Decoding Leakage Tendencies of Water Pipelines in Dolomitic Land: A Case Study of the City of Tshwane

Achieng G. Ogutu, Okuthe P. Kogeda, and Lall Manoj

Abstract— Increase in service demand due to rapid development in urban areas has extended the amount of pressure exerted on water distribution networks; thereby, causing an acute upsurge in the number of pipeline failures experienced. Failure occurrence in fiddly dolomite lands, however, may be fatal considering high levels of ground instability in such regions. It is therefore important to understand the tendencies of pipe failures in these regions, as this would enhance utilization of failure trends in predictive modeling of pipeline leakage. This however, may seem complicated, given the magnitude of inherent uncertainties accompanying pipe failure and scarcity of failure data. Nonetheless, such uncertainties can be addressed through strategic combination and utilization of facts, knowledge and auxiliary information, which can be tackled using predictive models like Bayesian Networks. In this paper therefore, we present an overview on how data uncertainty can be handled using BNs. We also identify failure tendencies and inherent uncertainties from a set of maintenance data from the City of Tshwane, and address how we intend to handle the uncertainties for effective failure prediction. Finally, we present a breakdown of data integrity concerns identified during the analysis, which warrants further research. Among the results we discover that most leakages occur along street corners and road intersections; and that small diameter pipes are most prone to failure. These analyses however, precede an intense uncertainty modeling process that is to be conducted. Information presented herein may be used to produce models for predicting pipeline leakage subject to historical failure.

Index Terms— pipeline failure, Dolomitic land, leakage prediction, distribution networks, Bayesian networks

I. INTRODUCTION

The importance of water distribution networks to urban settings that are thriving with countless number of socio-economic activities and a sharp population increase is a subject that warrants no assumption [1], [2]. These networks determine the productivity levels and also support the essence of communal safety and wellness in these regions [2], [3]. With such a high level of operational weight [3], they are likely to experience a number of breakdowns, which in most cases, are manifested through the number of leakages reported by water utilities [4], [5]. Pipeline failure refers to the “unintended loss of pipeline contents” [6], and leakages stand out as the most common form of failure that affect water distribution pipes [3]. They are inevitable [7] and are highly destructive [7–10], thereby, prompting utility management to be on the constant lookout for effective leakage minimization strategies and techniques [2]. Water leakage results in a variety of negative effects, including property damage, environmental pollution and disturbance to other infrastructure [4], [8], [11]. They also lead to a continuous reduction of the distribution networks’ reliability [12]. These effects consequently minimizes the ability of water facilities to meet their operational goals and targets [2].

Like a typical urban setting, the City of Tshwane (CoT) in Gauteng Province, South Africa boasts of a myriad of human activities and high levels of infrastructural development; underground pipeline facilities being one of them. Approximately, 23% of the total land in the city is Dolomitic [13], [14]. Dolomitic land refers to a piece of land that is underlain by dolomite rock, either directly or at shallow depths of possibly below 100m [13], [15]. These rocks are soluble in water [13–15]. Therefore, in the presence of pipe leakage, they may dissolve, leading to creation of voids and cavities within them. As a result, sinkholes may be formed, resulting into massive ground movements [13], [15] that are accompanied with fatal outcomes. Continued urban development however, has resulted in several sections of water distribution networks being lined in these Dolomitic regions [16], an aspect that calls for maximum attention. Determination of pipeline failure tendencies in this region is therefore very relevant, as this would encourage exploitation of these trends that can be included in predictive modeling of pipeline leakage, with the aim of leakage reduction.

According to the research reported in [28], there lacks an adequate level of understanding when it comes to the factors and processes that collectively lead to pipeline failure. In addition to this, timely identification of failures also comes as a challenge to a number of utilities [5], an aspect that lead to imprecision in capturing and recording of pipe failure history [5], [29]. These two aspects therefore, stand out as the highest contributors to uncertainty in pipe failure data.
II. FACTORS AFFECTING PIPELINE FAILURE IN DOLOMITIC LAND

There exists an extended body of literature that attempts to analyze various factors that influence pipeline failure [4], [17], [18], [19]. The different factors analyzed by the abovementioned collection of literature talks about the factors that may influence pipeline failure in general. However, additional factors, specific to Dolomitic lands are relayed in [17]. Categorized as a region with collapsible tendencies [20], the soil coverage in dolomite regions are considered to undergo abrupt volume reduction in the presence of sufficient triggering mechanisms. Because of the abrupt reduction in volume, massive ground movements may be experienced. This in turn, may cause pipe breakages and cracks [13], [17], [20]. Pipe bedding or backfill [17], [20] also plays an important role with regards to support and protection of a pipe. In addition, thickness of pipe walls is also identified as an important determinant of its failure tendencies [17]. This is because the thickness of the pipe determines the amount of pressure it may handle in with regards to external load [20], [22] as well as longitudinal and circumferential deflections [17]. Last but not least, the safe span distance of a pipe is also considered to influence failure, given its strong correlation to a pipes wall thickness and diameter; factors that equally affect failure [13], [17], [20], [22].

