to normalise data (to achieve a common scale of data) when performing cluster analysis and reported that overall accuracy measures do not always reflect the success of a model. Instead, the success should be based on the ability to prioritise pipes at high risk and therefore use a rank-based performance measure. Aslani *et al.* (2021) continued the use of failure clusters by using Getis–Ord (Getis & Ord 1992) to recognise spatial clusters of pipe failures in the city of Tampa, Florida, for a five-year dataset (2015–2020). The results of the cluster analysis were used as a variable for various machine learning models. The data were prepared

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Table 2 | Summary of the main probabilistic models discussed

Author	Model[a]	Network size	Failure history	Pipe materials[b]	Variables	Spatial level	Response
Andreou *et al.* (1987a, 1987b)	Cox PH	–	–	–	Pipe-intrinsic Operational		Probability of failure
Le Gat & Eisenbeis (2000)	Weibull proportional hazard model	1,243 km (1,212 pipes)	Nine years	AC, CI, PVC, ST	Pipe-intrinsic Environmental	Individual pipes	Number of failures
		688.7 km (6,966 pipes)	Since 1926		Pipe-intrinsic Operational		
Mailhot *et al.* (2000)	Weibull exponential model	352 km (2,096 pipes)	1976–1996	–	Pipe-intrinsic Environmental	Pipe groups	Failure age probability
Røstum (2000)	NHPP Weibull PH	7,627 pipes	1988–1996	CI, DI, plastic, others	Pipe-intrinsic Environmental	Individual pipes	Number of failures
Watson *et al.* (2004)	NHPP	–	–	–	Pipe-intrinsic Expert Knowledge	Individual pipes	Time-to-failure
Economou *et al.* (2008)	NHPP/ZINHPP (Bayseian framework)	1,349 pipes	1969–2003	CI	Pipe-intrinsic	Individual pipes	Failure rate
Yamijala *et al.* (2009)	Logistic regression	<1,600 km (85,000 pipes)	2000–2005	AC, CI, DI, PVC, ST	Pipe-intrinsic Environmental	Individual pipes	Probability of failure
Debón *et al.* (2010)	Cox PH Weibull accelerated lifetime model GLM	32,387 pipes	2000–2006	CI, DI, PE	Pipe-intrinsic Environmental Operational	Individual pipes	Time-to-failure
Alvisi & Franchini (2010)	Weibull proportional hazard model Weibull exponential model	2,400 km (23,000 pipes)	2000–2006	AC, CI, PE, PVC, ST	Pipe-intrinsic Environmental	Pipe groups	Time-to-failure (inter-arrival time)
Kleiner & Rajani (2010)	Zero-inflated Non-homogeneous Poisson process	146.6 km (1,091 pipes)	1961–2006	AC, CI	Pipe-intrinsic Environmental	Pipe groups	Probability of failure
Economou *et al.* (2012)	ZINHPP	1,349 pipes 532 pipes	1962–2003	CI	Pipe-intrinsic Operational	Individual pipes	Failure rate
Rajani *et al.* (2012)	NHPP	2,200 km (16,383 pipes)	1972–2001	CI	Environmental	Pipe groups	Number of failures
Kleiner & Rajani (2012)	Heuristic model, Naïve Bayes, Logistic regression and NHPP	370 km	1962–2006	AC, CI, DI	Pipe-intrinsic Environmental	Individual pipes (within a group)	Rank (pipe expected to experience the highest number of failures)
Scholten *et al.* (2013)	Weibull, lognormal and gamma models	322 km (3,643 pipes)	2001–2010	AC, CI, DI, PE, ST	Pipe-intrinsic Environmental Expert knowledge	Pipe groups	Time-to-failure (inter-arrival time)
Kimutai *et al.* (2015)	Weibull PH model Cox PH model Poisson model	4,281 km (31,662 pipes)	1956–2013	CI, DI, PVC	Pipe-intrinsic Environmental	Pipe groups	Time-to-failure (inter-arrival time)

(*Continued.*)

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Table 2 | Continued

Author	Model[a]	Network size	Failure history	Pipe materials[b]	Variables	Spatial level	Response
Kabir *et al.* (2016)	Bayesian Weibull PH Cox PH Poisson model	4,281 km (49,531 pipes)	From 1956	CI, DI, PVC, other	Pipe-intrinsic Environmental	Individual pipes	Time-to-failure (inter-arrival time)
Snider & McBean (2019)	Weibull PH xgboost	30,000 pipes	1960–2005	AC, CI, DI, PVC	Pipe-intrinsic	Individual pipes	Time-to-failure (inter-arrival time)
Motiee & Ghasemnejad (2019)	Logistic GLM	583 km	2004–2007	AC, CI, DI, PE	Pipe-intrinsic	Individual pipes	Probability of failure
Tang *et al.* (2019)	Bayesian network Guided/learning Bayesian network	–	1980–2017	AC, CI, DI, GRP, PE, PVC, ST	Pipe-intrinsic Environmental Expert Knowledge	Individual pipes	Probability of failure
Francis *et al.* (2014)	Bayesian belief networks	–	2010–2011	–	Pipe-intrinsic Environmental Census	Pipe groups (census)	Probability of failure

–, Unavailable information.

[a]GLM, Generalised Linear Model; NHPP, Non-Homogeneous Poisson Process; PH, Proportional Hazard; ZINHPP, Zero-Inflated Non-Homogeneous Poisson Process.

[b]AC, asbestos cement; CI, cast iron; DI, ductile iron; GRP, glass reinforced plastic; PE, polyethylene; PVC, polyvinyl chloride, concrete; ST, steel.

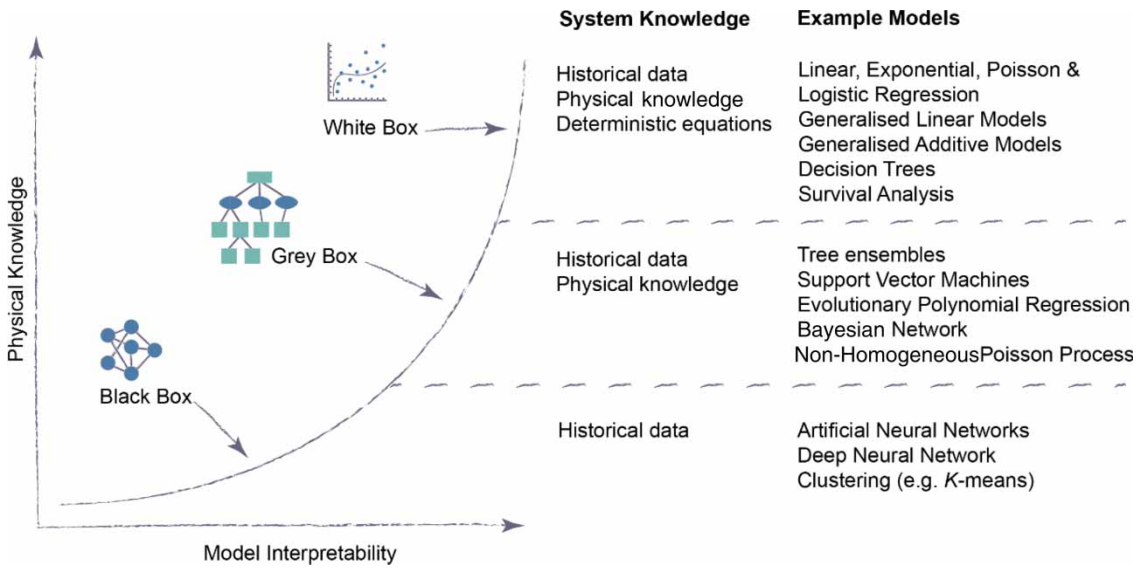

Figure 3 | Pictorial representation of white-box, grey-box, and black-box models, showing how the greater the physical knowledge that exists, the better the interpretability of pipe failures (adapted from Rezania (2008)).

using multiple imputations for missing values, and categorical data were processed into numeric data using one-hot-encoding. The accuracy of the models was presented as RMSE and rank-ordered break (similar to Chen & Guikema (2020)) and suggested poorly fit models, which is likely to be a result of the spatiotemporal extent of the model leading to data imbalance. However, the authors found the cluster variable to be significant and concluded that a boosted regression tree was the most reliable model.

K-means clustering assigns data samples into clusters around a centroid, using a function that iteratively minimises the dissimilarity between data (the Euclidean distance). The process then recalculates and moves the centroid reassigning the data, repeating the process until an optimal balance is achieved (Kleiner & Rajani 2012) (see Figure 4). *K*-means clustering is

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

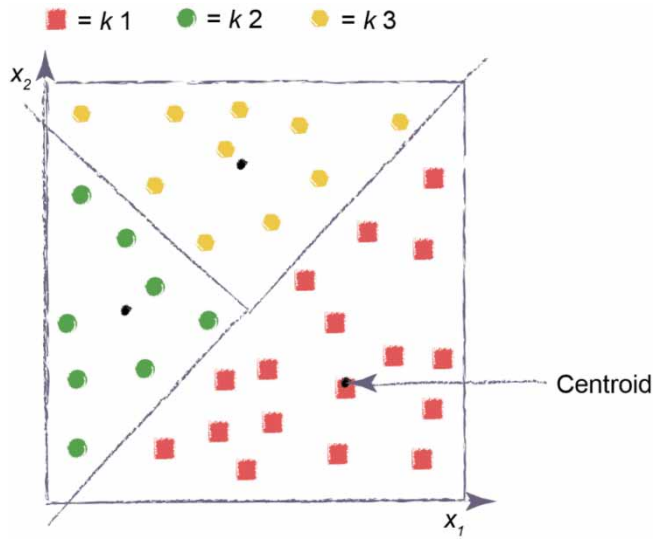

Figure 4 | Illustrative example of *K*-means clustering in a two-dimensional space, depicting three clusters of the *n* data samples, based on Euclidean distance to the cluster centroid.

typically applied during pre-processing failure models to group pipes with precision (Farmani *et al.* 2017), modelling each group individually to improve failure prediction accuracy. There is no specific study that suggests *K*-means clustering outperforms other grouping methods. However, Kakoudakis *et al.* (2017) compared an EPR model with a *K*-means-clustered EPR model, using a UK city WDN. The results showed better predictive accuracy in the clustered models when categorising pipe failure rates; 55% and 85% accuracy, respectively, for low, medium, high, and very high-risk categories. Cluster analysis is a valuable tool for identifying why failures may occur, yet it cannot predict the future pipe state since there is no response variable; therefore, it is not strictly in itself a failure model.

