Proposed probabilistic models of pipe failure in water distribution system

Thi Minh Lanh Pham^{1,2,}, Hai Ha Pham², Nguyen Anh Thu Do¹, and Dinh Hong Le¹*

¹Ho Chi Minh City of Technology University, 268 Ly Thuong Kiet Street, Ward 14, District 10, Ho Chi Minh City, Viet Nam

²Ho Chi Minh City of Architecture University, 196 Pasteur Street, Ward 6, District 3, Ho Chi Minh City, Viet Nam

Abstract. All pipes in water supply network are installed underground, so it is difficult to identify pipe failure location during the operation of a system. Prediction of the risk of pipe failure in the water distribution systems is necessary for preparation of reparations and displacement of a pipe network system. Based on the probability of pipe failure, it will be possible to save money and labor cost for water supply companies. Many studies have been conducted on this topic, some of which used experimental models, others used statistical models in which recently many authors used regression model, but almost all the models come up with calculating the pipe failure rate per unit length of pipe in a year. It is not a direct probability of pipe failure. This article reviews various methods to evaluate pipe failure in water distribution systems. Based on that, the authors proposed two models: Regression Logistic Model and Decision Tree Model that would support an effective decision making for detecting the pipe failure and proposing appropriate solutions.

1 Introduction

Most of pipes in urban water distribution system are set up underground or under pavements. When pipe failure occurs, a remarkable amount of water is loss without proper detection and it also causes pathway for microbial contamination of distribution system which leads to water-borne disease to end-users [1]. Although the number of failures on the water distribution systems are recorded daily, extra budgetary needs for inspecting pipe leaks and failures on water supply system where pipe diameter of 150,200 and 300mm are still remaining[2].

It is crucial to propose a reasonable plan for water supply system maintenance, as well as timely replacement of pipes that have the likelihood of failure or less productivity. The important factor of this process is the ability to precisely predict pipe failures on water distribution system and identify which contributing factors responsible for making the system stable and long lasting working.

Pipe failure is a process when pipe materials are becoming fatigue and cracks appearing on pipe wall and then expanded to a failure limit. There are many kind of cracks. Within this

* Corresponding author: phamthiminhlanh@gmail.com

study, the geometry of the cracks, which are vertical cracks, horizontal cracks and circumferential failures are neglected. Pipes failureis supposed to occur under normal conditions. Other abnormal incidents andfailures at the connection fittings are diminished.

- This paper focus mainly on:
- Reviewing, analysis and classification the development of various studies on pipe failuresand their methods to evaluate the probability of pipe failure in those studies
 - Proposing two models for studying pipe failure probability
 - The advantages and limits of proposing models.

2 Literature review

In Figure 1, it is obvious that pipe failures based solely on internal corrosion and surrounding soil conditions. In addition, pipe materials are degraded due to dead loads and live loads. Historical failure record is also one of contributing factors to assess future performance of water supply pipes. There have been many research projects conducted for better understanding of the magnitude of utility’s pipe failure problems and predicting pipe failures, the results from which are used as reference elements to the rehabilitation and maintenance of water mains. As time goes by, research topics and methods are also being extended and studied in an overall manner and the results are simulated to best fitted to practical conditions.

Fig. 1. Literature review on studies of pipe failures on water supply networks.

2.1 Analyzing the corrosion rate of water mains

Corrosion is the result of electrochemical oxidation of materials when being placed in natural condition. In his earliest published study in 1948, Bubbis discussed by T. Wengström[3] investigated on 40 year-old cast-iron pipes and had signs of extremely bad corrosion. The result of this study has been applied for later research to identify the reasons of corrosion on material surface. Following that, the impacts from surrounding environment where the pipes were laid and the working condition inside the pipes are most common causes of corrosion. Studies have been focused on analyzing relevant factors including soil conditions and water inside pipe networks. In R.B. Petersen study[4] analyzed factors that effect the corrosion of iron in soil include: soil type, moisture content, the degree of aeration, resistivity, pore water chemistry, microbiological activity and temperature.

