Journal of Engineering and Applied Sciences 12 (12): 3163-3173, 2017
ISSN: 1816-949X

A Review of Probabilistic Modeling of Pipeline Leakage using Bayesian Networks

G.A. Ogutu, P.K. Okuthe and M. Lall
Department of Computer Science, Faculty of Information and Communication Technology,
Tshwane University of Technology, Private X680, 0001 Pretoria, South Africa

Abstract: The increasing amounts of pressure and threat on pipeline infrastructure consequently represent an elevation in the number of pipeline failures experienced. These failures are accompanied with extensive damage leading to environmental, social and economic stress to municipalities and water utilities. Respective managers are therefore pressured to put in place reliable maintenance and rehabilitation strategies in effort of minimizing losses. Prediction of potential mishap is one way through which instigation of planned rehabilitation may be upheld. However, this is challenging, thanks to inherent uncertainties. One effective way of handling uncertainty is through collection and combination of auxiliary information and knowledge which can be tackled using probabilistic models like Bayesian Networks (BNs). In this study, therefore we present comprehensive review of how probabilistic models have been applied in different ways to predict pipeline leakage; we identify various gaps presented by these models and finally we highlight the current state of research as far as leakage prediction is concerned. We also propose a recommendation for future research work.

Key words: Pipeline leakage, failure, prediction, distribution network, Bayesian networks

INTRODUCTION

Water distribution networks stand out as one of the most important and most expensive infrastructure assets to water utilities and municipalities (Kabir *et al*., 2015). They determine the utilitie's operational cost, service quality and highly uphold the foundation of public health concerns (Kabir *et al*., 2015; Kleiner and Rajani, 2000). These networks are anticipated to yield to failure given that they are continually exposed to internal and external degradation factors (Makar and Kleiner, 2000). As discussed by Kabir *et al*. (2015), Rajani and Kleiner (2001), the risk of failure is a blend of the probability and the severity of a number of different processes (in most cases, never comprehensively understood) that negatively affect the ability of pipelines to achieve a number of operational objectives set by utilities. In addition, the complexities of these processes make modeling of detection and prediction systems that are sensitive and accurate enough for considerable leak notification quite daunting (Rajani and Kleiner, 2001). Leakages are the most prevalent form of failure that affects pipelines (Kleiner and Rajani, 2000). A Leak, according to Poulakis *et al*. (2003), refer to an outflow or escape of water or any other fluid from a local point or position in a pipe, serving as proof of damage. They are unavoidable and can be noticed in more than one location simultaneously.

Leakages have the potential of causing extensive direct damage (in terms of destruction of road and building foundations, flooding and other associated liability costs like the cost of repairing damaged pipes and cost of water loss quantified in terms of the cost of raw water treatment) and indirect damage to the economy and the environment (Makar and Kleiner, 2000; Yamijala, 2007). Generally, Pipeline failures reduce network reliability (Tabesh *et al*., 2009) which in turn affects utilitie's ability to meet their objectives (Kabir *et al*., 2015). They also increase utility expenditure; for instance, it is reported that close to 80% of all the investment in a water facility is spent on the distribution network with at least 60% of these funds used on the piping system (Poulakis *et al*., 2003; Stone *et al*., 2002). In addition, nearly half of the funds allocated for the pipe system are used for maintenance and rehabilitation. Beside these, global reports (Stone *et al*., 2002) still indicate that water distribution systems are increasingly at risk of failure. These consequences have created a need to develop proper proactive assessment methods and tools that are built upon a combination of scientific approaches and human expertise for risk of failure evaluation (Kabir *et al*., 2015). The tools can assist municipality and utility management to bring together both long and short term management strategies that in the long run could be useful to them with regards to cost cutting and improved service delivery (Christodoulou *et al*., 2009; Kabir *et al*., 2015).

Corresponding Author: G.A. Ogutu, Department of Computer Science, Faculty of Information and Communication Technology, Tshwane University of Technology, Private X680, 0001 Pretoria, South Africa

Predictive modeling is fundamental for decision making as it facilitates future planning of network assessment based on current annotations (Kleiner and Rajani, 2001; Mailhot *et al*., 2003). The general goal in predictive modeling is to make use of readily available data about the current network state to develop techniques that would adequately use the same data to make intervention plans on the network before failure is experienced (Mailhot *et al*., 2003). A good number of models quantifying pipeline risk of failure and deterioration factors given a significant amount of historical data have been reported (Clair and Sinha, 2012; Kabir *et al*., 2015; Kleiner and Rajani, 2001; Rajani and Kleiner, 2001). Kleiner and Rajani (2001) clasified probabilistic models into two categories the probabilistic single-variate that are described to derive probabilistic measurements on grouped data and probabilistic multi-variate models that are said to consider more variables during analysis and acts on individual pipe basis. Gat and Eisenbeis (2000) on the other hand, presented a methodology for using maintenance records to forecast pipe failure. The procedure utilized the Weibull Proportional Hazard Model (WPHM) for analysis and failure occurrence was forecasted using different data inputs including length of pipe, pipe age, diameter, soil type, method of pipe assembly, traffic level and the kind of supply made by the pipe. Two cities were used as a case to compare the predicted failures of which in one of the cities, critics (Clair and Sinha, 2012) pointed out that all the predicted failures turned out to have been over estimated.

Mailhot *et al*. (2003) on the other hand presented a method that described the Probability Density Functions (PDF) defining the time difference between consecutive breaks in a pipe. The model assumed that pipe ageing takes place in two distinct processes: non-exponential, characterized by uneven distribution of the duration between failures; and exponential, characterized by a uniform distribution of the time difference between failures. However, critics Andreou *et al*. (1987) argued that failure occur as a result of a combination of totally different aspects which cannot uniformly act on a pipe. Therefore, exponential increase of failure over time may not be possible. Kabir *et al*. (2015) also pointed out that there exists a widespread acknowledgement that the relationships among pipe failure parameters are not linear and therefore, require more sophisticated analytic procedures and not simple mathematical models.