III. REVIEW OF UNCERTAINTY MODELING IN PIPE DATA

In most cases, data recorded by utilities regarding pipeline failure tend to be either incomplete or contain unreliable information. Sometimes however, the data may be completely false [5], [29], [30]. These factors stand out as the greatest contributors to data uncertainty, hence creating difficulties in estimating failure trends in pipelines [30], [31]. Availability of predictive models fit enough to handle these uncertainties however, brings back hope to both developers and utility management. These models are dynamic enough to be used in incidences where data availability is limited, as well as in incorporation of both theoretic and expert knowledge during the modeling process [3], [30].

Through utilization of available historical records of pipeline data, these models are able to make predictions concerning the future behavior patterns of the pipelines [28], [29], [32]. The patterns identified are assumed to continue into the future, hence are used to predict failure probabilities in the future [29], [32]. This forms the basis of predictive modeling. Therefore, in situations where data availability is limited or incomplete, incorporation of expert or engineering knowledge would enhance modeling. One appropriate technique that may be employed to handle such a scenario is the use of Bayesian Networks (BNs).

A. Overview of Bayesian Networks

Bayesian Networks (BNs) are graphical models that principally present a system as a network created by interactions of different variables, commencing from a primary cause to the final outcome of an event. Their structures consist of nodes, which represent the system components; and arcs that link the nodes to show probabilistic dependencies between them [33 – 35]. BNs have the ability of integrating several scenarios, interactions and outcomes as well as swapping efficiently between investigation parameters. They also stand out, given their ability to utilize combinations of data and knowledge to model uncertainty, and in their ease of applicability in handling missing data [34], [35]. These properties have made BNs to be frequently considered in modeling environmental systems [35]. Typically, the ultimate system behavior (success or failure) of a BN is determined by the interaction among its components [34]. Fig. 1 illustrates how components of a BN interact for the ultimate success or failure of a system.

In Fig. 1, independent components (Corrosion, Performance, Number of Previous Failure (NoPF) and Manufacturing defect) collectively affect fracture. Dependent component(s) (Thinning of pipe wall) however, is a child to its causal parent, but nonetheless, equally affects fracture. Probabilities in BNs can be computed using the Bayes’ theorem, which also enhances appropriate assignment of conditional probabilities [3].

B. Review on Applications of Bayesian Networks in Pipeline Failure Modelling

There exists a significant body of literature that talks about how BNs have been utilized in predicting pipeline failure, given their enormous ability to handle data uncertainty. In [1], BNs are engaged in assessing deterioration rate of pipes. This is achieved by using Bayesian configurations together with prior knowledge of water assessment procedures, which are then applied to generate weights of influence of failure factors. In [3], BNs are applied to assess the risk of failure of metallic pipes, a process that highly embraced a large combination of numerous factors influential to failure through Bayesian inference. Combination of BNs together with other modeling techniques for an ultimate model creation is also demonstrated in [36] and [37] where BN are coupled with a data oriented Scoring model and Fuzzy Logic (FL) respectively, showcasing diversity and flexibility levels of BNs applicability. Additionally these models indicate that BNs are used to determine a diverse range of aspects regarding pipeline failure including risk of failure [36], [3], pipe deterioration rates [1], pipeline safety [37] among others. In this paper though, the analyses performed are carried out in preparation for the subsequent uncertainty modeling that is to be conducted, where BNs will be applied.

---

Figure 1: Simple Bayesian Network

---
to predict leakage incidents based on the dataset discussed below.

IV. THE CASE STUDY

Pipeline maintenance records and data pertaining to pipe network description were collected from Dorinkloof suburbs in the City of Tshwane (CoT), Gauteng province, South Africa. The reason for a regional focus was to decrease several potential variations in the data ranging from environmental to material differences, as well as construction practices so as to minimize uncertainty. In the subsequent sections, we provide a detailed description of the data obtained from the study area. To begin with, a brief history of the study area is provided. The approximate network capacity, number of recorded failure, frequency of failure in different pipe materials and sizes, among other descriptions are reported.

A. The Study Area

Located in fairly populated and sumptuously developed eastern regions of Pretoria, South Africa, the area contains an estimation of about 1,100 links of distribution pipes that form part of the water pipe network in the region. A link refers to a fraction or a pipe segment that basically forms part of a larger pipe length. In the area of study, the distance between the start and the end of any given link is considered as a complete pipe segment. Any rehabilitation, repair or replacement activity conducted on a link is recorded in the maintenance database, and is identified using the Municipal’s addressing blocks, also known as Erf numbers. The Erf numbers, together with street names and street numbers are then used to determine failure locations along a pipe. Fig. 2 shows the pipeline network in the study area.

