Comparative performance of regression and the Markov based approach in the prediction of the future condition of a water distribution pipe network amidst data scarce situations: a case study of Kampala water, Uganda

Jotham Ivan Sempewo **IWA**[a,*] and Lydia Kyokaali[b]

[a] Civil and Environmental Engineering Department, Makerere University, P O Box 7062, Kampala, Uganda

[b] Lydia Kyokaali, Engineer, National Water and Sewerage Corporation., P O Box 70255, Kampala, Uganda

*Corresponding author. E-mail: jsempewo@cedat.mak.ac.ug

Abstract

Globally, water utilities are grappling with the challenge of predicting the condition of deteriorating pipe network infrastructure amidst financial constraints and data-scarce scenarios. As a result, new innovative approaches such as statistical regression and Markov-based approaches have been introduced to aid water distribution pipe renewal decision making. However, comparison of the performance of these models under limited data has not been undertaken so far. In addition, the models have been applied elsewhere, in different environments and data availability scenarios. This paper addresses therefore the mentioned research gap and compares the performance of statistical regression and Markov models in the prediction of a condition of a pipe in a developing country. In addition, the criticality analysis of a block is studied. The data used for assessment is from Kampala water, the largest area in the National Water and Sewerage Corporation, Uganda. The results show that 78.26% of the prediction of the regression model is accurate in comparison to 88.4% for the Markov model. This means that the Markov-based approach is more superior than a regression model in a data scarce scenario. The approach will go a long way in helping water utilities in development of water decision pipe renewal plan amidst a limited budget and in data scarce scenarios.

Key words: decision making, markov chains, probabilistic learning, regression analysis, water distribution network

INTRODUCTION

Globally, many countries particularly of middle and low-income status, are struggling with the challenge of deterioration and the rampant failure of water distribution due to funding limitations and ageing infrastructure. Aging pipes often lead to increased pipe breaks, which affect system losses and the levels of service for the customers (Mutikanga 2012) resulting in customer dissatisfaction, potential health risks to the consumer due to contamination (Vairavamoorthy *et al.* 2007) and high operation and maintenance costs in terms of energy, repair costs and the carbon footprint for a water utility (González-Gómez *et al.* 2011; Echart *et al.* 2012).

A study by Folkman (2018) showed that in Canada and the United States, water main breaks increased by 27% from 7 to 9 breaks/(100 km)/year between 2012 and 2018. Currently, Uganda with a 15,000 km network experiences 1,175 breaks/100 km annually in contrast to 800/100 km/year in Sub-Saharan Africa (Bannerjee & Morella 2011) and 100 times higher than the Netherlands at 8/100 km/year (Vloerbergh & Blokker 2009). This break rate is also 200 times higher than the acceptable standard (Pelletier *et al.* 2003). In Canada and the USA, averages of 700 water main breaks are registered on a daily basis (ASCE 2017). If these pipes are to be replaced in the USA

alone to reduce the pipe breaks, the American Society of Civil Engineers (ASCE) estimates that over USD 1 trillion is required in the next couple of decades (Boulos 2017). In Uganda, an average of USD 7 million per annum is spent on network repairs and maintenance (NWSC 2014).

Rapid urbanisation and population growth coupled with ageing infrastructure amidst scarce resources has created the need for robust rehabilitation and main replacement decision support systems and approaches (Echart *et al.* 2012; Scholten *et al.* 2014). As a result, new innovative approaches have been developed to help identify optimal rehabilitation of pipes that will yield the biggest return and benefit to the water utilities (Rogers & Grigg 2009; Malm *et al.* 2012; Kabir *et al.* 2015). Most of these studies use models that are risk-based approaches for renewal decisions, including costs; see for example Malm *et al.* 2012; Francisque *et al.* 2013. Recently tools for pipe prediction modelling based on Markov chains under the computer-aided rehabilitation project such as CARE-W (Saegrov 2005) have been developed. However, these are criticized for requiring large data and skill that is not readily available, particularly in developing countries. The tools are based on risk approaches. Recently, water distribution main renewal decisions made by utility managers have been based solely on deterministic approaches (Shamir & Howard 1979; St. Clair & Sinha 2012; Wilson *et al.* 2017) and on a risk-based approach (Zhou 2018), particularly in data-scarce scenarios, and there is lack of skilled staff. Whereas these approaches are favoured because of their simplicity, they are criticized for yielding unrealistic results because the structural deterioration of water mains is probabilistic in nature, governed by several environmental and operational stresses that lead to deterioration (Kleiner & Rajani 2001). Most of these stresses are dynamic in nature, causing uncertainty in the deterioration.

Condition-based models such as Markov chains are probabilistic in nature and can be used to develop a pipe failure prediction approach (Ossai *et al.* 2016; Sempewo & Kyokaali 2016). With the uncertainties and imprecision in the water utilities' historical database, the Markov chain methodology is recommended because it depends only upon the asset's current condition (Baik *et al.* 2006; Uchwat & Macleod 2012). Many studies have investigated the use of predictive models for water infrastructure renewal decisions such as regression models (Asnaashari *et al.* 2009; Wang *et al.* 2009), time linear and exponential models (Le Gat *et al.* 2013), probabilistic models (St. Clair & Sinha 2012; Ji *et al.* 2017) and Bayesian models (Watson *et al.* 2004; Kabir *et al.* 2015). However, few studies have applied Markov chains for the prediction of the future condition of water pipes (Kleiner *et al.* 2006; Kim *et al.* 2015). Existing Markov models are criticized for being skewed towards modelling the deterioration of pipes without a failure history or those which have been recently repaired to restore the condition to as good as new. These approaches are therefore best suited for pipes with low failure rates such as large diameter transmission mains. The approaches cannot be used to model the deterioration of distribution mains, which experience recurrent failures (Economou *et al.* 2008). In addition, the mains, replacement strategies developed from the existing models do not take into account the issue of customer criticality in terms of water sales and are criticized for giving scattered priority groups of pipes, which makes implementation hard.