Christodoulou *et al*. (2009) applied Artificial Neural Networks (ANN) combined with fuzzy logic (Neurofuzzy) to analyze pipeline risk of failure and to develop a Decision Support System (DSS) in urban water zones. In their study, the city of Limassol, Cyprus and New York City were used as the case study. The study reported that pipe breakage history, pipe age, material and pipe diameter were the most significant factors that affected failure in the studied regions. Al-Barqawi and Zayed (2006) also proposed a model that utilized ANN to rate the condition of underground pipeline networks. ANN was used to develop the procedures for prioritization of network rehabilitation. Among the results, they concluded that the highest contributor to pipeline failure was the pipe break rate, followed by pipe age. To further understand and gather together some of the different models that tackle predictive modeling of pipeline failure, reference is made to a number of reviews that have been conducted to investigate pipeline failure. Clair and Sinha (2012), Kleiner and Rajani (2001) conducted different comprehensive reviews of general models for prediction of pipe condition, deterioration and failure rates while Colombo *et al.* (2009) and Engelhardt *et al.* (2000) conducted selective reviews about the same models. A summary of some of the models highlighted in the respective reviews is given in Table 1.

With regards to the articles cited in Table 1, predictive models are categorized under statistical and physical models (Clair and Sinha, 2012; Kleiner and Rajani, 2001). Statistically, they make use of available historical failure data to identify different failure patterns (Kleiner and Rajani, 2001). Physical models however are more geared towards analyzing the physical processes that lead to pipe failure (Rajani and Kleiner, 2001). Nonetheless, majority of these models assume that causal parameters are in one way or another independent, even though in reality, they are somehow connected to one another. Therefore, a wholesome outlook that presents the interconnection of all the different causal events is necessary for identification of the relationships between these events. One way of achieving this is through the use of Network Based Models (NBMs). Kabir *et al.* (2016) gives a rather clear discussion of a group of Network Based Modeling techniques.

These techniques include Cognitive Maps or Fuzzy Cognitive Maps (CM/FCM), Analytical Network Process (ANP), Credal Network (CNs), Bayesian Belief Networks (BBNs), Fuzzy Rule-Based Models (FRBM) and Artificial Neural Networks (ANN). Network models are considered to be quite effective for predictive modeling because of their ability to handle inherent data uncertainties. A comparison of their ability in uncertainty management while modeling is also performed in the said discussion. Additionally, discussions performed by the different researchers also bring out attention to a number of different characteristics of probabilistic models (discussed in the subsequent section) which demonstrates their relevance in describing pipeline failure.

Table 1: General failure prediction models modified after Clair and Sinha (2012)

Article reference	Title	Type of prediction	Method used
Predictive models			
Kleiner and Rajani (1999)	Using limited data to assess future needs	Break rate	Gumbel, Weibull andherz distribution
Gat and Eisenbeis (2000)	Using maintenance records to forecast failures in water networks	Failure rate	Weibull Proportional Hazard Model (WPHM), Monte Carlo simulation
Mailhot *et al.* (2003)	Optimal replacement of water pipes	Time to failure	Regression analysis
Christodoulou *et al.* (2003)	A risk analysis framework for evaluating structural degradation of water mains in urban settings, using neurofuzzy systems and statistical modeling techniques	Pipe failure	ANN and Fuzzy logic (neurofuzzy)
Rajani and Tesfamariam (2005)	Estimating time to failure of ageing cast iron water mains under uncertainties	Failure rate	Fuzzy sets theory
Silva *et al.* (2006)	Condition assessment and probabilistic analysis to estimate failure rates in buried metallic pipelines	Failure rate	Weibull probability distribution
Tesfamariam *et al.* (2006)	Probabilistic approach for consideration of uncertainties to estimate structural capacity of ageing cast iron water mains	Failure rate	Fuzzy logic
Al-Barqawi and Zayed (2006)	Condition rating model for underground infrastructure sustainable water mains	Failure rate and condition rating	ANN
Davis *et al.* (2007)	A physical probabilistic model to predict failure rates in buried PVC pipelines	Failure rate	LEFM theory, Weibull hazard function and Monte Carlo simulation
Achim *et al.* (2007)	Prediction of water pipe asset life using neural networks	Pipe failure	ANN
Rajani and Tesfamariam (2007)	Estimating time to failure of cast iron water mains	Failure rate	Fuzzy logic theory
Davis *et al.* (2008a, b)	Failure prediction and optimal scheduling of replacements in asbestos cement water pipes	Lifetime	Weibull and Herz distribution
Davis *et al.* (2008a, b)	Fracture prediction in tough polyethylene pipes using measured craze strength	Time to failure	Craze strength (CDNT) tests and empirical method (deterministic modeling)
Dehghan *et al.* (2008a)	Probabilistic failure prediction for deteriorating pipelines: nonparametric approach	Failure rate	Nonparametric
Dehghan *et al.* (2008b)	Statistical analysis of structural failures of water pipes	Failure rate	Statistical modeling
Savic (2009)	The use of data-driven methodologies for prediction of water and wastewater asset failures	Failure rate	Evolutionary Polynomial Regression (EPR)
Wang *et al.*, (2009)	Prediction models for annual break rates of water mains	Annual break rate	Regression analysis
Christodoulou *et al.* (2009)	Risk-based asset management of water piping networks using neurofuzzy systems	Risk of failure	Neuro fuzzy
Fares and Zayed (2010)	Hierarchical fuzzy expert system for risk of failure of water mains	Risk of failure	Hierarchical Fuzzy Logic
Xu *et al.* (2011)	Pipe break prediction based on evolutionary data-driven methods with brief recorded data	Break prediction	Genetic Programming (GP) and Evolutionary Polynomial Regression (EPR)
Xu *et al.* (2013)	Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network	Break rate	Genetic Programming (GP)