In Fig.2, an Erf is represented by the symmetrical shapes (mostly rectangular blocks or divisions). Every Erf is placed strategically along a street, basically to ease accessibility. All distribution pipes (in blue) are lined along the street (along pedestrian sidewalks). Service pipes (short or small red connectors) are then used to create a connection between the Erf and the distribution networks. Bulk or trunk water mains are however represented using solid (thick) red color scheme as illustrated in Fig.2.

The water pipe network in the study area is composed of four different types of pipe materials. These materials include Asbestos-Cement (AC) pipes, High-Density Polyethylene (HDPE) pipes, Steel (ST) pipes and unplasticized Polyvinyl Chloride (uPVC) pipes. However, different flavors or versions of these pipes are collectively used in the network as shown in Fig. 3.

AC pipe have been considered to be able to resist attacks from various aggressive corrosive effects of water that commonly affect metallic pipe material [23]. However, these pipes also undergo leeching, a degenerative chemical reaction, which may be compared to electrochemical corrosion in iron pipes [22], [24].

HDPE pipes are however mentioned as the most extensively used plastic pipe material by utilities around the world [25], [26] and are highly popular in transportation and distribution of natural gas. About 80% of new pipe installations currently are of HDPE pipes [26], an aspect that is also noticed in our study area. uPVC and steel pipes are equally reported to be of high exploitation in water distribution networks, thanks to their abilities to withstand high levels of water pressure [27].

According to the values presented in Fig.3, uPVC pipes collectively account for an approximated 45% of the total network material composition, followed by HDPE pipes. Steel and AC pipes account for the lowest composition, representing 15% and 5% respectively.

The oldest pipes in the region were installed in the late 60s. Fig.4 represents the material composition with respect to age. AC-D and ST-A4.5 pipes stand out as the oldest pipe materials in the network, installed between the late 60s all through the 70s. no HDPE and uPVC pipes are observed in this period. uPVC-12 and HDPE-100-12.5 pipes however dominate in 2010 and 2011 periods, a period where no AC and minimal Steel pipes are observed. This may be concluded, with assumptions, to be a result of pipe replacement procedures.
8. Data description

Pipe maintenance records of the study area for a period of 5 years were obtained for analysis. Over the period in question, approximations of 350 leakage incidents were reported; among which about 59% involved repairs of different levels of distribution pipes. Fig.5 gives more details about the reported leakage incidents.

Fig.5 indicates that the highest number of repairs involved correction or repair of distribution pipes (59%), repair of service pipes (21%) and repair of hydrants (10%) respectively. The lowest performed repairs however involved repair and replacement of valves, which accounted for 0.63% and 0.94% respectively, as well as repair of 300-600mm pipes, which also accounted for 0.94% of the total reported breakdown. However, the focus of this research is on leakage along water distribution mains, of which about 126 incidents were reported, and respective repairs conducted are shown in Fig.6.

Water distribution pipes in the area of study are classified in three categories based on the size of their diameters as depicted in Fig.6. They include 300-600mm pipes, 200mm and 100mm piped. Out of the three categories of the pipes, the highest number of repairs activities was performed on pipes with 100mm diameters, accounting to about 93% of the total repairs conducted. The lowest number of repairs however, was performed on pipes with diameter size of 300-600mm, accounting for only 1% of the total repairs.

All the repair activities performed on the distribution networks as highlighted in Fig.6 took place within the periods specified in Fig.7, within the five year time frame of the data collected. Majority of repair were performed from 2011 to present, followed by the quantity of repairs performed between the 90s and 2010.

C. Observations

In addition to the data described herein, we discovered that majority of failures occurred along the corner sections of various streets as well as at intersections of busy traffic roads. This aspect may be explained to as a result of additional external pressure (due to traffic load) exerted upon pipelines [17], [20], [22]. It equally noteworthy to mention that majority of leakages repaired along distribution pipes involved small diameter pipes of 100mm, while the lowest repairs were in pipes of 300-600mm diameter size. This somehow confirms a number of conclusions from various literature [20], [22], [28], that Small diameter pipes are more prone to failure compared to large diameter pipes, given the fact that they are often subjected to extra load as they are commonly used in residential regions, with an added possibility that they may be weaker in construction. Finally, majority of repairs are noticed on pipe sections that were installed in the period between 1969 and 1979, and then later to pipes installed form 2010.