4.3.2. Artificial neural networks

ANN was one of the first machine learning models adopted in pipe failure models. It was used to improve the sometimes poor predictive accuracy seen from traditional models and is often likened to a biological neural network based on how the algorithm processes information (Kalogirou 2004). Considering this, ANNs comprise several process units known as neurons l arrayed in each layer. Each neuron within a layer is connected to each neuron within the joining layer. The neuron calculates a weighted sum w_{ij} of each input variable from the first layer to the second layer (hidden layer), which is made of activation units that perform the activation function f, transforming the data into non-linear outputs (such is the hyperbolic symmetric sigmoid function (Kutyłowska 2015), though many other functions exist). The third layer is a single activation unit that takes weighted outputs from the second layer and predicts y. An ANN takes the general form $y_j = f\left(\sum_{i=1}^{l} w_{ij}x_i + a\right)$ where a is a threshold (Wei *et al.* 2018). The process is illustrated in Figure 5.

Some of the early ANN models were combined with fuzzy logic; an algorithm used to handle imprecise information and approximate reasoning. In this respect, fuzzy logic is a technique that can return a logic gradient of between 0 and 1, and when combined with ANN, provides a human-style logic with the neural-style learning structure (Kleiner *et al.* 2004). Fuzzy logic alone has been used to determine a pipe deterioration procedure (Kleiner *et al.* 2004) and translate inspection investigations into pipe condition ratings (Rajani *et al.* 2006). One of the first ANN models was developed by Christodoulou *et al.* (2003), who used an ANN in a hybrid model with fuzzy logic. The model was used with Weibull distributions and Kaplan–Meier survival curves to explain the time-to-failure on groups of pipes. Twelve variables were used in the model and applied to a network in New York covering a total length of 365 km. Pattern recognition using fuzzy logic was employed to rank the risk of failure and finally assembled in a GIS platform for visualisation. The authors suggested that the model successfully ranked the risk of failure and identified material, diameter, length, and the number of failures as the most influential variables. Building on this model, Christodoulou & Deligianni (2010) further developed the use of the ANN model again with fuzzy logic for two small WDNs (365 km in New York, and 795 km in Limassol, Cyprus) of individual pipes using only five significant variables identified as important from previous studies, including traffic (not typically used in

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

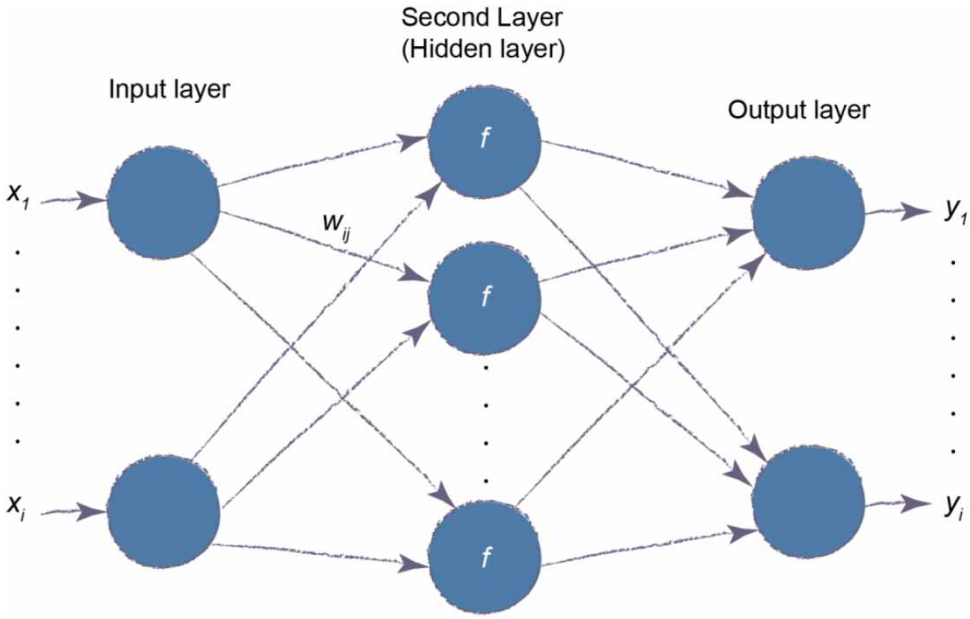

Figure 5 | Simple Artificial Neural Network architecture illustrating only three network layers (adapted from (Wei *et al.* 2018)).

models). The fuzzy logic approach is linked to a management tool that predicts the risk of failure over time, a useful output for industry. The authors attest to the success of the model, which was rolled out into a pilot scheme in two cities and highlighted the strength of ANN to facilitate incomplete time-sensitive multi-parameter data suitably. Tabesh *et al.* (2009) used an Adaptive Neuro-Fuzzy Inference System (ANFIS), a method that obtains the if-then rule automatically through the learning capability of the ANN model instead of through fuzzy rules. The ANFIS was compared with an ANN and applied to a network of 580 km of pipe and five pipe-intrinsic variables. The authors concluded that the ANN model provided more reliable results since it had a lower magnitude of variation in the results.

Other notable studies include Jafar *et al.* (2010), who employed different ANN models (split by material and a global model for comparison) to predict failure rates and replacement times of individual pipes on a dataset of 162 km (4,862 pipes) in northern France between 1991 and 2004. The authors used 11 variables that were standardised (since ANN models can be sensitive to outliers) to their minimum and maximum values after categorical data were converted into separate variables using dummy coding, including soil and pressure. A variety of variables were used, with length, age, and the number of previous failures the most significant. Like other studies, the global model performed well for cement, plastic, metallic and global models, R^2 of 0.589, 0.671, 0.522, 0.671 for the testing data. The main advantage of this research was its use of a benefit index for optimising investment, identifying 5% of pipes in the global model for replacement, potentially reducing failures by 53%. Nishiyama & Filion (2014) predicted the total number of failures over a two-year and five-year period applied to a WDN in Ontario, Canada. Pipe failure records were collected between 1998 and 2011 and included diameter, age, length, and soil-type variables. The results showed an approximate AUC of 0.78 and overall accuracy of 85%, yet the true positive predicted only ~40%. Harvey *et al.* (2014) only found acceptable results when predicting time-to-failure for asbestos cement, cast iron and ductile iron, with an adjusted *R*-value of 0.70, 0.70 and 0.81, respectively, when using data from a medium WDN of 5,850 km over a pipe failure dataset recorded between 1962 and 2005.

Kakoudakis *et al.* (2018) advanced the use of ANN by using the model to distinguish between a day with and without failure, a short prediction interval not previously attempted. The model used six clusters of pipes determined using *k*-means clustering, cross-validation for tuning and a binary response. The authors observed an AUC measure of 0.184, and the results help alert water companies to potential failure areas. Sattar *et al.* (2019) proposed a novel Extreme Learning Machine model (ELM), an alternative to ANN with a basic structure of a single-layer network, making it computationally fast and having a high generalisation capacity. The authors predicted time-to-failure using pipe coatings, material, length, and diameter and applied the model to a network in Scarborough, Greater Toronto, which contains 6,342 water mains over 1,000 km, and

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

failure records were collected from 1962. Comparing the model to feed-forward ANN, SVM and non-linear regression, the ELM had superior predictive accuracy. One significant advantage of this study was the long duration of failure records, which resulted in 44% of the pipes having had at least one failure.

When comparing ANN with other models, there are, however, mixed results. Asnaashari *et al.* (2013) predicted the failure of two models and found ANN (R^2 of 0.94) to have a better predictive capability than multiple linear regression (R^2 of 0.75, or 0.63 with cross-validation) when applied to a 5,850 km WDN with a long failure record history (since 1960). Giraldo-González & Rodríguez (2020) compared the probability of failure in four machine learning models and found that neural networks poorly predicted the minority class (failures). Over-fitting is the main disadvantage of ANN since the network must be identified *a priori* (e.g. the number of neurons, activation functions, training epochs, hidden layers, learning rate and momentum term) (St. Clair & Sinha 2012). In a review of literature, Wilson *et al.* (2017) excluded ANN models, suggesting they had high data requirements and were therefore unsuitable for large-diameter pipe models.

4.3.3. Evolutionary polynomial regression

EPR is a data-driven hybrid regression technique that uses mathematical structures based on evolutionary computing developed by Giustolisi & Savic (2006) and is widely used for pipe failure modelling. EPR is a two-step process: firstly, it looks for the best model structure using a Multi-Objective Genetic Algorithm (MOGA); secondly, it finds the parameter estimation using the ordinary least squares method (Berardi *et al.* 2008). In this respect, EPR builds symbolic models by integrating features of genetic programming (Xu *et al.* 2011), numerical regression, and symbolic regression, establishing parsimony between the polynomial terms and the variable. The general EPR form is $y = \sum_{j=1}^{m} F(X, f(X), c_j) + b$, where c_j is the constant, f is a user-defined function and F an EPR constructed function (Giraldo-González & Rodríguez 2020).

Berardi *et al.* (2008) used an EPR model to predict pipe failures and to derive performance indicators as a means of representing the propensity of a pipe to fail. The study assessed individual pipe criticality, defined as the failure likelihood and the expected damage. Pipes were grouped by diameter and age and only included pipes <250 mm in diameter, since the model was only economical for small-diameter pipes. The variables used in the model included age, length, diameter, and the number of properties. The accuracy measure for EPR was the Coefficient of Determination (CoD). The results showed that the EPR model overall fitted with a CoD of 0.822, but over-estimated failure rates in pipes with no failure history because of the limited failure history. The approach was simple and produced understandable relationships between failures and variables, yet the study did not validate the predictions. The main disadvantage in this study was that the timing of failures was unknown. Laucelli *et al.* (2014) used EPR to understand dynamic time-related climate variables. Using Rajani *et al.* (2012) as a starting point, the authors extended the range of dynamic temperature variables and applied EPR on a dataset for Scarborough (Ontario, Canada) focusing on 150 mm diameter CI pipes with a failure period between 1962 and 1985. Overall, the model showed reasonable accuracy (CoD=0.79); however, when investigating the relationship between climate data, the authors found good accuracy for failure rates during the winter, but not for the summer, since the phenomena that result in high failure rates for cast-iron pipes during the winter cannot explain the failures during the summer. However, the models determined warm and cold seasons by applying a threshold for the number of freezing days rather than distinct seasons, which may have influenced the results. The sensitivity of the model to these thresholds was not explored in the study.

Farmani *et al.* (2017) and Kakoudakis *et al.* (2017) explored the use of *k*-means clustering with EPR models, and both studies identified an improved performance when grouping the pipes and modelling each group with a separate EPR model. The authors presented the response as failure rates, and the risk of failure was categorised successfully by using Natural Jenks, an approach rarely used for pipe failure models. This premise was continued in a further study by Kakoudakis *et al.* (2018), where the authors used EPR with *k*-means clustering to partition the training data into different groups for the EPR. The pipes were divided into six clusters based on diameter and age, showing reasonable results from very low- to very high-risk categories for long-term predictions (with an accuracy of between 46% and 87%) and good performance for the short-term probability model (AUC of 0.814). Comparing EPR with other models, Giraldo-González & Rodríguez (2020) found that Poisson regression outperformed EPR (R^2 of 0.927 and 0.885, respectively) when predicting groups of pipes for a medium-sized WDN. The authors used *k*-means clustering to create pipe groups and observed its benefits. EPR models are intuitive and can include expert judgement in the process. The approach has mainly been used for groups of pipes and not for individual pipe models (Berardi *et al.* 2008).