MATERIALS AND METHODS

Characteristics of predictive models: As inferred from the different reviews conducted herein, predictive models are depicted as: highly data oriented: they require a wide range of data used for prediction purposes. In most cases, availability of this data is limited (Clair and Sinha, 2012; Kleiner and Rajani, 2001; Mailhot *et al*., 2003; Poulakis *et al*., 2003; Tabesh *et al*., 2009; Yamijala, 2007). They are dynamic; they can be used even in cases where the available database has got little information. This is because they are able to incorporate expert knowledge together with theoretic knowledge during the modeling process (Heckerman, 1996; Kabir *et al*., 2015; Margaritis, 2003). The models are able to analyze how different parameters affect pipe performance and not just focus on previous failure history alone (Clair and Sinha, 2012). They use presently available and historical failure data to determine future behavior of assets and future failure patterns (Kleiner and Rajani, 2001; Mailhot *et al*., 2003; Rajani and Kleiner, 2001). These patterns are assumed to extend into the future and therefore used for analyzing future failure probabilities (Kleiner and Rajani, 2001; Mailhot *et al*., 2003). Additionally, the models are able to forecast failure in individual pipes and in a network or a grouping of pipes (Clair and Sinha, 2012; Rajani and Kleiner, 2001).

Nonetheless, the issue of uncertainty is still prevalent and one major contributor of uncertainty is data incompleteness (Mailhot *et al*., 2003; Makar and Kleiner, 2000; Margaritis, 2003). Failure data recorded by utilities in most cases are incomplete or contain unreliable and sometimes, false information. These characteristics creates unpleasant modeling problems with the most apparent one being difficulty in estimating failure rate of pipes due to incompleteness or even unavailability of the data itself (Gat and Eisenbeis, 2000; Margaritis, 2003; Tabesh *et al.*, 2009). As discussed by Gat and Eisenbeis (2000) and Margaritis (2003), a lesser quantity of more precise data can produce better results than a more complete or a large quantity but uncertain data. Therefore, when provided with limited but reliable or incomplete data, a safer route is to rely on expert or engineering knowledge to enhance modeling. One suitable technique for such a scenario is the use of Bayesian Networks (BNs).

Bayesian network modeling: Bayesian Networks (BNs) are graphical models used to present knowledge about uncertain domain or used for reasoning under uncertainty. They are made up of nodes and arcs; where the nodes represent system components and the arcs link the nodes indicating probabilistic dependencies or relationship between them. BN modeling therefore is a probabilistic approach used to model and forecast the behavior of a system based on observed proceedings (Ben-Gal *et al.*, 2007; Doguc and Ramirez-Marquez, 2009; Fenton *et al.*, 2002; Heckerman, 1996; Margaritis, 2003). In a typical Bayesian network, the interaction among the system components leads to the ultimate system behavior in terms of success or failure. The arcs basically run from a parent to a child node, signifying that the probability of success of a child depends on or is conditional to its association with the parents, determined by their strength of influence to the child (Doguc and Ramirez-Marquez, 2009). In case of absence of a link then the system component are considered independent variables. Uncertainty is therefore represented by associating probabilities with the links between the components. As demonstrated by Fenton *et al.* (2002), the probabilities conform to three basic maxims:

- P (A), the probability of an event A lies between 0 and 1
- P (A) = 0 means that A is impossible while P (A) = 1 means that A is definite
- P (A or B) = P (A)+P (B), provided A and B are disjoint

In addition, a BN is only complete when all the conditional probabilities are computed and represented in the ultimate model (Ben-Gal *et al.*, 2007). A simple illustration of BNs is shown in Fig. 1. The illustration depicts how system components interact, leading to system success or failure.

In Fig. 1, the topmost components, A_1, A_2 and A_4 are independent while the others are the dependent components and probabilities can be computed using the Baye's theorem. Baye's theorem also enhances appropriate assignment of conditional probabilities. Basing on an illustration embraced by Kabir *et al.* (2015), given a situation or a scenario comprising n number of mutually exclusive parameters A_i (i = 1, 2, ..., n) and when given observed data Y then the probability can be updated by Eq. 1:

$$P(A_i|Y) = \frac{p(Y|A_i)\times p(A_i)}{\sum_j p(y|A_j)p(A_j)} \tag{1}$$

Where:

P(A|Y) = The posterior occurrence of the probability of A given the condition that Y occurs

P(A) = The prior occurrence probability of A

P(Y) = The marginal (total) occurrence probability of Y which is considered constant given the data at hand and finally

P(Y|A) = The conditional occurrence probability of Y given that A occurs too and is viewed as the likelihood distribution

BNs together with Bayesian analytical techniques facilitate combination of expert, domain or engineering knowledge and data. This knowledge refers to our prior belief regarding the subject and can be incorporated through the use of causal semantics within BNs that make it possible and more forward to program prior knowledge. This is very critical in situations where data is scarce (Ben-Gal *et al.*, 2007; Doguc and Ramirez-Marquez 2009; Francis *et al.*, 2014; Kabir *et al.*, 2015). BNs also allow us

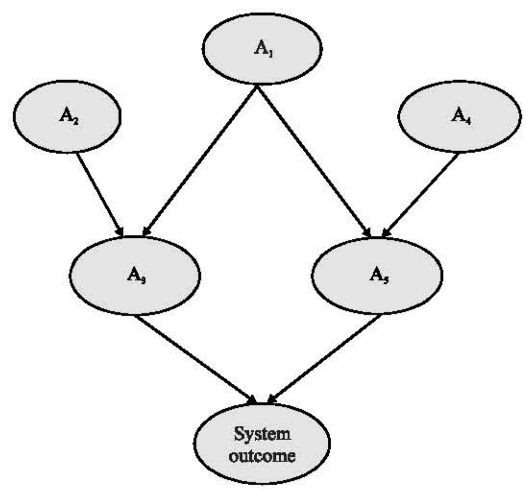

Fig. 1: A simple Bayesian Network modified after (Doguc and Ramirez-Marquez, 2009)

to learn about causal relationships within the data which not only provide understanding about the domain but also allows us to make predictions. Most significantly with BNs we are able to handle incomplete data sets (Ben-Gal *et al*., 2007; Doguc and Ramirez-Marquez, 2009; Francis *et al*., 2014; Margaritis, 2003). In addition, every arrival of new information means that the prior probability can be updated (Fenton *et al*., 2002).