Vladimir Kuceraproposed a method to calculate corrosion rate based on the age of pipe and focused on analyzing two variables, which is (i) total time of wetness in a year and (ii) the chemical and physical properties of the corrosion products as discussed by RehanSadiq[5]. In this method, the author used a number of constant values and those needed to adjust in accordance to specific areas. According to the previous studies, pipe failure caused by external corrosion is obvious, within which soil in mixture of clay and sand has the major impact[6].

Rajani (2000) [7] conducted his study in a more detail way when taking into account both internal and external corrosion on pipes. The author used exponential fast-developed corrosion model and linear slow-developed corrosion model. In another related study results of those models are compared, it was proposed that corrosion pit depth was developed rapidly at the beginning stage but gradually slow down at later stage[8]. A similar conclusion on the probability of pipe failure was also stated in H. Rezaei[9], which was: 'Pipe failures increase at the beginning working stage and gradually decrease later on'. There are many other contributing factors to pipe failures as well, including material properties, manufacturing defects, improper installation and laying location of water mains[6].

The probability of pipe failure due to corrosion is calculated using time as a representative variable. Some studies yielded the average age of ductile cast-iron pipes tend to lower up to 3 to 7 years when compared to other materials and the age of failure for ductile cast-iron is 13 years, whilst in another study is 32 to 40 years. Wengström [3] concluded that uPVC material is more susceptible to failure at the beginning stage than steel material.

2.2 Assessment of water mains loading

The age of pipes decreases due to factors like corrosion and failures associated with its wear, as well as changing in the material from which the pipe network is made. Those changes occur when the material is bended from external impacts and loads, some of which are earth moves, natural hazards or other random events [10] and internal pressure fluctuations. Many studies were conducted on grey cast-iron and being developed by Sadiq[11]. The author used excavated cast-iron pipes from the water distribution network of Toronto, Ontario. Mechanical properties of materials were tested including tension, compression, ring bearing and full-scale longitudinal test. The author concluded that aged pipes are more susceptible to failure under loading than the others. The probability of pipe failure due to loads which are constant factor.

There are big differences between pipe working under dead loads and live loads. Pressure fluctuation, which is understood here as any alteration in pressure level inside the pipe network, has complex impacts to pipe materials [9]. Hossein Rezaei proved that longitudinal cracks have a positive correlation to pressure variation, whilst circumferential cracks have

negative correlation. In addition, if cyclic loadings are strong enough, there might occur main failures and shorten pipes age.

2.3 Assessment of physical conditions

As discussed by T. Wengström[3], diameter, length, material and laying location of pipelines are not considered as impact factors that would contributed to the longlife of pipes. Those paramaters are used to classify the working capacity of pipes. Studies before 1982 focused on grey cast-ion pipes only due to the fact that 80 to 95% of materials used in pipe networks are grey cast-iron. Other studies at that time focused on seperated pipe lines which were lack of comprehensive results. It was until 1990 that Kootmann extended his study to other pipe materials and proved that cast-iron and iron pipes are most susceptible to failures.

uPVC material is easy to failure when early installed; Cast iron tends to appear more failures than material of Amiang concrete, PVC, Di, PE and others on the data collected during 10 years[9]. Broken pipes are related not only to the pipe material variable but also to diameter, length and pipeagevariables. The possibility of pipe failures in pipeline networks of big cities are higher when diameter of pipes is bigger than 300 mm. However, this result depends on the percentage of residents living in that city and the area of which pipes are laid[3].Bayesian modelis used to evaluation the probability of pipe failure depend on pipe length and pipe age as variables [12].

2.4 Historical failure record

Previous failure pipe record numbers are considered as significant factor to the probability of pipe failure occurrence. A survey conducted at the pipelines network of Birmingham showed that bad sections of pipe are needed to replace, otherwise repairing cost would be much higher than replacement cost[3]. Based on this finding, Walski [13] studied the correlation of failures rate with the frequency that next failures would occur. The results revealed that there is a correlation of failures frequency with number of previous failures. A similar finding from Gorjibandpy [14] again stated that the period of future failures occurrence would be shortened in compared to previous failures.