According to literature, models for predicting the future condition of a pipe can be classified as probabilistic (Kleiner & Rajani 2001), deterministic (Ana & Bauwens 2010) and data-driven models (Xu *et al.* 2013). A critique for the pros and cons for these models can be found in literature (Wilson *et al.* 2017). These models are data dependent, require lots of skill and are tied to areas where the analysis has been done. Like many developing countries, Uganda also faces issues of data availability. In addition, the performance of these models in different environments and data availability scenarios has not been evaluated. Existing assessments and comparisons are qualitative and not quantitative and are limited to literature review (Kleiner & Rajani 2001; Wilson *et al.* 2017).

Sempewo & Kyokaali (2016) presented a Markov probabilistic approach for predicting the future condition of a water distribution pipe network per block. A block is the smallest 500 m by 500 m geographical reference in which pipe infrastructure is located and managed. The approach can help water engineer planners and engineers to predict pipe failures, which is required to inform

optimisation of repair and maintenance decisions. However, the model is limited to use for prediction of future condition and does not consider the ranking of the rehabilitation options. Moreover, the approach is more suitable for areas with vast amounts of data on historical pipe breaks or those with a large geographical area, which must be fairly homogeneous with respect to the factors influencing the deterioration of pipes. The available failure data in urban water systems often present a short failure history and incomplete records (da Costa Martins 2011).

In Uganda, there are limited studies on the application of statistical and probabilistic models to predict the future condition of a pipe. Moreover, assessment of the performance of the aforementioned models in a water utility with limited data has not been undertaken so far. This paper therefore addresses the aforementioned research gap and compares the performance of the statistical and Markov model in the prediction of the future condition of a pipe under limited data. This paper presents a risk-based pipe renewal decision-making approach that combines probability of failure and consequence of failure to develop a pipe renewal plan. Due to the limitations of the previous studies, the innovation of this research is an approach for predicting the future condition of a water distribution network using Markov-based models, taking into account the break history of the pipes and development of a main replacement strategy that takes into account the criticality of the different parts of the pipe network and for which pipes with the same priority are in the same vicinity.

METHODS

The main objective of this paper is to compare the performance of a Markovian probabilistic process to statistical regression models in the prediction of the future condition of a water distribution network. This was done in five steps, and these are outlined below:

1. Prediction of future condition of pipe based on a Markov Chain
2. Statistical regression analysis-evaluation of the condition of a pipe network based on age, diameter and break history.
3. Comparative assessment of predictive performance of the Markov and statistical regression models.
4. Development of a main replacement strategy based on the future condition and criticality of the pipe network.
5. Implementation of a case study using findings from 1, 2, 3 and 4 above as proof of the concept.

Details of the methodology followed for each of the five steps above are elaborated in the sections below:

Prediction of future condition of pipe based on a Markov Chain

The Markov approach proposed by Sempewo & Kyokaali (2016) was followed as the basis for prediction of the future condition of a pipe in this study. The approach was applied on the training data sets to obtain the transition probability matrix and to predict the future condition state. The methodology involved (i) Markov Property verification (ii) computation of Historical State Transition Matrices and (iii) prediction of future condition state, and is as shown in Figure 1.

The results from the transition probability matrix from test data were compared with the actual observed conditions of a pipe and block.
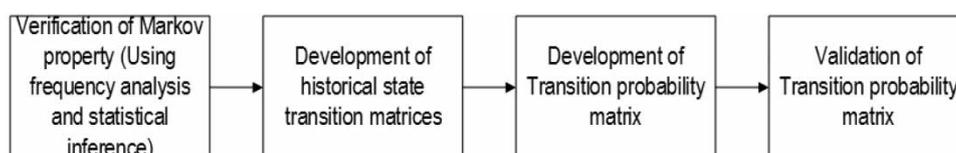
Figure 1 | Schematic layout of the Markov prediction approach.

Statistical regression analysis

The methodology proposed by Kleiner & Rajani (1999) was the basis for the statistical regression analysis. The methodology was applied because it is simple, enables future trends to be predicted based on historical data, allows for an evaluation of the 'fit' of the model prediction and the approach has been widely applied for predicting pipe performance (El-Abbasy *et al.* 2014; Scheidegger *et al.* 2015). The method was applied following the steps below:

- Grouping data on water main breaks (depending on pipe material and diameter) into homogeneous groups to achieve greater statistical significance when predicting break rates (Shamir & Howard 1979; Walski & Pelliccia 1982).
- Establishing the breakage rate patterns for the various groups.
- Validation of the analysis by establishing if the model could reasonably predict data that was not used in its development.

Comparative assessment of predictive performance of the Markov and statistical regression models

To assess the quality of predictions obtained using the Markov and statistical regression models, the condition data were divided into two data sets: a training data set and a test data set. The future condition states for the same pipe network per block were estimated using the training data set and the results were compared for the observed data sets for the two methods. One was entered for a correct prediction while zero was entered for an incorrect prediction. The percentage of correct predictions for each method was compared to assess the predictive performance of each option.