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

4.3.4. Support vector machines

SVMs are linear models that maximise the width of separation between data classes while minimising the error (Louppe 2014, p. 22). Good separation between the data classes is achieved when a large distance, also known as the margin, exists on either side of a hyperplane (see Figure 6). A linear hyperplane satisfies $w^T x - b = 0$, and the upper and lower margins $w^T x_i - b \geq 1$ if y_i=1 and $w^T x_i - b \leq 1$ if y_i=−1 respectively, where w is the weight of the variable space, and b is a bias term. SVMs can implicitly model non-parametric data using a kernel function (standard kernel functions commonly used include polynomial, hyperbolic tangent, and radial), mapping the variables into a high-dimensional space, separated by multiple hyperplanes (Louppe 2014; Robles-Velasco *et al.* 2020). SVM is useful for handling outliers by allowing misclassifications, letting some data fall within the margin to obtain better results, but limiting their use in the model via a slack variable $\in_i$ applied to the outlier (Giraldo-González & Rodríguez 2020). Put together for a range of input variables, the basic SVM model is $y_i(w^T.\emptyset(x_i) - b) \geq 1 - \in_i$ for $i = 1, \ldots, N$ (Robles-Velasco *et al.* 2020).

In one of the first pipe failure studies using SVM, Shirzad *et al.*(2014) used regression to predict the failure rate of pipes on two WDNs in Iran, Mashhad WDN with 580 km of pipe and Mahabad WDN with 140 km of pipe. The dataset was limited to one year, with no further historical data, and only one pipe material used in each WDN, asbestos cement and polyethylene, respectively. Comparing the results with an ANN, the authors found that the ANN showed better statistical accuracy when predicting polyethylene pipes (R^2 of 0.963 and 0.775 respectively), but similar results when comparing AC pipes (R^2 of 0.995 and 0.997 respectively). However, the authors suggested that the ANN did not generalise well since it was not consistent with the observed data, and the SVM was considered more suitable. Kutyłowska (2019) also used an SVM regression model on a WDN dataset in Poland, including data collected between 2008 and 2014. The results showed a low relative error of between 4% and 14% for the different pipe materials (cast iron, PVC, and polyethylene) and compared four kernel functions (linear, polynomial, sigmoidal and radial), finding that failure frequency was predicted well employing a linear kernel for all materials.

Robles-Velasco *et al.* (2020) compared a logistic regression and an SVM probabilistic classifier to predict the likelihood of failure in a medium-sized WDN of 3,800 km of pipes in Spain, recording failures between 2012 and 2018, using sampling techniques to balance the data, *k*-fold cross-validation for tuning, and pipe-intrinsic data and pipe pressure, but without the inclusion of any environmental variables. The study revealed that a global pipe model performed well (better than many material-specific models) and that the logistic regression showed similar results to the SVM, AUC 0.873 and 0.872, respectively. The logistic model prevented 34.09% of all failures by replacing 3.16% of the WDN, whilst the SVM prevented 29.52% of failures by replacing 3.84% of the WDN. The study managed data imbalance by under-sampling and removing data from the majority class, which can be a limitation since some information may have been lost. Giraldo-González & Rodríguez (2020) also used an SVM probabilistic classifier, comparing it against ANN, Bayes, and Gradient Boosting Trees (GBT) for asbestos cement and PVC pipes using age, diameter and length for a medium-sized WDN (1,819 km) on data collected between 2012 and 2018. The results found SVM to perform better on asbestos cement pipes over PVC pipes; AUC 0.991

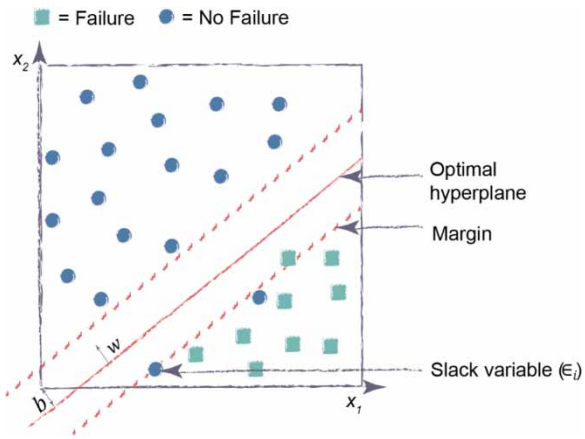

Figure 6 | A simple one-dimensional illustration of an optimal linear hyperplane maximising the separation between two categories of data, with a single slack variable within the margin.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

and 0.795, respectively. Both Bayesian models and GBT outperformed the SVM. SVM performs poorly on datasets with noise and overlapping classes (common in pipe failure datasets) and is perhaps why it compares poorly with other machine learning models. The authors identified that SVM is useful for handling outliers by allowing misclassifications, letting some data fall within the margin to obtain better results, but limiting their use in the model (Giraldo-González & Rodríguez 2020) via a slack variable $\in_i$ applied to each of these data points (see Figure 6). Since SVM assumes the data is within a standard range, it is important to standardise the covariates.

4.3.5. Tree models

Decision tree models are simple and can be used for regression and classification, and often as probabilistic classifiers (Robles-Velasco *et al.* 2021). Tree-based models have commonly been overlooked because they quickly become complex and prone to overfitting and high variance (more so than other machine learning models). Ensemble models significantly improve accuracy (Hastie *et al.* 2009), but are computationally expensive. With computational advances and quicker processing speeds, ensemble methods are more common and have shown better accuracy, even compared with several other data-driven pipe failure models (Chen *et al.* 2017; Giraldo-González & Rodríguez 2020). Tree ensemble models are simple in their approach and do not require mathematical pre-processing steps, and for this reason, they are an attractive machine learning model that is easily accessible (Winkler *et al.* 2018).

Ensemble approaches are best described in two parts, the decision tree and the ensemble method. A decision tree T partitions data into disjoint regions R_j through recursive partitioning along the axis. The regions are split to minimise prediction errors until the size of the tree reaches the terminal nodes J and stops based on a stopping criterion (the Gini index is commonly used); the process is visualised in Figure 7. A decision tree is formally described by Hastie *et al.* (2009) as

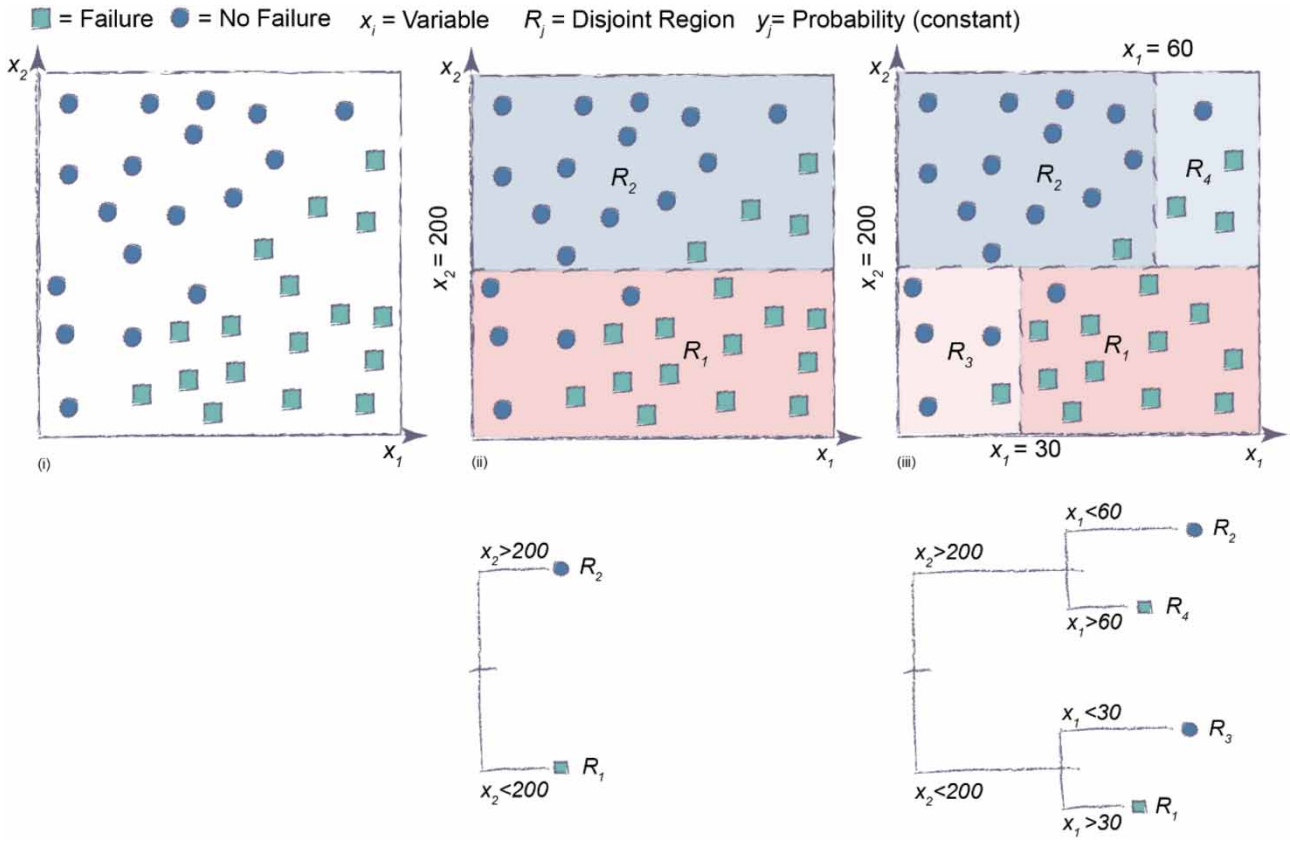

Figure 7 | An example of data partitioning by a classification decision tree: (i) shows the two-dimensional data space; (ii) shows the first condition for splitting the data by variable x_2 at 200, where disjoint region R_1 is $\leq$200 and disjoint region R_2 >200. This creates two additional nodes for further conditions; (iii) shows the second condition for splitting data by variable x_1 at both 30 and 60 to create two more disjoint regions R_3 and R_4 (adapted from Soguero-Ruiz *et al.* (2020)).

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

$\hat{f}(x) = \sum_{j=1}^{J} c_R I(x \in R_j)$, where I is an indicator function, equal to 1 if the condition is true (failure) or 0 otherwise (non-failure). A simple model like a constant c_R is applied to each partitioned region that determines the probability in that region.