V. PROPOSED METHODOLOGY

Having examined the failure history in the study area and identified the different variables affecting pipeline failure, in the dataset and in literature we intend to adopt factor weighting to account for uncertainty. Based on the knowledge gathered from existing literature, informative
prior weights will be attached to the failure factors. Investigations on factor weights reported in [1] indicate that the quantity of weight associated with a particular causal variable greatly affects the overall condition or state of a model. This process will precede general pipe deterioration condition estimation as illustrated in (1), a general equation for estimating the deterioration rate of pipelines [38]

\[
G = \sum_{i} \lambda_i \times x_i,
\]

Where \( G \) is the pipe deterioration rate, \( i \) is the respective variable; \( I \) the total number of variables, \( \lambda \) representing the variable weight and finally, \( x \), the factor position. Taking advantage of parameter fitting possibilities provided by BNs, we will incorporate this estimation into the Bayesian theorem for effective estimation. The unknown parameter \( \theta \), in this case, leakage incidence and any other unidentified value are taken to be random variables. Their distributions are derived from prior information and newly available data [39]. This is basically enclosed in a probability function (2):

\[
P(\theta | y) = \frac{P(\theta)P(y | \theta)}{P(y)} = \frac{P(\theta)P(y | \theta)}{\int_{\theta} P(\theta)P(y | \theta) d\theta}.
\]

Where \( \theta \) is the factor to be estimated, \( P(\theta | y) \) represents the posterior likelihood of \( \theta \), \( P(\theta) \) represents the prior probability \( \theta \), while \( P(y | \theta) \) is the probability function for the possibility of condition \( y \) occurring based on response from the assumed causal associations and the predictor variables.

VI. DATA QUALITY CONCERNS

Major quality concerns with the data obtained from the region were largely related to unknown or unclear timeline information, especially those pertaining to replacement procedures as well as location information. Replacement dates are basically not known even though it was clear that majority of pipelines had been replaced, given the change in pipe material patterns. In addition, information about material switch was equally missing and unidentified during the analysis period.

Identification of the failed sections of pipes was estimated using block and Erf addresses provided by various service providers and assigned contactors who detailed their repair activities as well as through the use the municipal pipeline network maps that were provided. However, this information alone was not sufficient enough for identification of actual failure location in a given pipe segment. In some cases, the failure location listed in the failure record did not match any neighboring pipes in the municipal network map. It was however, assumed in these instances, that the failure record accurately described the pipe location, which was used to update our records.

These assumptions however taken into consideration may not be very helpful or significant in situations where pipeline failure modeling is dependent on critical variables such as pipe material. This is because inclusion of these parameters may introduce biasness into a model given the fact that pipes with fuller records are the only ones that would provide positive values for a particular model.

However, through expert elicitation, missing installation and replacement dates may be assigned using different techniques. An example of how this may be achieved is through introduction of binary categorical variables to represent the missing data and assumed timelines, especially those that pertain to the replacement dates, and replaced materials. The categorical variables may therefore be used to account for the assumptions made.

VII. SIGNIFICANCE OF THE ANALYSES

Determination of pipeline failure trends especially in regions considered as high risk zones [21] is highly necessary, especially when it comes to prioritization of rehabilitation, repair and replacement activities on pipe facilities in these regions. Prioritization with regards to assessments as well as resource allocation and rehabilitation optimization is equally enhanced if the failure trends are known.

Pipeline failure trends also enhance the ability of utilities to set distribution network requirement specifications with regards to production, installation and operation of these facilities. The trends also improve our understanding of the various levels of uncertainty that accompany pipeline failure in a given region. This may therefore give room for determining various ways on how accuracy and precision during modeling of these failures can be achieved, especially when expert opinions are included.

VIII. CONCLUSION

The data presented for these analyses was availed from a single geographical location, which to a certain extent; represent a number of similarities to the network. These may be in the form of environmental concerns (pipe surrounding), installation guidelines and procedures, among others. However, some aspects of uncertainty were still spotted, especially regarding specific location of various leakages along pipe segments, and not forgetting incidents of missing data.

Nevertheless, this research highlights the importance of recording and keeping of historical failure data by utilities. Utilities are therefore encouraged to conduct leak detection campaigns as well as planned water audits have to enhance identification of possible failures. Additionally, application of basic management procedures involving inspection and data gathering are also of high importance.

Making assumptions during failure modeling is indispensable, given the number of situations that may be unattainable during model development. However, in uncertainty modeling, acknowledgement of assumptions is a necessity. Even so, availability of processes and techniques of achieving acceptable levels of assumption acknowledgement are scarce. We therefore recommend extended research aimed at determining various techniques of tackling assumption acknowledgement in uncertainty modelling.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science, Faculty of ICT at the Tshwane University of Technology (TUT), with whom this research was undertaken. The authors also thank the City of Tshwane (CoT) for providing the data that was used in this research work.
REFERENCES


[17] Nel 2011