Criticality analysis of the blocks

The risk-based approach for pipe renewal decision making was adopted as the methodology for criticality analysis of blocks. The approach was adopted because it has been widely applied in pipe renewal decision making (Engelhardt *et al.* 2000; Lusiba 2003; Zhou 2018). However, it is criticized for having been applied elsewhere and not in Uganda. The approach was also adopted because of its simplicity and the requirement for minimal data. The risk for each block was obtained using Equation (1)

$$\text{Risk of failure} = \text{probability of failure} \times \text{consequence of failure} \tag{1}$$

The probability of failure was obtained from the condition of the block/pipe whereas the consequence of failure was based on water sales impact data, which were obtained from the National Water and Sewerage Corporation billing system for the year 2015. Impact was based solely on impact index derived from water sales. Blocks with high water sales were assumed to have very critical customers such that any failure will cause a high impact in terms of lost revenue. For simplicity, it was assumed that impact of critical infrastructure such as hospitals was already taken care of in water sales. The impact index was obtained by dividing the water sales equally into five based on a Linkert scale from 1 to 5 (excellent to poor). Details of the method followed can be found in Kyokaali 2018.

APPLICATION OF THE METHODOLOGY ON A CASE STUDY

The developed methodology in Section 2 was applied to the Kampala City Centre Water Distribution System (KCCWDS) in the National Water and Sewerage Corporation (NWSC). NWSC is a government parastatal in charge of water and sewerage services in the country. KCCWDS has the oldest distribution pipes in NWSC with the majority of the pipes having been installed during the colonial

times, dating as far back as 1928. The KCCWDS services area supplies water to the central business district of Uganda's capital city, Kampala. The KCCWDS area encompasses an area of about 100 km² with an estimated water coverage of 77%. The city centre supply area lies in the central part of Uganda at the coordinates of 00 19N, 32 35E. Population estimates based on 2014 National Census Data estimate that 1.5 million inhabitants live within the service area and constitute a combination of upscale business and residential neighbourhoods as well as low-income settlements (UBOS 2014). With an annual growth rate of over 2%, Kampala is one of the fastest-growing cities in the world. Rapid population growth, urbanization, is responsible for increased water demand that has outstripped the capacity of the existing system, resulting in frequent bursts, water losses and intermittent supply. Currently, the greatest number of customer complaints is attributed to service interruptions caused by frequent mains bursts. KCCWDS has 228 km of distribution mains, 59.3% of which are steel pipes, most of which have long exceeded their service life. The network also has plasticised polyethene chloride (uPVC), high density polyethylene (HDPE), cast-iron (CI), ductile iron (DI) and galvanized iron (GI) pipes. The branch has 76 blocks and covers an area of approximately 0.76 km². KCCWDS has been selected as a case study because it is located in the heart of the Capital City and is prone to severe consequences in the event of pipe failures. Consequences may be due to service interruptions and destruction of property. The area has average monthly water sales of 630,403 m³, which translates into monthly revenue collections of close to two billion Ugandan shillings, which is the highest in Kampala water. Pipe failures therefore result in high revenue losses.

The total length of the existing water supply transmission mains is about 125 km in diameters ranging from DN 200 mm to DN 900 mm, while the distribution system is about 2,578 km long, ranging in diameter from DN 50 mm to DN 450 mm, mainly made of steel, DI, uPVC and HDPE materials. The average number of failures reported during the year 2018 was 1,384 breaks/100 km/year, an increase of 8% from 1,175 breaks/100 km/year in 2010 (Mutikanga 2012). This is not only very high compared to global values but also shows an increasing trend in pipe breaks. Non-revenue water averages about 36% of system input volume or 22 million m³ per year (NWSC 2014).

To apply the developed methodology on a case study, the pipe condition data for KCCWDS was first divided into a training data set and a test data set. The training data set used consisted of pipe condition data before 2010. The test data was obtained from observed data of the condition of a pipe and blocks in 2015. The data for 2015 was found to be appropriate for validating the matrix because it was not used in the development of the matrix. Information captured for each block relates to pipe diameter, location in terms of the block, the date on which the failure was reported and that on which it was repaired. The data used in the study was for a period of 25 years split into 5 year intervals because of availability and completeness of data. The 5-year time step was considered because (i) it is sufficient time after which a shift in the condition of a pipe can be observed (ii) the period matches with the rehabilitation planning and implementation cycle for NWSC. The data was obtained from job cards and the network maps at the Kampala water offices. The historical condition states for the data set used are summarized in Table 1.

The pipe breaks per block were converted into condition states and classified equally into five condition states ranging from 1 to 5 (excellent to poor) based on a Linkert scale following Sempewo & Kyokaali (2016).

Prediction of future condition of pipe based on a Markov chain

In this section, the results of application of a Markov approach on the data for KCCWDS to forecast condition states for the blocks are presented. Table 2 presents the results of the forecasted condition states for the blocks using Markov chain modelling.

The results show that 61 out of the 69 forecasts (88.4%) for the condition states for the blocks by the Markov model were accurate. The results agree with the findings of Sempewo & Kyokaali (2016).