The ensemble method fits several trees, improving weak classifiers through bagging or boosting and averaging the results. Bagging decreases the variance by generating additional data from the training set, producing multiple original datasets. Bagging is illustrated in Figure 8, such that: $\hat{f}_{avg}(x) = 1/B \sum_{b=1}^{B} \hat{f}^b(x)$, where $\hat{f}^b(x)$ are the independent classifiers from B_i separate training sets run in parallel, and the average of the trees is returned (Winkler *et al.* 2018).

Boosting is conceptionally like bagging but uses the residuals of the previous tree to improve learning (Hastie *et al.* 2009). Taking a regression model as an example, gradient boosting aims to minimise the loss function in the existing collection of trees by adding, at each step, another tree that best reduces the loss function. The loss function is the residual of the response minus the fitted probability mean (Elith *et al.* 2008). The final regression gradient boosting model is depicted in Figure 9 and can be expressed as (Hastie *et al.* 2009) $\hat{f}_M(x) = \sum_{m=1}^{M} T(x; \gamma_m)$, where $\hat{f}_M(x)$ is the average of the tree values, M is the boosting iteration, T is a tree, and x is the multivariate argument characterised by a set of parameters γ.

Winkler *et al.* (2018) studied a variety of ensemble tree methods, applying the model to a medium-sized WDN in Austria of 851 km of individual pipes (39,637 pipes), with a failure record of >30 years. The predictions were made over a long timeframe (five years and ten years) and compared RF with decision trees, Adaboost and RUSboost. The results found the RUSboost boosting method to be marginally better than the RF bagging method (AUC of 0.93 and 0.92, respectively). However, all models showed excellent results, partly due to dividing the pipes into street sections (having more equal lengths of pipe), dealing with the imbalanced data for all models through sampling methods, and using stratified sampling to represent the materials in both training and test datasets. Snider & McBean (2018) proposed using xgboost, an extreme gradient boosting model not previously used in the literature, to predict time-to-failure for a 16,866 ductile iron water mains dataset, recorded between 1960 and 2017. The xgboost model is faster than other gradient boosting models, increasing its appeal to industry, yet the model requires extensive pre-processing of data, with categorical data requiring one-hot-encoding. The variables used in

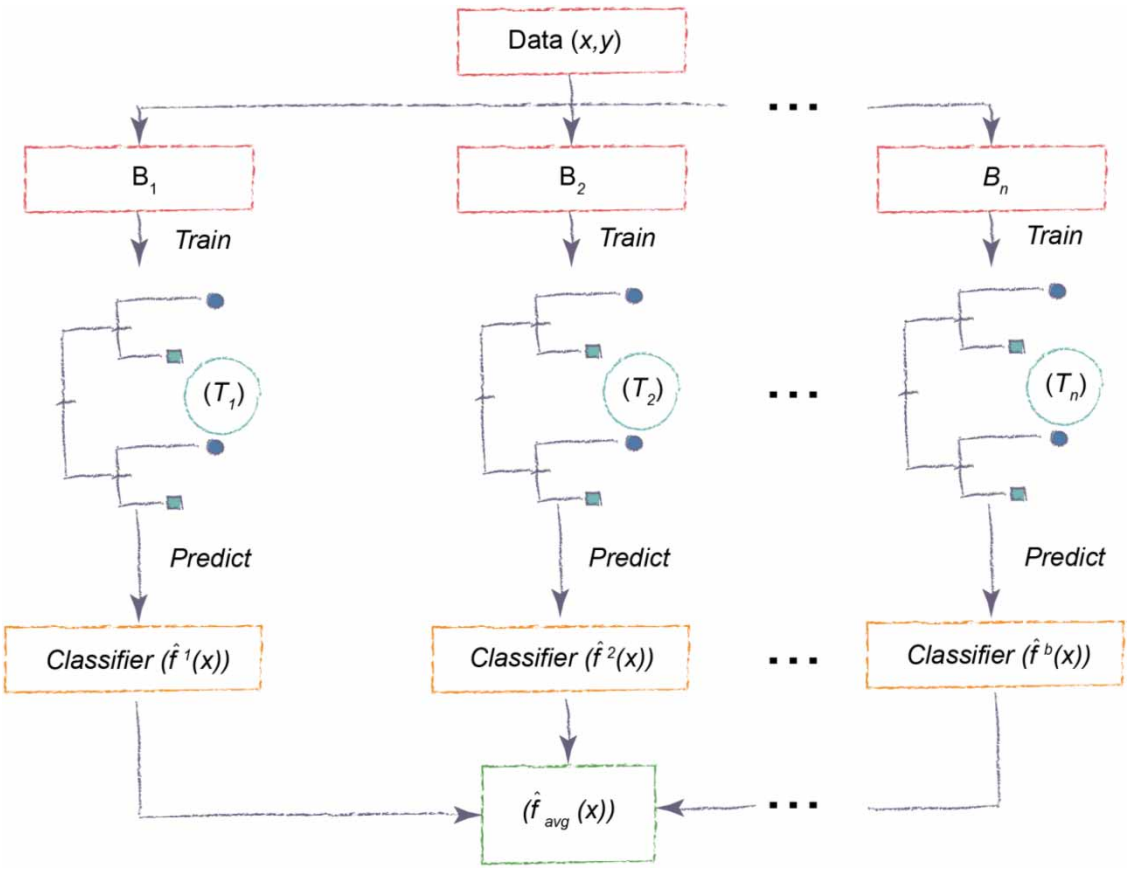

Figure 8 | Depiction of the bagging process, showing the parallel trees returning classifiers averaged for the final model.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

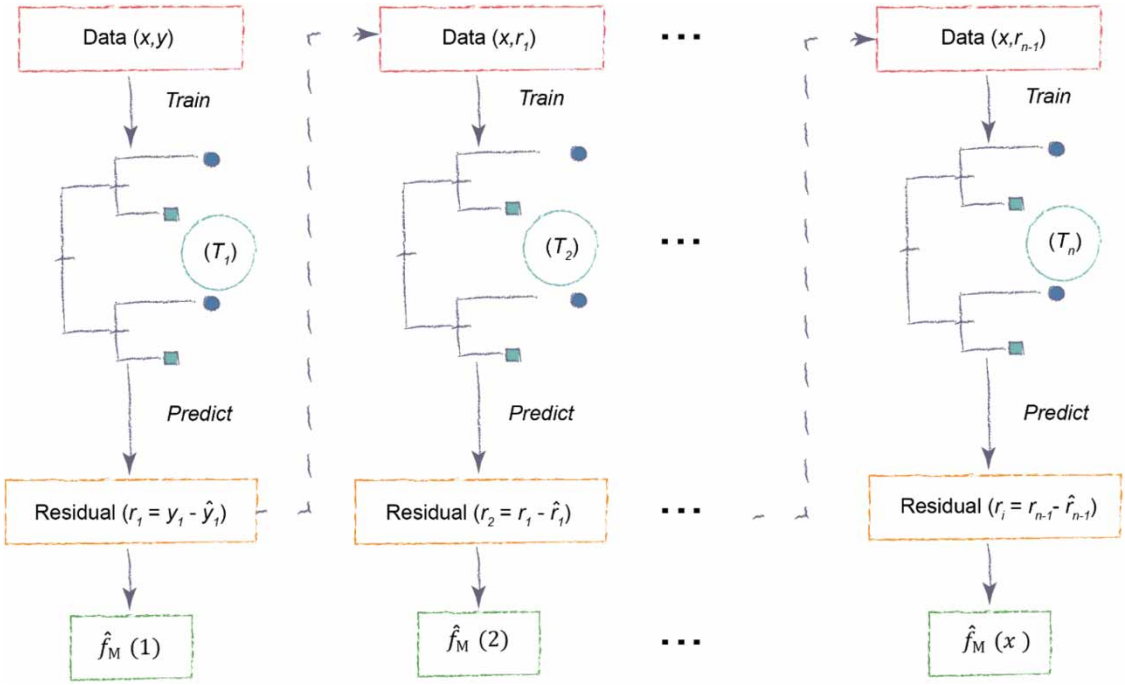

Figure 9 | Depiction of Boosting, like bagging, but using the residuals of the previous tree to improve the learning. The final model is the average of all trees.

this study included the year of construction, district, pipe material, length, diameter, soil, lining, and cathodic protection variables. The xgboost was compared with RF and ANN models using RMSE (test RMSE of 5.81, 5.90 and 7.32 respectively) and it was found that the xgboost had a 1.2% improvement over the RF model and a 25.9% improvement over the ANN model. In this study, the variable importance heavily leaned towards the previous failure date, whilst soils, district, and cathodic protection had only limited importance. Overall the authors found xgboost to be a reliable option, following with further research (Snider & McBean 2019). Chen *et al.* (2019) compared many models, including GLM, GAM, RF, GBT and Generalised Linear Mixed Models (GLMM), on pipes divided by a road network. The authors concluded that the GBT model performed best, returning the lowest Brier score (a more accurate measure for probabilistic predictions). The main limitations here are the monthly prediction interval and limited access to variables correlated to pipe failure. GBT models are useful, yet they can be computationally expensive and prone to overfitting if the hyperparameters are not tuned correctly. No calibration curve is provided. Giraldo-González & Rodríguez (2020) compared SVM, ANN, Bayes and GBT for a medium-sized WDN over a long prediction-interval (five years). Predicting the probability of individual pipe failures, the authors found the GBT to have the best performance (AUC 0.998 for AC pipes and 0.990 for PVC pipes), although again, all models showed excellent results. The limitation here is that the authors present the probability of failure and but did not provide a calibration curve to show how well the probabilities are observed with the outcomes in the sample test data.

The main limitation in most GBT studies is a lack of information on the model's performance, particularly the absence of a calibration plot or the Brier score to show the accuracy of the probabilities, which is helpful for models that are probabilistic classifiers. Further issues include the difficulty tuning hyperparameters and the risk of overfitting due to the additive process and the regularisation criteria used to control complexity (Wei *et al.* 2018).

4.3.6. Comments

Adopting machine learning for pipe failure modelling appears an obvious choice since these approaches are capable of modelling complex relationships between variables and complement the increased data collection over the past decade. Machine learning has been shown to improve predictive accuracy compared with deterministic and probabilistic approaches (Konstantinou & Stoianov 2020), although some studies show similar accuracy (Chen *et al.* 2019; Robles-Velasco *et al.* 2020), and when methodologies include large groups of pipes, deterministic models show comparable accuracy, as shown by Giraldo-González & Rodríguez (2020). Machine learning models can be used for regression, classification, or as probabilistic classifiers.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Supervised machine learning models are preferred since they perform well on structured data and are interpretable through variable importance and partial dependency plots, amongst other widely used accuracy metrics. The predictive capability of machine learning models works best with larger datasets. However, there is a greater need for advanced pre-processing methods, such as data transformation, data infusion, encoding categorical data through one-hot-encoding or dummy coding and standardisation (e.g. maximum, and minimum values), which are essential since the models are often sensitive to outliers. Pre-processing methods can be extensive and can consume some 60%–80% of the effort (Kahn 2021). Cross-validation and hyperparameter tuning are essential for reducing overfitting and bias. Machine learning is a complex mathematical framework that could deter its use within industry.