2.5 Prediction of pipe failure using statistics method

Pipe failures are the results of the above-mentioned factors. In order to assess those factors using statistics method, researchers normally use exponential function model, linear regression, Bayesian diagnostic and Possion. Each of the model has its own strength and depends on which type of survey data.

It was assumed at the beginning that the frequency of pipes failures is associated with pipe working age. The author [2]proposed an exponential model to describe this correlation. This model was continuously developed by Ossman in 2011 [15]. He based on the historical failures records to predict the frequency of pipe failures in future. However, in those surveys, factors related to internal working conditions of pipes were neglected, for example: pressure, internal and external environment.

Using Artificial Neural Network Model (ANN) to identified operation of the water distribution systems ability and failiure rates was proposed byAl-barqawi[16]. The author proposed to addclassification analysis in ANN to increase the efficiency of the model[17], the study yielded that the age of pipe have contributed tothe failure conditions at about 20.95%, pipe material is at 17.49% and failure rate is at 13.13%.The disadvantages of ANN are time to run models took so long and the cost for data collection is expensive. Those are the reasons why this model was applied only for academic research purpose.

Regressive Linears Model is applied solely on evaluating the capacity of operational pipe. The result of Andreou, 1987 show that this model can exactly evaluate the risk of broken about 70% [3] and pipe failures depend on pipe diameter, location of laying pipe, prior the number of pipe failure. The age and failure rate of pipe are effected by the second failure.

When comparing effective regression linear model, general regression logistic model, poisson model and exponential model, S. Yamijala [18] proposed using poisson model and exponential model to predict the risk of pipe failure due to the fact that the number of pipe failures can be counted. As such, in those models, the contributing factors are pipe diameter, pipe material, land use purposes (forest, transport area, agricultural land), temperature, rain fall and wet soil. Those factors are the reason leading to pipe failures. From the last failure failure rate, the probability of broken increase with time variable.

Bayesian model has been widely applied in recent studies due to its precise results in compare to previous models. The uncertain variables are used as inputs of the model to simulate effective factors of pipe conditions and soil environment for each materials [19]. It is revealed that physical condition is more important than soil environment. Also, cast iron and ductile iron material have high risk of failure than other factors like temperature and rain fall.

3 Materials and methods

It is difficult to determine the corrosion rate on actual implemented pipes so current studies limit to two materials only, which are cast iron and steel. In addition, most of the water supply pipes are buried underground, so the cost of sampling is very expensive. Moreover, the representation of the samples is not high when the samples of broken are only some of many pipes on the water distribution system. Pipe failure due to corrosion or noncorrosion should be studied more in the future.

The parameter representing corrosion is the pipe age and it has been shown to be correlated with the age of failure and the probability of pipe failure. Moreover, pipes which early failure have remaining age shorter than pipe later failure. Factors like pipe material, soil environment and pipe position are more important than pipe age when evaluating the possibility of failure.

Diameter of pipe is one of the most reliable information to predict the possibility of failure. The authors conducted surveys and the results revealed that the rate of pipe failure increases with diameter for 300mm pipe, but this result is noted that depending on the size of the city surveyed. The operating pressure in the pipe also requires a more detailed consideration of the extent to which the pipe is capable of acting.

From the above contents of the topic, the research step is proposed as shown in Figure 2. The content of the study will focus on effective model and the ability to apply in practical.

Fig. 2. Proposal content of research about the probability models of pipe failure.

4 Results

4.1 Data collection

The data collection is statistical about the pipes on water distribution system including the broken information: position, type of crack, time. From this data, number of broken and the pipe span at the time failure will be specified to input the models. However, records showed date of pipe failure only without any data on pipe installed. Therefore, the age of the failure pipe needs to clarify using information from Geological Information System (GIS). In the GIS, date of installing is available as well as information about the physical condition of pipe networks (material, diameter, length,...). All the data collected will be checked again with the CAD file drawing pipe plan.