Table 1 | Historical pipe break rates per block

Block	Diameter (mm)	Length (m)	Age	Breaks/km/year					
				0	5	10	15	20	25
2129	100	230	35	1.70	2.18	2.72	3.32	3.99	4.72
2130	100	200	34	1.70	2.18	2.72	3.32	3.99	4.72
	50	415	34	–	0.04	0.15	0.36	0.66	1.05
	40	282	34	2.21	–	0.04	0.17	0.40	0.72
	100	426	34	1.61	2.08	2.61	3.20	3.85	4.57
	100	160	34	1.61	2.08	2.61	3.20	3.85	4.57
2030	100	95	33	1.61	2.08	2.61	3.20	3.85	4.57
	100	353	28	1.53	1.98	2.50	3.07	3.72	4.42
	80	83	32	1.77	2.30	2.92	3.61	4.38	5.22
2029	50	226	25	1.87	2.52	–	0.04	0.15	0.36
	80	295	25	1.14	1.57	2.08	2.66	3.32	4.06
	100	215	24	1.53	2.12	2.81	–	0.04	0.15
	50	190	35	0.87	1.21	1.61	2.08	2.61	3.20

Table 2 | Forecasted condition states for the blocks using Markov chain modelling

SNo	Condition state	Total No of blocks in a condition state	No of blocks with correct observed state 2015	Percentage prediction accuracy
1	1	11	10	90.91%
2	2	21	20	95.24%
3	3	17	15	88.24%
4	4	9	5	55.56%
5	5	11	11	100.00%
Total		69	61	88.41%

The performance occurs because the Markov model takes care of uncertainties and probability trends by use of transition probability matrices. This makes the model suitable for prediction of the future condition of a network with a good degree of accuracy amidst data-scarce scenarios. However, there is a need to constantly update data as more information is collected.

Statistical regression analysis – evaluation of the condition of a pipe network based on age, diameter and break history

In this section, the statistical regression analysis for the regression analysis is presented. To improve the prediction accuracy, the KCCWDS were divided into three major groups according to the most common pipe materials, which are PVC, PE and steel. The results presented are for steel pipes because these constitute over 50% of the pipe network. Pipe data were used to create a scatter plot of pipe age versus observed historical break rates. Figure 2 shows the breaks versus age profile for all the steel pipes in KCCWDS.

The developed exponential model equation takes the form:

$$y = 0.014e^{0.076x}, \quad R^2 = 0.627 \tag{2}$$

where y is the observed breaks/km/year and x is pipe age in years.

The R-squared for the above regression is 0.627. When one outlier was eliminated, the R-squared increased to 0.884. The results are comparable and even better than the findings of Kettler & Goulter
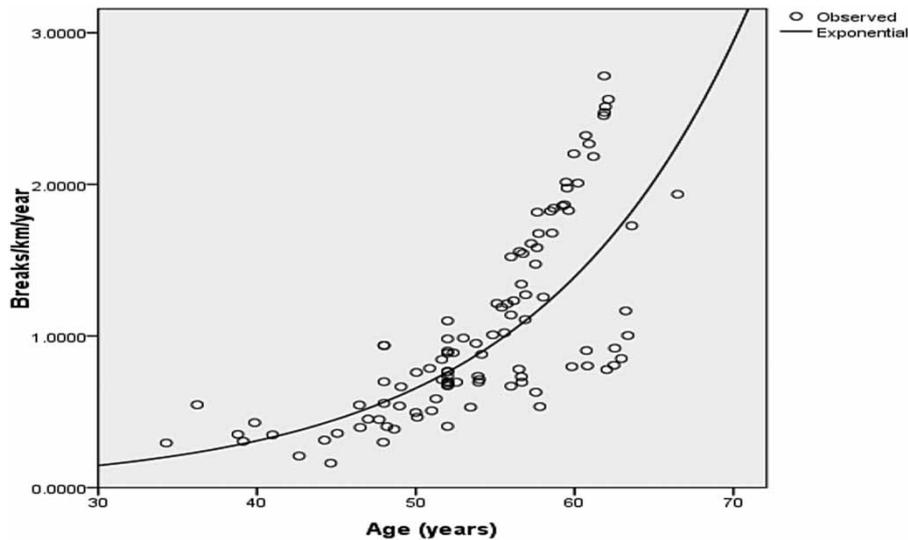
Figure 2 | A plot of age versus break rate for all steel pipes in the city centre.

(1985) who found a moderate R-square of 0.563 and 0.103 for asbestos cement and cast iron pipes respectively. The R-square of .627 is also higher than the R^2 of 0.47 reported by the Clark *et al.* (1982) exponential regression model. A significance of trend test using a t-test was applied to the model, as recommended by Kleiner & Rajani (2001). The results showed that that age has a significant effect on break history and the break rate increases rapidly after a certain period. This could be attributed to many factors that were not investigated in this study.

The model was used to estimate break rates for 2015 and these were compared with the actual observed break rates using a paired samples t-test. Results for the t-test indicate a significance level of 0.001, which is less than the desired significance level of 0.05. This shows that there is a significant difference between the observed and the expected break data. The results mean that the trends established in the 2010 break data were not carried forward into 2015. The model in Equation (2) above is invalid and could not be used for forecasting. To improve results, the steel pipes were divided into three diameter groups of (40–80 mm), (100–150 mm) and (200–600 mm) and the regression analysis and statistical testing process were repeated.