Evidence of modeling infrastructural systems using Bns was seen from the late 80s and has since been embraced by different researchers in literature (Kabir *et al*., 2015). Complexities of the models and the modeling techniques have also evolved to try and keep up with the present systems that gradually become more composite (Doguc and Ramirez-Marquez, 2009). Traditionally, experts had a difficult task of building accurate models because the process was tedious and full of errors (Kabir *et al*., 2015). To deal with such challenges, BNs were proposed as an alternative as they provided graphical belief environment for predicting risk of failure in complex systems (Ben-Gal *et al*., 2007; Doguc and Ramirez-Marquez, 2009; Francis *et al*., 2014; Kabir *et al*., 2015). In the subsequent sections of this study we examine how BNs have been applied in different ways to model pipeline failure; we also identify the various gaps presented by these models and highlight the current state of research as far as leakage prediction is concerned.

Recent models: Babovic *et al*. (2002) applied two advanced data mining techniques to determine the risk of pipe bursts using a database of previous burst events from the Danish capital, Copenhagen. The first technique was a scoring model which was described as the purely data mining method used with the intention of establishing an association between inputs (predictors) and the outputs. Determination of the association was necessary for modeling the behavior of different cases for instance pairing two cases that showed similar behavior. The score quality was measured using a quantification method called the Coefficient of Concordance (CoC). CoC was determined in five steps: involved assigning scores to each identified case in a data set after which cases bearing the same score were grouped together. Different cases (‘good’ or ‘bad’) in each group were then counted and then ordered in descending order of their scores. Finally, the CoC was calculated as a percentage of the cases. The second modeling technique was a BN (a data and knowledge model) that used pipe particulars, soil surrounding and pipe pressure as input parameters. As a result, the model produced an estimate of the pipe history and three other limit functions given as: hoop stress limit, a function that described pipe failure due to hoop stress. Fatigue limit that described the pipes capability to endure stress and shear stress limit, describing pipe failure due to shear stress. Part of their results, indicated that the scoring model had produced a lower maximum burst risk than the Bayesian model and was described as more homogenous in its performance.

This research basically presented an introduction on how to model underground pipelines using data mining methods. Both the BN model and the scoring model were not exhaustively discussed, hence creating need for further investigation if reliable models are to be produced from them. However, the ability of application of prior knowledge and expert experience for BN modeling was comparatively highlighted. The data requirements for this study included pipe details (age, length, diameter, material, number of previous bursts, year installed, traffic frequency in pipe surrounding and house count along the pipeline) for the scoring model. The BN model made use of pipe material, depth and mean diameter, method of installation, previous repairs, temperature, soil type surrounding pipe and rainfall amount. It is noteworthy to mention that the BN model proposed in this case only estimated the pipe history and three other limited state functions.

Poulakis *et al*. (2003) used a Bayesian technique coupled with hydraulic simulations to develop a model for detecting leakages in water pipes. The methodology was used because of its ability to handle uncertainties in measurements and modeling. The modeling procedure involved an assumption that it would be possible to detect a leak by correlating the changes in flow characteristics to the changes in the hydraulic patterns for a given network. It was also highlighted that observed changes in the hydraulic model were indicative of pipe damage hence pointing out the location and extent of the damage. Generally, the proposed strategy was based on an argument that pipeline leakage (in one or more location) involves liquid outflow in the leakage location. This outflow was believed to change the flow characteristics of a pipe, for instance, causing a change in the flow rate, pressure heads and even a change in the acoustic signals. However, the magnitude of the deviation in pipe flow is dependent on the position and severity of the said damage. Given the flow measurements obtained from sensor readings and a data management system, estimates of the probability of leakage events were obtained with the most probable event identified as that with the highest probability. The estimation also included the magnitude and location of leaks. Other leak events were ordered according to their relative probabilities.

The study abovementioned incorporated the use of sensor measurements, hydraulic simulation and Bayesian models for the determination of leak locations and leak extent or severity of a leakage. The Bayesian models were modified to fit in with the other measurement aspects for the purpose simulating a possible leak event, after which they were applied to a pipe network. Device noise, leak severity, error modeling and sensor configuration issues were also addressed given that sensor usage was incorporated in the study. For quality measurement reading and modeling precision, the researchers suggested that most favorable sensor location plans ought to be improved as this would improve the reliability of the predicted estimates. Data used basically included information derived from flow test data.

Doguc and Ramirez-Marquez (2009) presented a method for constructing a BN for forecasting system reliability using historical data available. Their objective was to minimize the need of experts in the development of BNs, reason being that humans are prone to errors both intentional and unintentional which could lead to miscalculations. The proposed method involved the use of K2 algorithm. K2 algorithm is a machine-learning algorithm that uses canonically ordered sets of variables to discover relations among them through the use of predefined scoring functions and heuristic techniques and is used for association rule mining. It searches the best set of parents for a node with the heuristic properties to help reduce the search space from exponential to quadratic. Conditional probabilities were then computed and stored in Conditional Probability Tables (CPT) in a BN and later the Baye's rule implemented to get the overall success of the system

It is a fact that human intervention is prone to mistakes which may lead to unreliable results and finding a good number of experts for construction of BN models is costly, limited and difficult at the same time. However, as pointed out in the study, in order to improve the accuracy of the K2 algorithm, the researchers recommend that existing associations should be taken into consideration associations, an aspect which would require human intervention. The correctness and accuracy of reliability estimation are dependent on the resulting BN model and the accuracy of the model also requires more input which may not be available. The data from which the network was constructed and tested was however not indicated and for the improvement of its precision, human expertise is still a requirement.