Beside the broken data, the model runs better when adding variables on pressure flow inside pipe. This information is getting from the pressure sensor laid on the pipe networks. Sensor is a device to record the pressure value on water distribution on the second time step, if the systems have more sensors then the result will be more precise. It is easy to load the pressure data from the computer but it is not easy to classify each pipe in the system and need have a long time to do this.

4.2 Regression logistic model

In the literature review about failure predictors, the correlation between the risk of pipe failure and the factors that influence on it is mainly analyzed. Analytical objects are often represented by binary variables such as failure / non-failure, occurring / not occurring. Influence factors are continuous variables (pipe age, diameter, length, number of nodes, ...). In previous studies, the regression linear model was modified by converting the pipe failure variable into a continuous variable that is the number of failures per kilometer in a year [19], however, this method has not yet determined the probability failure for each pipe. It only estimates the correlation between the factors of influence and the rate of pipe failure on the network.

Logistic regression is a classification algorithm used to predict binary variables, which may be considered as a special case of linear regression model. In logistic regression, the Y variable has only two values, which are 0 (unbroken pipe) and 1 (broken pipe). D.R. Cox was the first author to introduce a logistic regression model in 1969, a model commonly used in natural as well as social studies. The significance of this model includes:

- Describe the relationship between the independent variable (continuous variable) and the dependent variable (probability of failure pipe F is a binary variable), not complying with the standard distribution rule.
- Priority control (historic pipe failure)
- Development of predictor model. The model does not use the least squares method to predict the parameters that use the maximum estimate method

Fig. 3. Regression logistic model.

It is possible to generalize the logistic regression model applied to the evaluation the probability of pipe failure as shown in the following diagram:

The implementation steps of the logistic regression model are summarized in Fig. 3. From the above advantages, the application of the logistic regression model to determine the probability of failure is appropriate.

4.3 Decision tree model

The decision tree model has been developed since the 80s by Breiman, Freidman, Oshem and Stone[6]. Models are capable of handling big data in a short time by classifying statistical data into classes from which the decision tree is constructed. The model will ask whether the tube is damaged or not, thereby dividing survey data into subclasses.

Decision tree (DT) algorithmic will classify and present variables that can lead to pipe failure events and assign corresponding values to the most appropriate division of data, which will be repeated until the class can no longer be broken up. A decision tree can have multiple classes and each class can have hundreds of variables. The DT model is called a non-parametric model, so when used, there are no constraints between the variables as well as for each variable and even the null data sets. Because DT can represent all discrete values, sometimes sensitivity to training is susceptible to interference, inappropriate attributes[20]. Decision tree is built with 3 steps:

- The first step identified node on the DT;
- The second step built on the DT;
- The third step select the DT.

The final result DT is the shape as Fig 4. Decision tree model is a new direction for predicting pipe failure, with advantage isolation outlier variables from system, each of this value is isolated at alone node, it is not standardized data, resolve number variables and alphabet variables, small memory is used when running this model, time of running model is shorter. It is perfect to apply this model for the data collection on water distribution systems.

Fig. 4. The result of Decision Tree Model.

Fig. 5. The result of Decision Tree Model.

Table 1. Classification.

Classification 1: 3064/62	The original data has 3064 rows (F=0) + 62 rows (F=1) = 3126 rows (F)with 3064 rows F=0 & P $\geq$ 6.053; 62 rows F=1
Classification 2: 3064/50 0/12	Node 2 has 3064 rows(F=0) + 50 rows (F=1) = 3114rows(F) with P $\geq$ 6,053&L<201.1 Node 3 has 0 row (F =0)+12 rows (F=1)=12 rows(F) with P<6.053.
Classification 3: 3027/41 37/9	Node 4 has 3027 rows (F =0)+41 rows(F=1)= 3068(F) with L<107.7& P $\geq$ 6.053 Node 5 has37 rows (F =0)+ 9 rows(F=1) =41 rows (F) with L>107.7 & P $\geq$ 12.6