Figure 3 shows the breaks versus age profile for 200–600 mm, 100–150 mm and 40–80 mm diameter steel pipes in KCCWDS. Using linear regression analysis, the following mathematical models were obtained for the relationship between the observed breaks/km/year and age. The mathematical models for each of the three sub-groups of 40–80 mm, 100–150 mm and 200–600 mm diameter pipes in the KCCWDS network and the coefficients of determination R^2 are as shown in Equations (3)–(5) respectively:

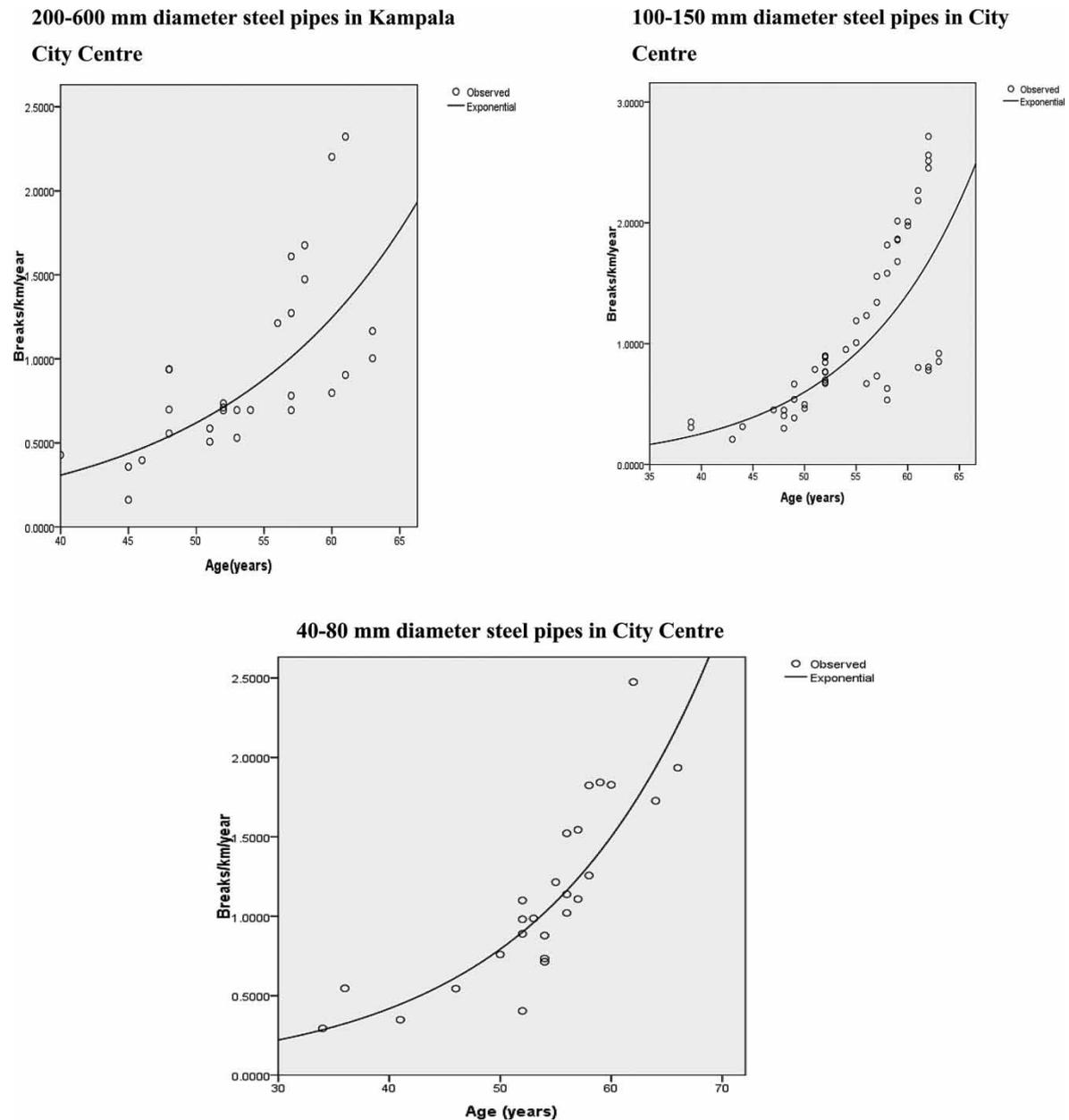40–80 mm diameter pipes:

$$y = 0.031e^{0.064x}, \quad R^2 = 0.756 \tag{3}$$

100–150 mm diameter pipes:

$$y = 0.008e^{0.086x}, \quad R^2 = 0.656 \tag{4}$$

200–600 mm diameter pipes:

$$y = 0.018e^{0.07x}, \quad R^2 = 0.534 \tag{5}$$

where y is observed breaks/km/year and x is pipe age in years.

Figure 3 | Breakage patterns for the different diameter steel pipes in City Center.

The R^2 is low for the pipe cohorts because there are independent variables other than physical indicators that can explain pipe breaks in the literature and these include bedding condition and traffic load (Skipworth *et al.* 2002), surface permeability and ground water condition (Zhou 2018), buried depth (Davies *et al.* 2001) and temperature (Kutyłowska & Hotloś 2014). It was not possible to investigate this data in this study due to data availability constraints. However, further research could be undertaken to include these variables in the regression for data-rich water systems.

The results show differences in R^2 for the different pipe cohorts. The 40–80 mm diameter pipes cohort had the highest R^2 and thus were the best-performing regression model for the KCCWDS and the 200–600 mm diameter pipe cohorts had the lowest R^2 and thus were the worst performing regression model. Wide disparity exists between the R^2 for 200–600 mm diameter pipes and 40–80 mm diameter pipes. The disparity in results is attributed to the difference in the size of dataset used in the regression analysis for 200–600 mm diameter pipes being less than that for 40–80 mm diameter pipes. It is worth noting that the R^2 increases with reduction in pipe diameter. The R^2

observed for the pipe cohorts are comparable and better than the findings of Kettler & Goulter (1985), who found a moderate R-square of 0.563 and 0.103 for asbestos cement and cast-iron pipes respectively. The developed regression models were used to predict the future condition states for the pipe network per block and the results show 54 out of the 69 forecasts (78.26%) for the regression model.