Clustering is a straightforward process and is used to identify spatial relationships between failures. These relationships were used successfully by de Oliveira *et al.* (2011) to target areas for investigation. However, this approach is backwards-looking and cannot predict future failures. Nevertheless, spatial relationships are essential, and Chen & Guikema (2020) used the information as a variable in a dataset, suggesting beneficial results.

ANN is widely used for its flexibility. Firstly, the algorithm automatically adjusts the connection weights to reflect non-linearities in the data, and secondly, the ANN can be developed using multiple training algorithms. This flexibility continues through extensions of alternative algorithms to help improve results, such as the fuzzy logic used by Christodoulou & Deligianni (2010) and the neuro-fuzzy inference system used by Tabesh *et al.* (2009). The ANN framework can generalise well but is also prone to overfitting if not tuned correctly. Some studies have reported poor generalisation (Shirzad *et al.* 2014), but this may be the result of the limited data typically seen in pipe failure datasets, especially considering ANN ordinarily performs better on high-dimensional unstructured data. The complexity of the model structure is problematic for two reasons: the optimum structure of an ANN is identified *a priori* (i.e. inputs, hidden layers and data transformation within the activation units), such that a time-consuming trial-and-error approach is typically required, and knowledge of the weight terms or bias is generally unknown, which limits interpretability (Xu *et al.* 2011).

EPR models provide transparent and well-structured relationships from a simple, easy-to-use formula that water companies can apply. Unlike other machine learning models, EPR has few parameters to tune, can include expert knowledge and is reasonably fast. However, EPR, like other machine learning models, is computationally expensive and prone to overfitting, resulting in poor accuracy and overestimating failures in pipes with no recorded failures (Berardi *et al.* 2008). Another major disadvantage is that only a single output can be achieved since the model is regression, and therefore works best on groups of pipes to return the failure rates (Giraldo-González & Rodríguez 2020).

SVM is a flexible model that can be a regression or a probabilistic classifier and improves linear models by modelling linear relationships with a single hyperplane. Non-linear relationships can be implicitly modelled using a kernel function, which can be difficult. This is advantageous since the approach is efficient with high-dimensional data. The algorithm fails to cope with noisy data where the clear separation between classifications is not easily defined; poor quality has been observed in pipe failure data (Debón *et al.* 2010; Kabir *et al.* 2016; Tang *et al.* 2019). Considering this, SVM is often outperformed by different machine learning models (Giraldo-González & Rodríguez 2020) and also simplistic mathematical frameworks like the logistic regression model (Robles-Velasco *et al.* 2020), which is perhaps why the model is not often used.

Ensemble tree methods consistently outperform other machine learning models, as found in several comparison studies (Chen *et al.* 2019; Giraldo-González & Rodríguez 2020). Bagging and boosting are two approaches commonly used in the literature, and a comparison of different methods by Winkler *et al.* (2018) has shown that gradient boosting is statistically more accurate. All the methods in this study showed good results predicting the probability of failure, although accuracy measures, such as calibration plots or Brier scores, were not presented. The methodology for many ensemble studies has included stratification when dividing testing and training datasets and predicting over long periods, typically five years, adding to the improved accuracy (Winkler *et al.* 2018; Giraldo-González & Rodríguez 2020). Another advantage of ensemble methods is that data normalisation is not required, making the pre-processing slightly easier. Table 3 shows a summary of the main machine learning models discussed.

5. SUMMARY AND STATISTICAL MODEL DECISION AID

This review has focussed on providing an overview of the standard methods and models used to predict pipe failures. The evolution of pipe failure models is clear, from simplistic single-variate regression models at a network scale to the use of

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Table 3 | Summary of the main machine learning pipe failure models

Author	Model[a]	Network size	Failure history	Pipe materials[b]	Variables	Spatial level	Response
Christodoulou *et al.* (2003)	ANN Fuzzy logic	365 km	1982–2002	AC, CI, DI	Pipe-intrinsic Environmental	Pipe groups	Number of failures (inter-arrival time)
Berardi *et al.* (2008)	EPR	172 km (3,669 pipes)	1986–1999	–	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Tabesh *et al.* (2009)	ANN ANFIS	580 km	–	AC, CI	Pipe-intrinsic Operational	–	Failure rate
Christodoulou & Deligianni (2010)	ANN Fuzzy logic	365 km 795 km	10 years	CI, Plastic, ST	Pipe-intrinsic	Individual pipes	Risk of failure
Jafar *et al.* (2010)	ANN	162 km (4,862 pipes)	1991–2004	Cement, CI, DI, Plastic	Pipe-intrinsic Environmental Operational	Individual pipes	Number of failures
Nishiyama & Filion (2013)	ANN	670 km	1998–2011	CI	Pipe-intrinsic Environmental	Pipe groups	Number of failures
Asnaashari *et al.* (2013)	ANN	784 km	1959–2004	CI, Concrete, DI, AC, PVC	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Harvey *et al.* (2014)	ANN	1,021 km (6,346 pipes)	1962–2005	AC, CI, DI, PVC	Pipe-intrinsic Environmental	Individual pipes	Time-to-failure (inter-arrival time)
Laucelli *et al.* (2014)	EPR	679 km (6,879 pipes)	1962–2003	CI	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Shirzad *et al.* (2014)	ANN SVR	580 km 140 km	1 year	AC, PE	Pipe-intrinsic Operational	Pipe groups	Failure rate
Farmani *et al.* (2017)	EPR	300.63 km (7,987 pipes)	–	CI	Pipe-intrinsic Environmental	Pipe groups	Number of failures
Kakoudakis *et al.* (2017)	EPR	–	–	AC, CI, DI, PE, PVC	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Kakoudakis *et al.* (2018)	ANN EPR	647 km (18,872 pipes)	2003–2013	CI	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Winkler *et al.* (2018)	Adaboost RUSboost RF GBT	851 km (39,637)	>30 years	CI, DI, PE, ST	Pipe-intrinsic Operational	Individual pipes	Probability of failure
Snider & McBean (2018)	GBT (xgboost) RF ANN	3,042 km	1960–2017	Cement, CI, DI, PVC	Pipe-intrinsic Environmental Operational	Individual pipes	Time-to-failure
Sattar *et al.* (2019)	Extreme Learning Machine (ELM) ANN SVM Non-linear regression	>1,000 km (6,342 pipes)	From 1962	AC, CI, DI	Pipe-intrinsic Environmental	Individual pipes	Time-to-failure

(*Continued.*)

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Table 3 | Continued

Author	Model[a]	Network size	Failure history	Pipe materials[b]	Variables	Spatial level	Response
Kutyłowska (2019)	SVM	17 km 14 km	2008–2014	CI, PE, PVC	Pipe-intrinsic Operational	Individual pipes	Failure rate
Chen *et al.* (2019)	GAM RF DT GBT GLMM	–	2010–2014	–	Environmental Census	Individual pipes (road segment) Pipe group (census track)	Probability of failure
Giraldo-González & Rodríguez (2020)	Linear, Poisson, and EPR GBT, Bayes, SVM, ANN	652 km (20,793 pipes)	2015–2017	AC, CI, DI, PE, PVC	Pipe-intrinsic Environmental Operational	Pipe groups Individual pipes	Failure rates Probability of failure
Robles-Velasco *et al.* (2020)	SVM Linear	3,800 km	2012–2018	Cements, metal, plastics	Pipe-intrinsic Operational	Individual pipes	Probability of failure
Chen & Guikema (2020)	GLM GAM RF GBT	680 km (12,092 pipes)	2008–2017	CI, other	Census Pipe-intrinsic Environmental Operational	Individual pipes	Number of failures
Aslani *et al.* (2021)	RF GBT Multivariate adaptive regression splines ANN	76,000 pipes	2015–2020	CI, DI, PE, PVC	Pipe-intrinsic Environmental Operational	Individual pipes	Failure rate

–, Unavailable information.

[a]ANN, Artificial Neural Network; DT, Decision Tree; EPR, Evolutionary Polynomial Regression; GAM, Generalised Additive Model; GBT, Gradient Boosting Tree; GLM, Generalised Linear Model; GLMM, Generalised Linear Mixed Effect Model; RF, Random Forest; SVM, Support Vector Machine.

[b]AC, asbestos cement; CI, cast iron; DI, ductile iron; HY, hyprescon; PE, polyethylene; PVC, polyvinyl chloride; ST, steel.

more complex machine learning, able to accommodate large multivariate datasets and finer individual pipe-scale methods. Each category of statistical model has advantages and disadvantages, detailed in Table 4.

Static variables are fundamental to pipe failure models, and throughout the literature, pipe-intrinsic factors such as pipe material, pipe age, pipe length, and the number of failures are consistently among the most significant. The temporal scale of predicting pipe failures has moved towards predicting shorter time scales, motivated by the need to assist operational management. Shifting towards predicting failures over shorter time intervals can use time-dynamic variables that capture the effects of seasonal or annual variations (and potentially future climatic variations). However, modelling time-dependent dynamic variables is more complex, especially for models that attempt to predict response for individual pipes, because pipe failures are rare and result in imbalanced data. In this scenario, the assumptions of the Poisson distribution are not met; therefore, predicting the number of failures or failure rate has resulted in poor accuracy. Predicting the probability of failure or the time-to-failure is reported as more useful for individual pipes and often provides enough information for decision-makers. Zero-inflated models may be used, but these have not necessarily proved to be more accurate. Alternatively, models should group pipes for greater accuracy if short prediction intervals are necessary. The Poisson distribution can return the number of failures or failure rate if the data is such that enough failures are captured by pipe group for statistical significance. However, in such instances, pipes that regularly fail may also be included in a group with pipes that never fail, distorting the true failure rate for many pipes. Therefore, grouping must be considered carefully, with some authors choosing census-level groups to provide more accurate predictions. Concisely, time-dependent dynamic variables can be included in pipe failure models, though the statistical analysis may prove to be more challenging.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Table 4 | Advantages and disadvantages of deterministic, probabilistic and machine learning methods

Category	Advantages	Disadvantages
Deterministic	• Simplistic mathematical frameworks. • Easy to implement within industry. • Easy to develop and quick to run. • Interpretable results. • Easy to develop and interpret. • Computationally efficient.	• It cannot account for randomness. • Does not easily handle large complex datasets. • Cannot be used to return the probability of failure. • Not accurate for individual pipe models. • Not flexible (cannot be tuned)
Probabilistic	• Survival analysis can predict different outcomes and cope with left-truncated data. • Can account for randomness. • The probability of failure is more accurate for individual pipe models. • Bayes models can incorporate expert knowledge.	• Some models have complex mathematical frameworks. • Not easy for water companies to implement as it would require expert, specialised knowledge. • Survival analysis requires extensive failure records. • The probability of failure is not always required.
Machine Learning	• Can describe complex relationships for large datasets. • Generally, it improves accuracy for individual pipe models. • Can be used to return failure rate, time-to-failure and as probabilistic classifiers. • Is flexible and can be tuned using hyperparameters.	• Requires complex mathematical frameworks. • Not easy for water companies to implement. It would require expert knowledge. • Computationally expensive to run and during pre-processing. • Prone to overfitting. • Results are harder to interpret.