Figure 4 show that node 1 get 100% pipe failure data & $P \geq 6.1$. Node 2 has 100% ($F=0$) in the original data, with condition $P \geq 6.1$ & $L < 201$ then 98% pipe failure and 2% pipe nonfailure. Node 3 ($P < 6.1$) have not pipe failure data so it has 0% ($F=0$). Classification 3 statistic data are divide in to 2 parts , 1 part into node 4 is 98% ($F=0$) and 1 part into node 5 is 1% ($F=0$). Continue for the next classification, original pipe nonfailure ($F=0$) will divide smaller into nodes, the steps are repeated to the last classification. The value % in one node show that the number of variables in pipe failure and no failure data correspond with classified conditions, as node 4 will have 99% ($F=0$) and 1% ($F=1$) with $P \geq 6.1$ & $L < 108$; node 5 classify rate is $F(0|1)$ 80% and 20% $P \geq 13$ & $L > 201$.

5 Conclusion

This paper have analyzed prior researches and proposed another direction to resolve the problem of predicting the probability of pipe failure. With the factors of corrosion, load, physical condition and historic pipe failure, the author suggest that data is collected at company where manage pipe network systems. Data collection is pipe fracture in the past, information on GIS and pipe network system drawings. Based on data collections, it totally identified variables used as inputs for two models: Regression Logistic Model and Decision Tree Model to evaluate the probability of the pipe failure on water distribution systems.

References

1. S. Mansour-Rezaei, G. Naser, R. Sadiq, Predicting the potential of contaminant intrusion in water distribution systems, 105–115 (2014)
2. A. Wood, B.J. Lence, J. Infrastruct. Syst. **15-2**, 111–119 (2009)
3. T. Wengström, *Comparative analysis of Pipe Break Rates* (Chalmers Univ. Technol., 1993)
4. R.B. Petersen, R.E. Melchers, Corros. Prev. **23**, 1–12 (2012)
5. E.S.-M. Sadiq, Y.K.R. Sadiq, Urban Water J. **27-1**, 1–12 (2008)
6. M. Helena, F. Miquel, P. Amparo, D. Vera, CFD Simulation of pathogen intrusion during hydraulic transients in water distribution networks, International Environmental Modelling and Software Society (2014)
7. V.M. Rajani, J.B. Makar, S. McDonal, C. Zhan, S. Kuraoka, C-K. Jen, *Investigation of grey cast iron water mains to develop a methodology for estimating service life* (Denver, CO, 2000)
8. R. Vera, O.T. De Rincon, A. Di Sarli, J. Alberto, R. Valenzuela, J. Tidblad, Int. J. Corros. **2012**, 85 (2012)
9. H. Rezaei, B. Ryan, I. Stoianov, Procedia Eng. **119-1**, 253–262 (2015)
10. K. Pietrucha-Urbaniak, Eng. Fail. Anal. **57**, 137–142 (2015)
11. M.V. Seica, J.A. Packer, F. Asce, Mechanical Properties and Strength of Aged Cast Iron Water Pipes, 69–77 (2004)
12. G. Demissie, S. Tesfamariam, R. Sadiq, J. Pipeline Syst. Eng. Pract. **7-2**, 4015028 (2016)
13. T.M.W.A. Pelliccia, *Water main repair/replacement for Binghamton, N.Y.* (Vicksburg, 1981)
14. M. Gorji-bandpy, M. Shateri, GMSARN Int. J. **2**, 117–124 (2008)
15. H. O. Bainbridge, K. Bainbridge, J. Perform. Constr. Facil. **25**, 259–266 (2011)
16. H. Al-barqawi, T. Zayed, Performance, 305–318 (2008)

17. D. Achim, F. Ghotb, K.J. McManus, J. Infrastruct. Syst. **13-1**, 26–30 (2007)
18. S. Yamijala, *Statisticcal estimation of water distribution system pipe break risk* (Texas A&M University, 2007)
19. G. Kabir, S. Tesfamariam, J. Loepky, R. Sadiq, ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng. **1-3**, 4015007 (2015)
20. I.H. Witten, E. Frank, M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques* (Google eBook, 2011)