Comparing performance of Markov chain and regression models in prediction of pipe condition

The validated regression models and the new Markov model developed were used to predict the future condition states for the pipe network per block and the results compared. The Markov model had the highest prediction rate at 88.4% and thus is the best performing pipe condition prediction model in comparison to the regression model, which has a prediction rate of 78.26% and was the worst performing. The Markov model performed better because it takes care of factors and probabilistic analyses that that regression models do not account for, as elaborated by Kleiner *et al.* 2006.

Development of a priority-based mains replacement strategy

Figure 4 below shows the results of the criticality analysis for the blocks. The classification of the criticality follows the pipe prioritization decision models that plot pipe condition against significance in technical literature for rehabilitation of pipes (Engelhardt *et al.* 2000; Lippai & Wright 2005). Results show that there are four priority groups.
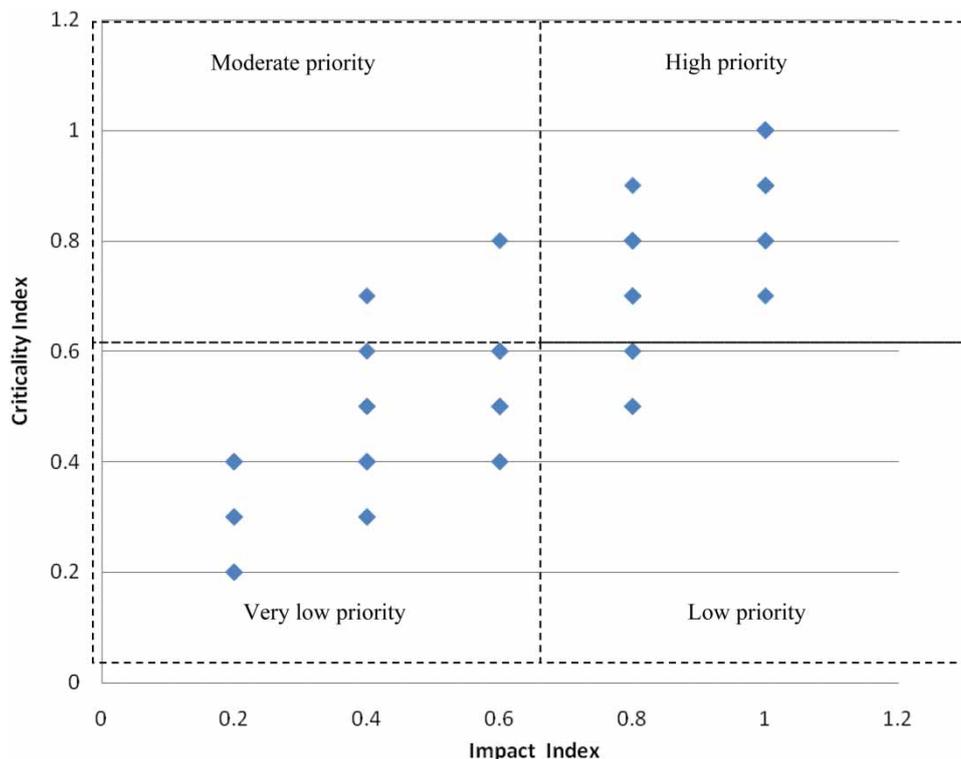
Figure 4 | Criticality vs. impact.

Based on the plot, a reactive or proactive strategy can be recommended. The results show that 36% of the pipes can be classified to have very low replacement priority, 16% low replacement priority, 31% high replacement priority, and 3% moderate replacement priority. In case of limited resources, rehabilitation options should commence with the high priority group. Prioritization of pipes' criticality is a suitable approach for identifying pipes and pipe cohorts amidst limited financial resources.

The decision maker, depending on the available funds, may choose a minimum risk index as a threshold for which the blocks above that factor are considered for rehabilitation and replacement. The most critical priority group of high replacement priority consists of 31 blocks. Of the 31 blocks, 1818 is the most critical with a water sales rank of 1 and block 2221 the least important. Block 1818 has average monthly water sales of 52,000 m^3, which translate into USD 44,200 based on a converted NWSC water tariff of USD 0.85/m^3 being considered. Block 2221 has average monthly water sales of 14,152 m^3, worth USD 12,029. The big difference in revenue collections between the two blocks shows that block 1818 is much more critical than block 2221.

CONCLUSIONS

The study focused on comparing the performance of regression modelling and the Markov-based probabilistic (risk-based) approach in the prediction of the future condition of water mains amidst data-scarce situations. The study also developed a risk-based mains replacement strategy that combines future condition and criticality of the pipe network to prioritize rehabilitation requirements for a block. All the developed approaches have been applied to a case study as a proof of concept. The findings, discussions and conclusions of the study are outlined in the sections that follow.

Statistical regression analysis-evaluation of the condition of a pipe network based on age, diameter and break history

Statistical regression analysis for the Kampala City Centre Water Distribution System (KCCWDS) pipe cohorts of 40–80 mm, 100–150 mm and 200–600 mm produced good prediction results that are better and comparable to those in previous studies. The regression model for the 40–80 mm diameter pipes appeared to be the best performing of the three pipe cohorts' statistical analysis. It should be noted that R^2 increases with reduction in pipe diameter, which could be attributed to differences in sizes of data sets used. The developed regression models were used to predict the future condition states for the pipe network per block and the results show 54 out of the 69 forecasts (78.26%) for the regression model. The statistical regression model has to be an appropriate model for predicting pipe conditions in a water network. The approach can be used by planners and engineers to predict the future condition of pipes particularly in data scarce scenarios.