The most statistically accurate temporal scale for model accuracy on individual pipes remains long-term prediction intervals, annual or greater, and in many recent machine learning efforts, individual pipes are predicted over five or more years (Nishiyama & Filion 2013; Harvey *et al.* 2014; Snider & McBean 2019; Robles-Velasco *et al.* 2020). Long-term prediction intervals show favourable accuracy where enough failures accumulate for statistical significance but at a loss of inter-annual information. Survival analysis lends itself well to long-term planning because there are options to predict inter-arrival times of failures. Advances in survival analysis have shown valuable results based on the Weibull distribution used for time-to-first failure and the exponential distribution for subsequent failures. Typically, a long pipe failure dataset is required. Alternatively, Bayesian models return the probability of failure and accommodate explicit expert domain knowledge using prior distributions. The graphical nature of Bayesian models provides a useful means of viewing the interactions between failures and the variables. If the objective is to analyse the results, then Bayesian networks are appropriate due to their interpretability. Yet, a lack of failure data means the influence of prior information is greater, resulting in bias introduced from opinions.

Machine learning models are not strictly probabilistic but can be termed probabilistic classifiers and have also proved to be useful when predicting the probability of failure for individual pipes over long periods (Winkler *et al.* 2018). Supervised models are the obvious choice for the structured nature of pipe failure datasets, and comparing machine learning models, the gradient boosting model has shown great promise in terms of accuracy (Giraldo-González & Rodríguez 2020), yet SVM and ANN have also shown useful results (Robles-Velasco *et al.* 2020). There is a level of complexity in machine learning that can deter its use in industry, and therefore very few models have transferred into industry working models. The complexity is associated with pre-processing, often, extensive datasets since the models are more sensitive to outliers; one-hot-encoding or dummy coding and standardisation (e.g., maximum and minimum values) are two examples. Tuning hyperparameters using cross-validation can be difficult but is essential to avoid overfitting. However, machine learning models are still subject to the same difficulties when considering the spatiotemporal scale and the limited number of failures observed in pipe failure datasets. Classification models predicting probability can use various sampling techniques (e.g., under- or oversampling) to balance the data, yet there are limitations to these techniques, and the results are not always favourable, nor do they necessarily improve the model accuracy (Fan *et al.* 2022).

Choosing the correct model is a complex process and can directly impact the usefulness of the response for decision-makers. No single model is superior, and so the choice of model should be one that carefully considers four key factors: (1) the type of variables available, (2) the time interval for management interventions, (3) the spatial level of the model (e.g., pipes or grouped pipes), and (4) the response type and level of inference required. Practitioners may find it useful to understand the most appropriate model; thus, Figure 10 shows a flow diagram devised to map these general assumptions, leading to a model selection that

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

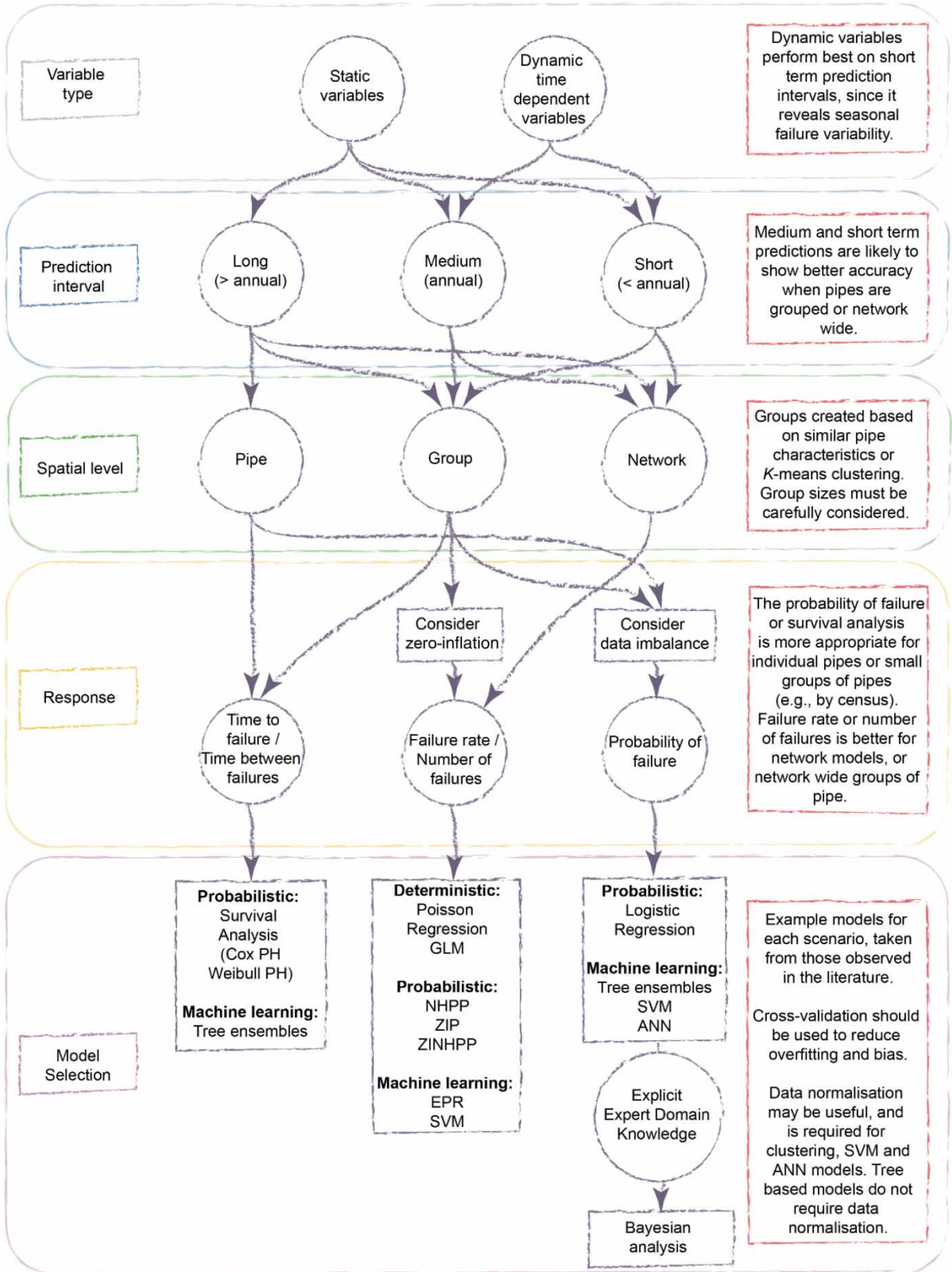

Figure 10 | Prediction model decision aid to determine the most useful approach in different decision-making contexts, considering four important key factors: (1) the type of variables available, (2) the time interval for management interventions, (3) the spatial level of the model, and (4) the response type and level of inference required.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

is likely to be the most useful in different decision-making contexts. The model shows five steps, with the first four steps working through the key factors and the final step suggesting potentially suitable statistical models.

6. CONCLUSIONS

From the literature, several general conclusions are deduced:

- There is much discussion on the benefits of more complex models, such as Bayesian and machine learning; however, these models are still subject to poor and limited data. Considering this, simple models are often useful and more intuitive.
- Early failure models separated pipe materials for more accurate results due to their unique behaviour within environments. Yet, recent literature has shown improved results for global models, where failure can be accumulated to avoid non-convergence due to too few failures. Using stratified sampling, the global model ensures representative materials are found in both the training and test datasets. This benefits from developing a single model instead of multiple models for various materials, saving time and providing a denser dataset for modelling.
- Data quality is generally poor, yet only a handful of studies explicitly identify data quality as a limiting issue. Furthermore, many studies do not include clarified assumptions or study limitations; for example, the approach to pre-processing the bursts data is often unknown, and yet there can be several issues in locating pipe failures accurately along with a network. These issues should be detailed to provide a better understanding of the limitations.
- It has been acknowledged that failure records collected by water companies are typically short (Yamijala *et al.* 2009; Chen *et al.* 2019). From the studies used here (noting this is not an exhaustive list and that not all literature had pipe failure record length), some records are as short as three years; however, approximately 60% had failure records greater than ten years, and 35% greater than 30 years. The studies with longer failure records report more useful results; however, it is unclear if the length of failure records or the length of the prediction interval improves the results; perhaps a combination of both.
- Water utilities should endeavour to collect regular and consistent data in such a manner to facilitate long-term studies and modelling efforts.
- Some studies do not demonstrate how the reported model can be transferred into a practical tool and utilised by water companies. Furthermore, most of the models were not implemented in ways whereby the outcomes could be easily integrated into a GIS framework for visualisation and further spatial analysis.
- Other than clustering algorithms that are not strictly failure models, there is a limited scope to address the spatiotemporal relationship between pipe failures. More research could be usefully conducted in this area, especially incorporating the use of cluster analysis.
- Limited studies have been undertaken on large networks, and many studies focus efforts in urban areas, typically cities. More studies could be completed using rural areas or a combination of both.
- Few studies use a comprehensive list of pipe-intrinsic, environmental, and operational covariates, which shows a lack of available data rather than shortfalls in the studies themselves.
- Few studies include data from decommissioned pipes, which would increase the dataset and number of failures to train the models.

The following is a list of recommendations related to improving the practice of pipe failure modelling for future methodologies:

- Data is essential to modelling failures correctly. At a minimum, pipe-intrinsic data should be collected. Further research on census data, environmental and socio-economic factors, could provide interesting insight.
- Data collection and management protocols are needed to improve data quality and quantity.
- Additional research is required on effective data collection and information extraction from existing data.
- Data collection and temporal and spatial scales should be tied to the model that is used. Further research is needed to establish appropriate pipe groupings that improve accuracy whilst describing the influences of more localised conditions.

ACKNOWLEDGEMENTS

This work was supported by the UK Natural Environment Research Council [NERC Ref: NE/M009009/1] Anglian Water plc, who had no role in this study, and the participants. The authors are grateful for their support.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

COMPETING INTERESTS

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish.

AUTHOR CONTRIBUTIONS

Neal Barton: Conceptualisation, Methodology, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualisation and Project Administration. Stephen Hallett: Writing – Review and Editing, Supervision, Funding Acquisition. Simon Jude: Review, Supervision, Funding Acquisition. Trung Hieu Tran: Review and Editing.