Wang *et al*. (2010) applied Bayesian inference to assess the deterioration rate of pipes. The study however was more focused on the quantification of factors that majorly affected pipe deterioration. To achieve this, an approach using Bayesian configuration against the pipe condition together with Prior knowledge of water assessment procedures was applied to generate the weight of influence of the failure factors. The process was divided into three steps. First, pipe data was used to generate relative weights of the factors. This was done by utilizing expert recommended values and use of Bayesian inference for the weight generation. The Markov Chain Monte Carlo (MCMC) method was then employed to numerically solve the Bayesian posterior distributions and for the Bayesian fitting. Secondly, evaluation of the influence of each factor on model performance was done, one at a time. This assisted in the determination of how each given factor contributed to pipe condition. Lastly, simplification of the model was attempted. Here, a realistic predictive model with the least factors was obtained, dropping out factors that had the least influence on pipe condition. A test was then carried to check if the model obtained accurately fit pipe condition without the excluded factors. Among the results, pipe diameter, the inner and outer coatings were found to be quite influential and significant for assessment. Trench depth, electric recharge and the number of road lanes had small weights and therefore were considered not so significant.

The feasibility of using Bayesian theory combined with expert knowledge for pipe assessment was demonstrated in the study. A statistical understanding of how different factors contribute to pipe failure was relayed. This shows that during analysis of factors affecting deterioration, pipe data can be effectively exploited without basically looking for pipe failure mechanisms that may be difficult to find. Results produced by this work were quite consistent to those found in other studies, for instance pipe age was found to be the most influential factor affecting pipe condition. However, the study reported used data from literature, no actual application of the model to a working distribution network. The data used included pipe details from three pipe materials: cast iron, steel and ductile cast iron. Selected properties included pipe material, size, age, inner and outer coating of pipe , pipe bedding condition, soil condition, electrical recharge, trench depth, pipe operational pressure and number of road lanes close to pipe.

Francis *et al*. (2014) presented a knowledge model for pipe breaks using BBNs and utilizing pipe break data from mid-Atlantic United States (US). For learning the BN structure, the researchers used the Grow-Shrink (GS) algorithm in the constraint-based and the RSMAX2 algorithm in score-based methods. This integration of methods was to help in improving the fit and interpretation of the BBN. The models were then evaluated using their negative log-likelihood. However, the results presented suggested that in the dataset there were relationships between variables that were not well-fit. The data set used was both zero-inflated at the individual pipe segment level and aggregated at the census tract level and therefore, individual pipe segment models were not able to be constructed. In addition, differentiation between signal and noise was quite difficult.

At the time of the research, the researchers reported unavailability of supervised discretization technique. This made it difficult to work with continuous data while dealing with continuous discrete models. Moreover, the researchers highlighted the need for standardization of data collection to direct utility operators during data collection. This work however, apart from not totally including pipe characteristic information, did not really show how pipe characteristics like age, diameter and material among others contribute to failure. A wider range of other variables were however used including location data, soil characteristics, population details and weather condition.

Kabir *et al.* (2015) proposed the use of Bayesian Belief Networks (BBNs) to assess the risk of failure of metallic pipes. In their model, different factors affecting pipe deterioration were classified into four main categories and incorporated in the modeling procedure. They included: the Hydraulic Capacity Index (HCI), Structural Integrity Index (SII), Water Quality Index (WQI) and Consequence Index (CI). Graphical representation of the belief network was generated using Netica (commercially available software). Failure risk was then categorized in three levels: low, medium and high risk levels. These risk levels were then used to check the BBN model using three hypothetical scenarios: depicting a situation where the criteria used is in the worst state condition, where the criteria is of average condition and where the criteria is of favorable condition. Among the results it was discovered that failure risk of a main would be very high if it has poor structural condition in terms of very large diameter, a very long length and a maximum age; poor hydraulic condition due to low water pressure and low water velocity; poor soil condition in terms of high pH, low resistivity, poor drainage condition and low reduction and oxidation (redox) potential, all which increased soil corrosiveness; and finally, populated area which meant maximum land use.

In this reserch, factors leading deterioration and eventual failure as well as the consequences of the failures were studied. The model construction made use of these robust attributes whose information was collected from different types of documents, ranging from manufacturing and pipe design information to pipe maintenance reports, visual inspection among others. Among the three broad information category used the specific data used included water Location data, soil characteristics, population details and weather condition data. There are so many direct and indirect consequences that accompany failure which may not be adequately quantified. The work reported examines a rather safe way through which failure consequence may be qualified and analyzed without being intrusive.

Kabir *et al.* (2015) for a second time, proposed to develop a Fuzzy Bayesian Belief Network (FBBN) model for assessing the safety of Oil and Gas pipelines (O&G) failure which was done by incorporating fuzzy logic into BBNs. Their aim was to build a novel and efficient model for safety assessment for evaluating the pipelines failure in dealing with uncertainties. The variables incorporated in their study included linguistic variables and fuzzy number based probabilities instead of only using crisp probabilities that are usually required for Bayesian inference. Fault Trees (FT) were produced and then later transformed into the FBBN by first; directly transforming the primary events, intermediate events and the top event of the FT into parent nodes, intermediate nodes and child nodes in the corresponding BBN in the same order. Secondly, expert analysts were invited to define the likelihood using linguistic terms, after which transformation of the fuzzy number into a crisp value was then performed. The BBN based safety assessment model was however constructed using commercially available software package, Netica. Data was collected from a large number of sources and their results indicated that construction defects, mechanical damage, overload, poor installation and worker experience or quality of works were the factors that mostly affected the failure of oil and gas pipelines.