Break pattern analysis assumed that trends witnessed in the available break data from the past can be used to estimate pipe breaks in the future. However, such an assumption may not hold for very prolonged forecasts because the developed regression models implicitly consider the wear out phase in the bathtub curve and may not give accurate pipe break estimates for pipes in earlier phases of their lifecycle. It is advisable for water utilities to keep records of break data to eliminate the need for filling missing data.

Filling in missing historical pipe break data using regression analysis necessitates the availability of reasonably significant observed break data to enable credible pattern analysis and to render the approach credible. This is a major challenge because networks with few pipes may not have enough data registered.

A limitation of the proposed model is the requirement for large amounts of quality data. Management particularly in developing countries should develop and strengthen data collection; for example, the years of installation, the service life for different pipe materials under local conditions, as well as the break data (date and location) of the pipes. Data collection should be disaggregated as much as possible to improve the accuracy of the models. The regression is limited to physical indicators such as age. Further research is required where additional parameters such as environmental and

operational indicators are included in the regression model. However, collection of this data can be very expensive.

Prediction of the future condition of a pipe based on a Markov chain

The Markov model approach has shown both conceptually and through statistical analyses, to be an appropriate model for predicting pipe cohort conditions in a water network. The results show that 61 out of the 69 forecasts (88.4%) for the Markov model were accurate. The study methodology can be used to model the future condition states of a water distribution network. Application of the model for long term forecasts may be inaccurate because deterioration-causing conditions, such as climate, over-burden, pressure regimes, can change since buried infrastructure assets have typically long lives. The approach can help water utility managers optimize maintenance and repair decisions amidst budget limitations whilst taking into consideration both current and future states of the pipe network.

Comparison between performance of Markov and statistical regression models in the prediction of the future condition of a pipe

In this paper, the challenge of which model to use by water utilities with limited data is addressed. We examined the performance of Markov and statistical regression models in the prediction of the future condition of a pipe when data is scarce. The study revealed that a Markov-based approach is more accurate than conventional regression models in predicting future condition of a pipe network. However, the down side to the Markov approach is that it requires engineers with strong mathematical skills, which may be rare. The study shows that with limited data Markov models are able to yield more accurate results than statistical models.

REFERENCES

Ana, E. & Bauwens, W. 2010 Modeling the structural deterioration of urban drainage pipes: the state-of-the-art in statistical methods. *Urban Water Journal* **7**, 47–59.

ASCE 2017 *2017 Infrastructure Report Card*. American Society of Civil Engineers.

Asnaashari, A., McBean, E. A., Shahrour, I. & Gharabaghi, B. 2009 Prediction of watermain failure frequencies using multiple and Poisson regression. *Water Science & Technology: Water Supply* **9**(1), 9. doi:10.2166/ws.2009.020.

Baik, H.-S., Jeong, H. S. & Abraham, D. M. 2006 Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. *Journal of Water Resources Planning and Management* **132**, 15–24.

Bannerjee, S. & Morella, E. 2011 *Africa's Water and Sanitation Infrastructure: Access, Affordability, and Alternatives*. The World Bank, Washington, DC, USA.

Boulos, P. F. 2017 Optimal scheduling of pipe replacement. *Journal of the American Water Works Association* **109**, 42–46.

Clark, R. M., Stafford, C. L. & Goodrich, J. A. 1982 Water distribution systems: a spatial and cost evaluation. *Journal of the Water Resources Planning and Management Division* **108**, 243–256.

Da Costa Martins, A. D. 2011 *Stochastic Models for Prediction of Pipe Failures in Water Supply Systems*. Instituto Superior Tecnico, Lisboa, Portugal.

Davies, J., Clarke, B., Whiter, J. & Cunningham, R. 2001 Factors influencing the structural deterioration and collapse of rigid sewer pipes. *Urban Water* **3**, 73–89.

Echart, J., Ghebremichael, K., Khatri, K., Mutikanga, H., Sempewo, J., Tsegaye, S. & Vairavamoorthy, K. 2012 *The Future of Water in African Cities: Why Waste Water?* Integrated Urban Water Management, Background Report.

Economou, T., Bailey, T. & Kapelan, Z. 2008 *A Hidden Semi-Markov Model for the Occurrences of Water Pipe Bursts*. no. July, 2008.

El-Abbasy, M. S., Senouci, A., Zayed, T., Mirahadi, F. & Parvizsedghy, L. 2014 Condition prediction models for oil and gas pipelines using regression analysis. *Journal of Construction Engineering and Management* **140**, 04014013.

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

Folkman, S. 2018 *Water Main Break Rates in the USA and Canada: A Comprehensive Study*. Utah State University Buried Structures Laboratory

Francisque, A., Shahriar, A., Islam, N., Betrie, G., Siddiqui, R. B., Tesfamariam, S. & Sadiq, R. 2013 A decision support tool for water mains renewal for small to medium sized utilities: a risk index approach. *Journal of Water Supply: Research and Technology – AQUA* **63**, 281–302.

González-Gómez, F., García-Rubio, M. A. & Guardiola, J. 2011 Why is non-revenue water so high in so many cities? *Water Resources Development* **27**, 345–360.