DATA AVAILABILITY STATEMENT

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

REFERENCES

Alvisi, S. & Franchini, M. 2010 Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system. *Civ. Eng. Environ. Syst.* **27**, 1–22. https://doi.org/10.1080/10286600802224064.

Andreou, S. A., Marks, D. H. & Clark, R. M. 1987a A new methodology for modelling break failure patterns in deteriorating water distribution systems: theory. *Adv. Water Resour.* **10**, 2–10. https://doi.org/10.1016/0309-1708(87)90002-9.

Andreou, S. A., Marks, D. H. & Clark, R. M. 1987b A new methodology for modelling break failure patterns in deteriorating water distribution systems: applications. *Adv. Water Resour.* **10**, 11–20. https://doi.org/10.1016/0309-1708(87)90003-0.

Aslani, B., Mohebbi, S. & Axthelm, H. 2021 Predictive analytics for water main breaks using spatiotemporal data. *Urban Water J.* **18** (6), 433–448. https://doi.org/10.1080/1573062X.2021.1893363.

Asnaashari, A., McBean, E. A., Shahrour, I. & Gharabaghi, B. 2009 Prediction of watermain failure frequencies using multiple and Poisson regression. *Water Supply* **9**, 9–19. https://doi.org/10.2166/ws.2009.020.

Asnaashari, A., McBean, E. A., Gharabaghi, B. & Tutt, D. 2013 Forecasting watermain failure using artificial neural network modelling. *Can. Water Resour. J.* **38**, 24–33. https://doi.org/10.1080/07011784.2013.774153.

Barton, N. A., Farewell, T. S., Hallett, S. H. & Acland, T. F. 2019 Improving pipe failure predictions: factors affecting pipe failure in drinking water networks. *Water Res.* **164**, 114926. https://doi.org/10.1016/j.watres.2019.114926.

Barton, N. A., Farewell, T. S. & Hallett, S. H. 2020 Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water* **3**, 31. https://doi.org/10.1038/s41545-020-0077-3.

Barton, N. A., Hallett, S. H. & Jude, S. R. 2021 The challenges of predicting pipe failures in clean water networks: a view from current practice. *Water Supply* **22**, 527–541. https://doi.org/10.2166/ws.2021.255.

Berardi, L., Giustolisi, O., Kapelan, Z. & Savic, D. A. 2008 Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinformatics* **10**, 113–126. https://doi.org/10.2166/hydro.2008.012.

Carrión, A., Solano, H., Gamiz, M. L. & Debón, A. 2010 Evaluation of the reliability of a water supply network from right-censored and left-truncated break data. *Water Resour. Manage.* **24**, 2917–2935. https://doi.org/10.1007/s11269-010-9587-y.

Chen, T. Y., Beekman, J. A. & Guikema, S. D. 2017 Drinking water distribution systems asset management: statistical modelling of pipe breaks. In: *Pipelines 2017* (A. Pridmore & J. Geisbush, eds), ASCE, Reston, VA, USA, pp. 173–186.

Chen, T. Y.-J. & Guikema, S. D. 2020 Prediction of water main failures with the spatial clustering of breaks. *Reliab. Eng. Syst. Saf.* **203**, 107108. https://doi.org/10.1016/j.ress.2020.107108.

Chen, T. Y.-J., Beekman, J. A., Guikema, S. D. & Shashaani, S. 2019 Statistical modeling in absence of system specific data: exploratory empirical analysis for prediction of water main breaks. *J. Infrastruct. Syst.* **25**, 04019009. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000482.

Christodoulou, S. & Deligianni, A. 2010 A neurofuzzy decision framework for the management of water distribution networks. *Water Resour. Manag.* **24**, 139–156. https://doi.org/10.1007/s11269-009-9441-2.

Christodoulou, S., Aslani, P. & Vanrenterghem, A. 2003 A risk analysis framework for evaluating structural degradation of water mains in urban settings, using neurofuzzy systems and statistical modeling techniques. In: *World Water & Environmental Resources Congress 2003* (P. Bizier & P. DeBarry, eds), ASCE, Reston, VA, USA. https://doi.org/10.1061/40685(2003)134.

Clark, R. M., Stafford, C. L. & Goodrich, J. A. 1982 Water distribution systems: a spatial and cost evaluation. *J. Water Resour. Plann. Manage.* **108**, 243–256.

Cox, D. R. 1972 Regression models and life-tables. *J. R. Stat. Soc. Ser. B* **34**, 187–202.

Deadman, C. 2010 *Strategic Asset Management: The Quest for Utility Excellence*. Matador, Leicester, UK.

Debón, A., Carrión, A., Cabrera, E. & Solano, H. 2010 Comparing risk of failure models in water supply networks using ROC curves. *Reliab. Eng. Syst. Saf.* **95**, 43–48. https://doi.org/10.1016/j.ress.2009.07.004.

De Oliveira, D. P., Garrett, J. H. & Soibelman, L. 2011 A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Adv. Eng. Informatics* **25**, 380–389. https://doi.org/10.1016/j.aei.2010.09.001.

Dijkstra, E. W. 1959 A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271. https://doi.org/10.1007/BF01386390.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Dridi, L., Mailhot, A., Parizeau, M. & Villeneuve, J.-P. 2005 A strategy for optimal replacement of water pipes integrating structural and hydraulic indicators based on a statistical water pipe break model. In: *Proceedings of the 8th International Conference on Computing and Control for the Water Industry*, Exeter, UK, pp. 65–70.

Economou, T., Kapelan, Z. & Bailey, T. 2008 A zero-inflated Bayesian model for the prediction of water pipe bursts. In: *Water Distribution Systems Analysis 2008* (K. Van Zyl, ed.), ASCE, Reston, VA, USA. https://doi.org/10.1061/41024(340)61.

Economou, T., Kapelan, Z. & Bailey, T. C. 2012 On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. *J. Hydroinformatics* **14**, 872–883. https://doi.org/10.2166/hydro.2012.144.

Elith, J., Leathwick, J. R. & Hastie, T. 2008 A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Fan, X., Wang, X., Zhang, X. & Yu, X. 2022 Machine learning based water pipe failure prediction: the effects of engineering, geology, climate and socio-economic factors. *Reliab. Eng. Syst. Saf.* **219**, 108185. https://doi.org/10.1016/j.ress.2021.108185.

Farmani, R., Kakoudakis, K., Behzadian, K. & Butler, D. 2017 Pipe failure prediction in water distribution systems considering static and dynamic factors. *Procedia Eng.* **186**, 117–126. https://doi.org/10.1016/j.proeng.2017.03.217.

Fitchett, J. C., Karadimitriou, K., West, Z. & Hughes, D. M. 2020 Machine learning for pipe condition assessments. *J. AWWA* **112** (5), 50–55. https://doi.org/10.1002/awwa.1501.

Francis, R. A., Guikema, S. D. & Henneman, L. 2014 Bayesian Belief Networks for predicting drinking water distribution system pipe breaks. *Reliab. Eng. Syst. Saf.* **130**, 1–11. https://doi.org/10.1016/j.ress.2014.04.024.

Getis, A. & Ord, J. K. 1992 The analysis of spatial association by use of distance statistics. *Geogr. Analysis* **24**, 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Giraldo-González, M. M. & Rodríguez, J. P. 2020 Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. *Water* **12**, 1153. https://doi.org/10.3390/w12041153.

Giustolisi, O. & Savic, D. A. 2006 A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinformatics* **8**, 207–222. https://doi.org/10.2166/hydro.2006.020b.

Goulter, I. C. & Kazemi, A. 1988 Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Can. J. Civ. Eng.* **15**, 91–97. https://doi.org/10.1139/l88-010.

Harvey, R., McBean, E. A. & Gharabaghi, B. 2014 Predicting the timing of water main failure using artificial neural networks. *J. Water Resour. Plann. Manage.* **140**, 425–434. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000354.

Hastie, T., Tibshirani, R. & Friedman, J. 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn*. Springer, New York, USA. https://doi.org/10.1007/978-0-387-84858-7.

Jafar, R., Shahrour, I. & Juran, I. 2010 Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Math. Comput. Modell.* **51**, 1170–1180. https://doi.org/10.1016/j.mcm.2009.12.033.

Kabir, G., Tesfamariam, S., Loeppky, J. & Sadiq, R. 2016 Predicting water main failures: a Bayesian model updating approach. *Knowledge-Based Syst.* **110**, 144–156. https://doi.org/10.1016/j.knosys.2016.07.024.

Kabir, G., Balek, N. B. C. & Tesfamariam, S. 2018 Consequence-based framework for buried infrastructure systems: a Bayesian belief network model. *Reliab. Eng. Syst. Saf.* **180**, 290–301. https://doi.org/10.1016/j.ress.2018.07.037.

Kahn, C. 2021 The value proposition for likelihood-of-failure modeling. *J. AWWA* **113** (1), 30–38. https://doi.org/10.1002/awwa.1649.

Kakoudakis, K., Behzadian, K., Farmani, R. & Butler, D. 2017 Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with *K*-means clustering. *Urban Water J.* **14**, 737–742. https://doi.org/10.1080/1573062X.2016.1253755.

Kakoudakis, K., Farmani, R. & Butler, D. 2018 Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. *J. Hydroinformatics* **20**, 1191–1200. https://doi.org/10.2166/hydro.2018.152.

Kalogirou, S. A. 2004 Neural network modeling of energy systems. In: *Encyclopedia of Energy* (C. J. Cleveland, ed.), Elsevier, New York, USA, pp. 291–299. https://doi.org/10.1016/B0-12-176480-X/00536-2.

Kettler, A. J. & Goulter, I. C. 1985 An analysis of pipe breakage in urban water distribution networks. *Can. J. Civ. Eng.* **12**, 286–293. https://doi.org/10.1139/l85-030.

Kimutai, E., Betrie, G., Brander, R., Sadiq, R. & Tesfamariam, S. 2015 Comparison of statistical models for predicting pipe failures: illustrative example with the city of Calgary water main failure. *J. Pipeline Syst. Eng. Pract.* **6**, 04015005. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000196.

Kleiner, Y. & Rajani, B. 2000 Considering time-dependent factors in the statistical prediction of water main breaks. In: *American Water Works Association Infrastructure Conference (AWWA 2000)*, 12–15 March, Baltimore, MA, USA.

Kleiner, Y. & Rajani, B. 2001 Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* **3**, 131–150. https://doi.org/10.1016/S1462-0758(01)00033-4.

Kleiner, Y. & Rajani, B. 2002 Forecasting variations and trends in water-main breaks. *J. Infrastruct. Syst.* **8**, 122–131. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:4(122).

Kleiner, Y. & Rajani, B. 2010 I-WARP: Individual Water mAin Renewal Planner. *Drink. Water Eng. Sci.* **3**, 71–77. https://doi.org/10.5194/dwes-3-71-2010.