The study heavily relied on expert opinions and decision making for majority of its procedures. The fuzzy prior probabilities and fuzzy conditional probabilities of both the parent and child nodes were provided by experts based on their experience. Construction of the Bayesian model and the FT was solicited by experts, so was the obtaining of the conditional probabilities used in the BBN model. Experts were again used for the determination of the proper linguistic variables used in the modeling process and for the proper mapping of the FT to the FBBN. For weight generation of the most credible decision making and establishment of the normalization factor, experts were used and so on and so forth. The modeling required too much expertise that somehow, it proves to be very expensive. In addition, expert decisions and judgments are likely to conflict one another leading to confusion. A simplified summary of the models reviewed in this study is illustrated in Table 2.

Table 2: Summary of BBN models

Article reference	Title	Type of prediction	Method used	Parameters used
Bayesian Belief models for assessing pipe failure				
Babovic *et al.* (2002)	A data mining approach to modeling	Risk of failure of water supply assets	BN method and Scoring method	Age, length, diameter, material, number of previous bursts, year installed, traffic frequency in pipe surrounding and house count along the pipeline (scoring model). Pipe material, depth and mean diameter, method of installation, previous repairs, temperature, soil type surrounding pipe and rainfall amount (BN model)
Poulakis *et al.* (2003)	Leakage detection in water pipe networks using a Bayesian probabilistic framework	Leak location and magnitude	Hydraulic simulation and BNs	Pipe flow characteristics like the flow rate, pressure heads and sensor signals
Doguc and Ramirez-Marquez (2009)	A generic method for estimating system reliability using Bayesian networks	System reliability	K2 machine learning algorithm	No data indicated
Wang *et al.* (2010)	An assessment model of water pipe condition using Bayesian inference	Deterioration rate	Bayesian inference	Pipe material, size, age, inner and outer coating of pipe, pipe bedding condition, soil condition, electrical recharge, trench depth pipe operational pressure and number of road lanes close to pipe
Kabir *et al*, (2015)	Evaluating risk of water mains failure using a Bayesian belief network	Risk of failure	BBN	Hydraulic Capacity index (HCI), Structural Integrity Index (SII), Water model Quality Index (WQI) and Consequence Index (CI)
Francis *et al.* (2014)	Bayesian Belief Networks for predicting drinking water distribution system pipe breaks	Probability of failure	BBN learning	Location data, soil characteristics, population details and weather condition data
Kabir *et al.* (2015)	A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines	Safety assessment	FBBN	General pipe failure data

RESULTS AND DISCUSSION

The principal motivation behind construction of models using BBNs is in their ability to handle different levels of uncertainty. As indicated in the various articles studied herein, pipeline operation, pipe failure and pipe assessment all come with inherently risky properties. During operation, pipelines are exposed to several risky and uncertain scenarios causing deterioration and eventual failure. In addition to this when pipelines undergo operational assessment, uncertain situations such as determination of precise location of failure, actual causes of failure, proper assessment methods and tools assessor's qualification among others are encountered. This leads to the unveiling of equally uncertain, incomplete and or irregular records of data. The ability of BNs that is portrayed in the various ways on how they are able to handle diverse scenarios of uncertainty therefore makes them a pretty good candidate for predicting pipe conditions. Inclusion of prior information based on expert knowledge when using BNs ensures that prediction of uncertainties is controlled through proper formulation of the priors. These prior values represent the system dynamics that are not strange to human expertise.

However, excessive inclusion of expert judgment and opinions as illustrated in some studies (Kabir *et al*., 2015) also prove to be very complicated. When a large number of experts are involved in decision making, expert opinion overload arises. This leads to a conflicted and disorderly decision making process, even though it is argued by Kabir *et al.* (2015) that choices are made based on the expert's experience. A model is likely to have different beliefs about the variables to be included, connections, probability generation, among others. This consequently, further complicates the decision making process and increases the levels of uncertainty when it is the same uncertainty being tackled. Additionally, humans are also prone to mistakes both intentional and unintentional (Doguc and Ramirez-Marquez, 2008). Furthermore, finding a good quality and quantity of experts for opinions in almost every process of modeling is complicated, limited and costly. The universal objective of modeling of pipeline asset deterioration and failure is the production of the best possible management tools at the lowest possible cost, hence the proposition by Doguc and Ramirez-Marquez (2009) to minimize expert intervention.

BNs are used to determine different aspects of pipeline failure including risk of failure (Babovic *et al*., 2002; Francis *et al*., 2014) and deterioration rates (Wang *et al*., 2010), due to the fact that they can quantify factors that majorly influence pipe deterioration and also relate these factors to pipe condition. This is fundamental in the event that utilities are faced with difficulties such as lack of funding, insufficient manpower; lack of instrumentation, among other factors that collectively lead to information shortage. It then becomes easier to point out between the critically needed aspects and those that may be ignored. They are also used to analyze safety assessment methods (Kabir *et al*., 2015), failure probability (Fenton *et al*., 2002) and system reliability (Doguc and Ramirez-Marquez, 2009) among others. Although, different parameters or factors may be related to pipe deterioration condition, the assumption that leakages changes the flow characteristics of a pipe may entirely not be true because some leak incidents may not be so obvious. Cutting across majority of these models, the basic parameters exploited for modeling pipeline failure include the general pipe failure data comprising the diameter, age, material, soil corrosiveness or soil condition, break rate, pipe length, pressure and traffic type in that order. This is followed by other factors like pipe depth, previous repairs and weather condition among others. Pipe age and pipe break rate are however unanimously voted as the biggest contributors to failure.

Current state of research: Establishment of universally reliable and acceptable pipeline detection, deterioration or prediction models that are fit for use globally is not easy. Actually, it is close to impossible (Clair and Sinha, 2012). This is due to regional, environmental, economic, operational and even technological differences available to utilities in different regions. Therefore, for effective modeling of a pipeline networks there is need for customization (Clair and Sinha, 2012; Kleiner and Rajani, 2001; Wang *et al*., 2010). Modeling parameters and strategies should be tailored to different utilities, putting in mind their goals and pipe assessment conditions. Although, a number of factors could be similar to most utilities, differences still lie among them in terms of pipe conditions and operational conditions, information availability and regional mapping. Therefore, the greatest concern lies on location specific risks. Pipeline deterioration processes occur differently in particular regions due to specific regional risks affecting the pipes as well as the consequences of pipe leakage to these regions. In addition, modeling of failure in high risk zones has not been really tackled, neither are they depicted by the different models reviewed, except in the oil and gas pipelines.