Ji, J., Robert, D., Zhang, C., Zhang, D. & Kodikara, J. 2017 Probabilistic physical modelling of corroded cast iron pipes for lifetime prediction. *Structural Safety* **64**, 62–75.

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

Kettler, A. & Goulter, I. 1985 An analysis of pipe breakage in urban water distribution networks. *Canadian Journal of Civil Engineering* **12**, 286–293.

Kim, J. W., Choi, G., Suh, J. C. & Lee, J. M. 2015 Optimal scheduling of the maintenance and improvement for water main system using Markov decision process. *IFAC-PapersOnLine* **48**, 379–384.

Kleiner, Y. & Rajani, B. 1999 Using limited data to assess future needs. *Journal-American Water Works Association* **91**, 47–61.

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

Kleiner, Y., Sadiq, R. & Rajani, B. 2006 Modelling the deterioration of buried infrastructure as a fuzzy Markov process. *Journal of Water Supply: Research and Technology-AQUA* **55**, 67–80.

Kutyłowska, M. & Hotloś, H. 2014 Failure analysis of water supply system in the Polish city of Głogów. *Engineering Failure Analysis* **41**, 23–29.

Kyokaali, L. 2018 *Prediction of the Future Condition of A Water Distribution Network Using A Markov Based Approach: A Case Study of Kampala Water*. MSc dissertation, Makerere University, Kampala, Uganda.

Le Gat, Y., Kropp, I. & Poulton, M. 2013 Is the service life of water distribution pipelines linked to their failure rate? *Water Science and Technology: Water Supply* **13**(2), 386–393.

Lippai, I. & Wright, L. 2005 *Criticality Analysis Case Study: Zone 7 Water Distribution System*. ASCE.

Lusiba, J. M. 2003 *Prioritisation of Replacement for Mains in Aged Water Supply Networks: Case Study of Masaka Municipality Network, Uganda*. ITC.

Malm, A., Ljunggren, O., Bergstedt, O., Pettersson, T. J. & Morrison, G. M. 2012 Replacement predictions for drinking water networks through historical data. *Water Research* **46**, 2149–2158.

Mutikanga, H. E. 2012 *Water Loss Management: Tools and Methods for Developing Countries*. IHE Delft Institute for Water Education, Delft, the Netherlands.

Nwsc 2014 *NWSC Annual Performance Report: Financial Year 2014–2015*. National Water and Sewerage Corporation, Kampala.

Ossai, C. I., Boswell, B. & Davies, I. J. 2016 Application of Markov modelling and Monte Carlo simulation technique in failure probability estimation – a consideration of corrosion defects of internally corroded pipelines. *Engineering Failure Analysis* **68**, 159–171.

Pelletier, G., Mailhot, A. & Villeneuve, J.-P. 2003 Modeling water pipe breaks – three case studies. *Journal of Water Resources Planning and Management* **129**, 115–123.

Rogers, P. D. & Grigg, N. S. 2009 Failure assessment modeling to prioritize water pipe renewal: two case studies. *Journal of Infrastructure Systems* **15**, 162–171.

Saegrov, S. 2005 *CARE-W: Computer Aided Rehabilitation for Water Networks*. IWA Publishing, London, UK.

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

Scholten, L., Scheidegger, A., Reichert, P., Mauer, M. & Lienert, J. 2014 Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis. *Water Research* **49**, 124–143.

Sempewo, J. I. & Kyokaali, L. 2016 Prediction of the future condition of a water distribution network using a Markov based approach: a case study of Kampala water. *Procedia Engineering* **154**, 374–383.

Shamir, U. & Howard, C. D. 1979 An analytic approach to scheduling pipe replacement. *Journal-American Water Works Association* **71**, 248–258.

Skipworth, P. J., Engelhardt, M., Cashman, A., Savic, D., Saul, A. & Walters, G. 2002 *Whole Life Costing for Water Distribution Network Management*. Thomas Telford, London, UK.

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

Uchwat, C. & Macleod, D. 2012 Case studies of regression and Markov chain models. In *Proceedings of the 2012 Conference and Exhibition of the Transportation Association of Canada-Transportation: Innovations and Opportunities*, New Brunswick.

Uganda Bureau of Statistics (UBOS) 2014 National population and housing census. UBOS, Kampala, Uganda.

Vairavamoorthy, K., Yan, J., Galgale, H. M. & Gorantiwar, S. D. 2007 IRA-WDS: A GIS-based risk analysis tool for water distribution systems. *Environmental Modelling & Software* **22**, 951–965.

Vloerbergh, I. & Blokker, E. 2009 Failure data analysis-a Dutch case study. In: *Strategic Asset Management of Water Supply and Wastewater Infrastructures* (And, H. A. & Almeida, M. D. eds). IWA Publishing, London, UK, p. 433.

Walski, T. M. & Pelliccia, A. 1982 Economic analysis of water main breaks. *Journal-American Water Works Association* **74**, 140–147.

Wang, Y., Zayed, T. & Moselhi, O. 2009 Prediction models for annual break rates of water mains. *Journal of Performance of Constructed Facilities* **23**(1), 47–54.

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

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

Xu, Q., Chen, Q., Ma, J. & Blanckaert, K. 2013 Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network. *Journal of Hydro-Environment Research* **7**, 134–140.

Zhou, Y. 2018 *Deterioration and Optimal Rehabilitation Modelling for Urban Water Distribution Systems*. CRC Press/Balkema, Leiden, The Netherlands.