Kleiner, Y. & Rajani, B. 2012 Comparison of four models to rank failure likelihood of individual pipes. *J. Hydroinformatics* **14**, 659–681. https://doi.org/10.2166/hydro.2011.029.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Kleiner, Y., Sadiq, R. & Rajani, B. 2004 Modeling failure risk in buried pipes using fuzzy Markov deterioration process. In: *Pipelines 2004: What's on the Horizon*? (J. J. Galleher, Jr & M. T. Stift, eds), ASCE, Reston, VA, USA. https://doi.org/10.1061/40745(146)7.

Kleiner, Y., Nafi, A. & Rajani, B. 2010 Planning renewal of water mains while considering deterioration, economies of scale and adjacent infrastructure. *Water Supply* **10**, 897–906. https://doi.org/10.2166/ws.2010.571.

Konstantinou, C. & Stoianov, I. 2020 A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks. *Urban Water J.* **17**, 534–548. https://doi.org/10.1080/1573062X.2020.1800758.

Kutyłowska, M. 2015 Modelling of failure rate of water-pipe networks. *Period. Polytech. Civ. Eng.* **59**, 37–43. https://doi.org/10.3311/PPci.7541.

Kutyłowska, M. 2019 Forecasting failure rate of water pipes. *Water Supply* **19**, 264–273. https://doi.org/10.2166/ws.2018.078.

Laucelli, D., Rajani, B., Kleiner, Y. & Giustolisi, O. 2014 Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. *J. Hydroinformatics* **16**, 743–757. https://doi.org/10.2166/hydro.2013.082.

Le Gat, Y. & Eisenbeis, P. 2000 Using maintenance records to forecast failures in water networks. *Urban Water* **2**, 173–181. https://doi.org/10.1016/S1462-0758(00)00057-1.

Lin, P. & Yuan, X.-X. 2019 A two-time-scale point process model of water main breaks for infrastructure asset management. *Water Res.* **150**, 296–309. https://doi.org/10.1016/j.watres.2018.11.066.

Louppe, G. 2014 *Understanding Random Forests: From Theory to Practice*. PhD thesis, University of Liège, Liège, Belgium. https://arxiv.org/abs/1407.7502

Mailhot, A., Pelletier, G., Noël, J.-F. & Villeneuve, J.-P. 2000 Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: methodology and application. *Water Resour. Res.* **36**, 3053–3062. https://doi.org/10.1029/2000WR900185.

Motiee, H. & Ghasemnejad, S. 2019 Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. *Water Supply* **19**, 695–702. https://doi.org/10.2166/ws.2018.137.

Nishiyama, M. & Filion, Y. 2013 Review of statistical water main break prediction models. *Can. J. Civ. Eng.* **40**, 972–979. https://doi.org/10.1139/cjce-2012-0424.

Nishiyama, M. & Filion, Y. 2014 Forecasting breaks in cast iron water mains in the city of Kingston with an artificial neural network model. *Can. J. Civ. Eng.* **41**, 918–923. https://doi.org/10.1139/cjce-2014-0114.

Ofwat 2020 Innovation in the Water Sector. Available from: https://www.ofwat.gov.uk/regulated-companies/innovation-in-the-water-sector/.

Ponce Romero, J., Hallett, S. H. & Jude, S. 2017 Leveraging big data tools and technologies: addressing the challenges of the water quality sector. *Sustainability* **9**, 2160. https://doi.org/10.3390/su9122160.

Rajani, B. & Kleiner, Y. 2001 Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water* **3**, 151–164. https://doi.org/10.1016/S1462-0758(01)00032-2.

Rajani, B., Zhan, C. & Kuraoka, S. 1996 Pipe–soil interaction analysis of jointed water mains. *Can. Geotech. J.* **33**, 393–404. https://doi.org/10.1139/t96-061.

Rajani, B., Kleiner, Y. & Sadiq, R. 2006 Translation of pipe inspection results into condition ratings using the fuzzy synthetic evaluation technique. *J. Water Supply Res. Technol. – AQUA* **55**, 11–24. https://doi.org/10.2166/aqua.2005.069.

Rajani, B., Kleiner, Y. & Sink, J.-E. 2012 Exploration of the relationship between water main breaks and temperature covariates. *Urban Water J.* **9**, 67–84. https://doi.org/10.1080/1573062X.2011.630093.

Rezania, M. 2008 *Evolutionary Polynomial Regression Based Constitutive Modelling and Incorporation in Finite Element Analysis*. PhD thesis, University of Exeter, Exeter, UK.

Robles-Velasco, A., Cortés, P., Muñuzuri, J. & Onieva, L. 2020 Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliab. Eng. Syst. Saf.* **196**, 106754. https://doi.org/10.1016/j.ress.2019.106754.

Robles-Velasco, A., Muñuzuri, J., Onieva, L. & Rodríguez Palero, M. 2021 Trends and applications of machine learning in water supply networks management. *J. Ind. Eng. Manag.* **14**, 45–54. https://doi.org/10.3926/jiem.3280.

Rodríguez, G. 2007 Survival models. In: *Lecture Notes on Generalized Linear Models*. https://data.princeton.edu/wws509/notes/.

Røstum, J. 2000 *Statistical Modelling of Pipe Failures in Water Networks* Doctoral thesis,. Norwegian University of Science and Technology, Trondheim, Norway.

Sattar, A. M. A., Ertuğrul, Ö. F., Gharabaghi, B., McBean, E. A. & Cao, J. 2019 Extreme learning machine model for water network management. *Neural Comput. Appl.* **31**, 157–169. https://doi.org/10.1007/s00521-017-2987-7.

Scheidegger, A., Scholten, L., Maurer, M. & Reichert, P. 2013 Extension of pipe failure models to consider the absence of data from replaced pipes. *Water Res.* **47**, 3696–3705. https://doi.org/10.1016/j.watres.2013.04.017.

Scheidegger, A., Leitão, J. P. & Scholten, L. 2015 Statistical failure models for water distribution pipes – a review from a unified perspective. *Water Res.* **83**, 237–247. https://doi.org/10.1016/j.watres.2015.06.027.

Scholten, L., Scheidegger, A., Reichert, P. & Maurer, M. 2013 Combining expert knowledge and local data for improved service life modeling of water supply networks. *Environ. Modell. Software* **42**, 1–16. https://doi.org/10.1016/j.envsoft.2012.11.013.

Shamir, U. & Howard, C. D. D. 1979 An analytic approach to scheduling pipe replacement. *J. AWWA* **71** (5), 248–258. https://doi.org/10.1002/j.1551-8833.1979.tb04345.x.

Shirzad, A., Tabesh, M. & Farmani, R. 2014 A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. *KSCE J. Civ. Eng.* **18**, 941–948. https://doi.org/10.1007/s12205-014-0537-8.

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

Snider, B. & McBean, E. A. 2018 Improving time-to-failure predictions for water distribution systems using gradient boosting algorithm. In: *1st International WDSA/CCWI 2018 Joint Conference*, 23–25 July, Kingston, ON, Canada.

Snider, B. & McBean, E. A. 2019 Improving urban water security through pipe-break prediction models: machine learning or survival analysis. *J. Environ. Eng.* **146**, 04019129. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001657.

Snider, B. & McBean, E. A. 2020 Watermain breaks and data: the intricate relationship between data availability and accuracy of predictions. *Urban Water J.* **17**, 163–176. https://doi.org/10.1080/1573062X.2020.1748664.

Soguero-Ruiz, C., Mora-Jiménez, I., Mohedano-Munoz, M. A., Rubio-Sanchez, M., de Miguel-Bohoyo, P. & Sanchez, A. 2020 Visually guided classification trees for analyzing chronic patients. *BMC Bioinformatics* **21**, 92. https://doi.org/10.1186/s12859-020-3359-3.

St. Clair, A. M. & Sinha, S. 2012 State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models. *Urban Water J.* **9**, 85–112. https://doi.org/10.1080/1573062X.2011.644566.

Tabesh, M., Soltani, J., Farmani, R. & Savic, D. 2009 Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. *J. Hydroinformatics* **11**, 1–17. https://doi.org/10.2166/hydro.2009.008.

Tang, K., Parsons, D. J. & Jude, S. 2019 Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliab. Eng. Syst. Saf.* **186**, 24–36. https://doi.org/10.1016/j.ress.2019.02.001.

Wang, Y., Zayed, T. & Moselhi, O. 2009 Prediction models for annual break rates of water mains. *J. Perform. Constr. Facil.* **23**, 47–54. https://doi.org/10.1061/(ASCE)0887-3828(2009)23:1(47).

Watson, T. G., Christian, C. D., Mason, A. J., Smith, M. H. & Meyer, R. 2004 Bayesian-based pipe failure model. *J. Hydroinformatics* **6**, 259–264. https://doi.org/10.2166/hydro.2004.0019.

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M. & Zhao, X. 2018 A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **82**, 1027–1047. https://doi.org/10.1016/j.rser.2017.09.108.

Wilson, D., Filion, Y. & Moore, I. 2017 State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water J.* **14**, 173–184. https://doi.org/10.1080/1573062X.2015.1080848.

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W. & Tscheikner-Gratl, F. 2018 Pipe failure modelling for water distribution networks using boosted decision trees. *Struct. Infrastruct. Eng.* **14**, 1402–1411. https://doi.org/10.1080/15732479.2018.1443145.

Wols, B. A., Vogelaar, A., Moerman, A. & Raterman, B. 2019 Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* **19**, 404–416. https://doi.org/10.2166/ws.2018.085.

Xu, Q., Chen, Q. & Li, W. 2011 Application of genetic programming to modeling pipe failures in water distribution systems. *J. Hydroinformatics* **13**, 419–428. https://doi.org/10.2166/hydro.2010.189.

Yamijala, S., Guikema, S. D. & Brumbelow, K. 2009 Statistical models for the analysis of water distribution system pipe break data. *Reliab. Eng. Syst. Saf.* **94**, 282–293. https://doi.org/10.1016/j.ress.2008.03.011.

First received 3 August 2021; accepted in revised form 20 December 2021. Available online 19 January 2022

Downloaded from http://iwaponline.com/ws/article-pdf/22/4/3784/1040954/ws022043784.pdf
by CRANFIELD UNIVERSITY user

https://dspace.org/

An evolution of statistical pipe failure models for drinking water networks: a targeted review

Barton, Neal A.; Hallett, Stephen; Jude, Simon R.; Tran, Trung Hieu

2022-04-01
IWA Publishing
https://doi.org/10.2166/ws.2022.019
Barton NA, Hallett SH, Jude SR, Tran TH. (2022) An evolution of statistical pipe failure models for drinking water networks: a targeted review, Water Supply, Volume 22, Issue 4, April 2022, pp. 3784-3813

Downloaded from a DSpace repository. https://dspace.org/