Definition of risk is relative. However, risk is generally governed by the likelihood of an event (risk occurrence) and the magnitude or the degree of loss (in this case, failure and consequences), respectively (Buttrick *et al*., 2002). Risk determination is equally not easy, although we can estimate where it is most probable and the severity of its consequences. Risk assessment models for high risk zones ought to be developed, taking into consideration the severity of the consequences of the various forms of failure. Generally, inclusion of failure consequence in modeling has not been fully exploited. Kabir *et al*. (2015) however, incorporated consequence of failure parameters in their study, in the form of population density, land use and pipe diameter. Nonetheless, further research on identification of additional ways on how the impacts and severity of failure can be incorporated into predictive modeling is highly recommended. Advanced exploration on the use of both data based and knowledge based BBN modeling for determination of water network risk index for rehabilitation prioritization is also recommended. This will be essential in supporting municipalities and utilities with proactive decision making tools that addresses water pipeline failure in time even when under constrained financial limits.

CONCLUSION

In this study, a detailed review, however not exhaustive on how predictive modeling of pipeline failure using Bayesian Networks has been done. BNs are confirmed to be quite effective in handling different aspects of uncertainty that are associated with pipeline operation, failure and pipeline failure assessment. A number of gaps exhibited by the models studied herein have also identified and adequately relayed. In a nutshell, it is noteworthy to mention that: Pipeline failure is inevitable, nonetheless quite complex with inadequate comprehension of the failure mechanisms and processes. Additionally, availability of data that can be utilized to model these failure processes is limited. This is because utilities are faced with barriers such as lack of adequate investment in pipe maintenance leading to a shortage in pipe failure records. The restricted data availed by utilities additionally; fail to meet the standard requirements for data collection, recording and analysis procedures. These are the greatest contributors to uncertainty.

However, the usefulness of network analytical models is that they are able to overcome such problems (uncertainty) associated with data inadequacy. This on the other hand, does not mean that water utilities should stop or avoid the collection of available data and keep an inventory of pipe operation effectively. It is also depicted from the different articles studied herein that; a great deal potential lies in the utilization of existing data. Therefore, more research on information discovery from limited should be encouraged so as to engage in better decision making. There is lack of a standard definition of failure. Available definitions are basically based on suitability, even though evidence of failure is quite uniform, resulting in leakages. Identification of the kind of information necessary for modeling is quite challenging and warrants further research. A universally acceptable or a standardized level of modeling accuracy has not been clarified yet which is another area that probably requires further research.

ACKNOWLEDGEMENTS

Researchers take this opportunity to say 'thank you' to the Department of Computer Science, ICT faculty at the Tshwane University of Technology (TUT) from where this research was conducted. The researchers also Thank the University (TUT) for funding this work.

REFERENCES

Achim, D., F. Ghotb and K.J. McManus, 2007. Prediction of water pipe asset life using neural networks. J. Infrastruct. Syst., 13: 26-30.

Al-Barqawi, H. and T. Zayed, 2006. Condition rating model for underground infrastructure sustainable water mains. J. Perform. Constr. Facil., 20: 126-135.

Andreou, S.A., D.H. Marks and R.M. Clark, 1987. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Applications. Adva. Water Resour., 10: 11-20.

Babovic, V., J.P. Drecourt, M. Keijzer and P.F. Hansen, 2002. A data mining approach to modelling of water supply assets. Urban Water, 4: 401-414.

Ben-Gal, I., F. Ruggeri, F. Faltin and R. Kenett, 2007. Bayesian Networks. In: Encyclopedia of Statistics in Quality and Reliability, Fabrizio R. (Ed.). Wiley, Hoboken, New Jersey, USA., pp: 1-6.

Buttrick, D.B., A.V. Schalkwyk, R.J. Kleywegt, R.B. Watermeyer and N. Trollip, 2002. Proposed method for dolomite land hazard and risk assessment in South Africa. J. S. Afr. Inst. Civil Eng., 44: 27-36.

Christodoulou, S., A. Deligianni, P. Aslani and A. Agathokleous, 2009. Risk-based asset management of water piping networks using neurofuzzy systems. Comput. Environ. Urban Syst., 33: 138-149.

Christodoulou, S., P. Aslani and A. Vanrenterghem, 2003. A risk analysis framework for evaluating structural degradation of water mains in urban settings, using neurofuzzy systems and statistical modeling techniques. Proceedings of the Congress on World Water and Environmental Resources, June 23-26, 2003, American Society of Civil Engineers, Pennsylvania, USA., pp: 1-9.

Clair, S.A.M. and S. Sinha, 2012. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models!. Urban Water J., 9: 85-112.

Colombo, A.F., P. Lee and B.W. Karney, 2009. A selective literature review of transient-based leak detection methods. J. Hydro Environ. Res., 2: 212-227.

Davis, P., D.D. Silva, D. Marlow, M. Moglia and S. Gould *et al.*, 2008a. Failure prediction and optimal scheduling of replacements in asbestos cement water pipes. J. Water Supply Res. Technol. Aqua, 57: 239-252.

Davis, P., S. Burn and S. Gould, 2008b. Fracture prediction in tough polyethylene pipes using measured craze strength. Polym. Eng. Sci., 48: 843-852.

Davis, P., S. Burn, M. Moglia and S. Gould, 2007. A physical probabilistic model to predict failure rates in buried PVC pipelines. Reliab. Eng. Syst. Saf., 92: 1258-1266.

Dehghan, A., K.J. McManus and E.F. Gad, 2008a. Probabilistic failure prediction for deteriorating pipelines: Nonparametric approach. J. Perform. Constr. Facil., 22: 45-53.

Dehghan, A., K.J. McManus and E.F. Gad, 2008b. Statistical analysis of structural failures of water pipes. Proceedings of the Conference on Institution of Civil Engineers-Water Management Vol. 161, August 25-27 2008, Thomas Telford Ltd, London, England, pp: 207-214.

Doguc, O. and J.E. Ramirez-Marquez, 2009. A generic method for estimating system reliability using Bayesian networks. Reliab. Eng. Syst. Saf., 94: 542-550.

Engelhardt, M.O., P.J. Skipworth, D.A. Savic, A.J. Saul and G.A. Walters, 2000. Rehabilitation strategies for water distribution networks: A literature review with a UK perspective. Urban Water, 2: 153-170.

Fares, H. and T. Zayed, 2010. Hierarchical fuzzy expert system for risk of failure of water mains. J. Pipeline Syst. Eng. Pract., 1: 53-62.

Fenton, N., P. Krause and M. Neil, 2002. Software measurement: Uncertainty and causal modeling. IEEE. Software, 19: 116-122.

Francis, R.A., S.D. Guikema and L. Henneman, 2014. Bayesian belief networks for predicting drinking water distribution system pipe breaks. Reliab. Eng. Syst. Saf., 130: 1-11.

Gat, L.Y. and P. Eisenbeis, 2000. Using maintenance records to forecast failures in water networks. Urban Water, 2: 173-181.

Heckerman, D. 1996. Bayesian Networks for Knowledge Discovery. In: Advances in Knowledge Discovery and Data Mining, Fayyad, U., G. Piatetsky-Shapiro, P. Smith and R. Urthurusamy (Eds.). Chapter-11, AAAI Press/The MIT Press, ISBN:0-262-56097-6, pp: 273-306.

Kabir, G., R. Sadiq and S. Tesfamariam, 2016. A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines. Struct. Infrastruct. Eng., 12: 874-889.

Kabir, G., S. Tesfamariam, A. Francisque and R. Sadiq, 2015. Evaluating risk of water mains failure using a Bayesian belief network model. Eur. J. Oper. Res., 240: 220-234.

Kleiner, Y. and B. Rajani, 1999. Using limited data to assess future needs. Am. Water Works Assoc. J., 91: 47-62.

Kleiner, Y. and B. Rajani, 2000. Considering time-dependent factors in the statistical prediction of water main breaks. Proceedings of the Conference on American Water Works Association Infrastructure, March 12-15, 2000, Baltimore, National Research Council, Maryland, pp: 1-12.

Kleiner, Y. and B. Rajani, 2001. Comprehensive review of structural deterioration of water mains: Statistical models. Urban Water, 3: 131-150.

Mailhot, A., A. Poulin and J.P. Villeneuve, 2003. Optimal replacement of water pipes. Water Resour. Res., Vol. 39, 10.1029/2002WR001904

Makar, J.M. and Y. Kleiner, 2000. Maintaining water pipeline integrity. Proceedings of the Conference and Exhibition on AWWA Infrastructure, March 12-15, 2000, Baltimore, Maryland, Institute for Research in Construction, pp: 1-13.

Margaritis, D., 2003. Learning Bayesian network model structure from data. Ph.D Thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Poulakis, Z., D. Valougeorgis and C. Papadimitriou, 2003. Leakage detection in water pipe networks using a Bayesian probabilistic framework. Probab. Eng. Mech., 18: 315-327.

Rajani, B. and S. Tesfamariam, 2005. Estimating time to failure of ageing cast iron water mains under uncertainties. Masters Thesis, University of Exeter, Exeter, England.

Rajani, B. and S. Tesfamariam, 2007. Estimating time to failure of cast-iron water mains. Proceedings of the Conference on Water Management Vol. 160, June 14-16, 2007, Thomas Telford Ltd, London, England, pp: 83-88.

Rajani, B. and Y. Kleiner, 2001. Comprehensive review of structural deterioration of water mains: Physically based models. Urban Water, 3: 151-164.

Savic, D.A., 2009. The Use of Data-Driven Methodologies for Prediction of Water and Wastewater Asset Failures. In: Risk Management of Water Supply and Sanitation Systems, Hlavinek, P., C. Popovska, J. Marsalek, I. Mahrikova and T. Kukharchyk (Eds.). Springer, Berlin, Germany, pp: 181.

Silva, D.D., M. Moglia, P. Davis and S. Burn, 2006. Condition assessment to estimate failure rates in buried metallic pipelines. J. Water Supply Res. Technol. Aqua, 55: 179-191.

Stone, S.L., E.J. Dzuray, D. Meisegeier, A. Dahlborg and M. Erickson *et al*., 2002. Decision-support tools for predicting the performance of water distribution and wastewater collection systems. US Environmental Protection Agency, Cincinnati, Ohio.

Tabesh, M., J. Soltani, R. Farmani and D. Savic, 2009. Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. J. Hydroinf., 11: 1-17.

Tesfamariam, S., B. Rajani and R. Sadiq, 2006. Possibilistic approach for consideration of uncertainties to estimate structural capacity of ageing cast iron water mains. Can. J. Civil Eng., 33: 1050-1064.

Wang, C.W., Z.G. Niu, H. Jia and H.W. Zhang, 2010. An assessment model of water pipe condition using Bayesian inference. J. Zhejiang Univ. Sci. A., 11: 495-504.

Wang, Y., T. Zayed and O. Moselhi, 2009. Prediction models for annual break rates of water mains. J. Perform. Const. Facil., 23: 47-54.

Xu, Q., Q. Chen, J. Ma and K. Blanckaert, 2013. Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network. J. Hydro Environ. Res., 7: 134-140.

Xu, Q., Q. Chen, W. Li and J. Ma, 2011. Pipe break prediction based on evolutionary data-driven methods with brief recorded data. Reliab. Eng. Syst. Saf., 96: 942-948.

Yamijala, S., 2007. Statistical estimation of water distribution system pipe break risk. Masters Thesis, Texas A&M University, Texas, USA.