

# Modelling of Non-Visible Leaks to Improve Targeted Detection

Diyang QI

Melinda R Hodkiewicz

School of Mechanical and Chemical Engineering

Nazim Kham

Gopalan Nair

School of Mathematics and Statistics

Jon Langford

Joe Standring

CEED Client: Water Corporation

## Abstract

*Leakage is a common issue in pipe networks. Water leaks in potable water reticulation have significant impacts on economic, environmental and social wellbeing of water companies. The Water Corporation manages 34,156 km of water reticulation pipe networks and experiences 33 GL/annum real loss (leakage) in Western Australia. The Asset Management Branch has run an active leak control (ALC) program for four years. Although the program remains sustainable, there is a strong business drive for improvement of the effectiveness of ALC. This project seeks to model the likelihood of non-visible leaks in reticulation networks based on four years active leak detection (ALD) records. A heat map will be generated to assist planning of ALD; as a result, the ALD can target areas with high risk of non-visible leaks and cost effectiveness can be improved. This project collected records from WC database. Water pipe and environmental covariates were extracted from GIS. The statistical package R is used for data regression analysis and model construction. The deliverables of this project is a heat map in shape file in GIS ArcMap10 and a thesis.*

## 1. Introduction

The Water Corporation (WC) Asset Management Branch (AMB) Strategy & Integration team is carrying out an Active Leak Control (ALC) program which seeks to reduce the quantum of leakage and decrease asset repair cost by active leak detection (ALD) and subsequent proactive repairs. The ALC program was commenced in 2010 and, up to now, it has prevented approximately 4.5 GL leakage. The ALD is mainly targeted at non-visible leaks. Visible leaks are able to be seen emerging from the ground or pavement. Visible leaks generally have higher leakage rate than non-visible ones. Most are reported by customers soon after they occur. Non-visible leaks percolate into the surrounding ground or enter other conveyance facilities. Non-visible leaks produce more water loss than visible leaks because they have much longer duration. The funding and resources of the ALC program are finite. There is strong business need to achieve best cost efficiency of the ALD (Langford, 2013).

The ALD currently operates under suburbs based planning, which requests the contractors to survey the entire water network in a targeted suburb. The operators use acoustic methods to conduct an initial survey and record all suspect sounds. The suspected leak is rechecked afterwards. Confirmed leaks will be pinpointed by electronic correlators. The average survey rate is 12km per man day. The leak finding rate is approximately 6 major leaks per 100 kilometre pipe. The ALC effort was applied inefficiently because there was little or even no leakage exhibited in many sub-areas. Consequently, a statistical model predicting the likelihood of non-visible leakage is necessary to target the ALD optimally.

## 2. Methodology

### 2.1 Data Collection and Process

The ALD data (mostly non-visible leaks) is collected by contractors in a number of different report formats. A summarised report of all records is in Microsoft Excel. The visible leaks and break data are saved in SAP and GIS. The leak data covers the major water mains leaks (larger than 1L/Min) in the suburbs in the Perth Metro Area from 01/07/2010 to 30/06/2014. 1302 records are included in Leak dataset. The pipe data includes all water reticulation pipes in the detected suburbs and spatially joined with all available environmental and operational information. The population of the pipe dataset is 110731.

The “fish bone diagram” summarises the potential influential factors of non-visible leaks. The factors on the upper half of the fish bone diagram are believed to affect pipe failure. The lower half factors are likely to correlate with leaks being non-visible (Langford & Standring, 2014; Yamijala et al, 2009; etc.). The highlighted factors have collectable data in WC documents or GIS geodatabase. Some factors, such as meteorological variables, have available data. However, those factors are not incorporated into this model since the ALD leak data cannot indicate the time when the leak started.

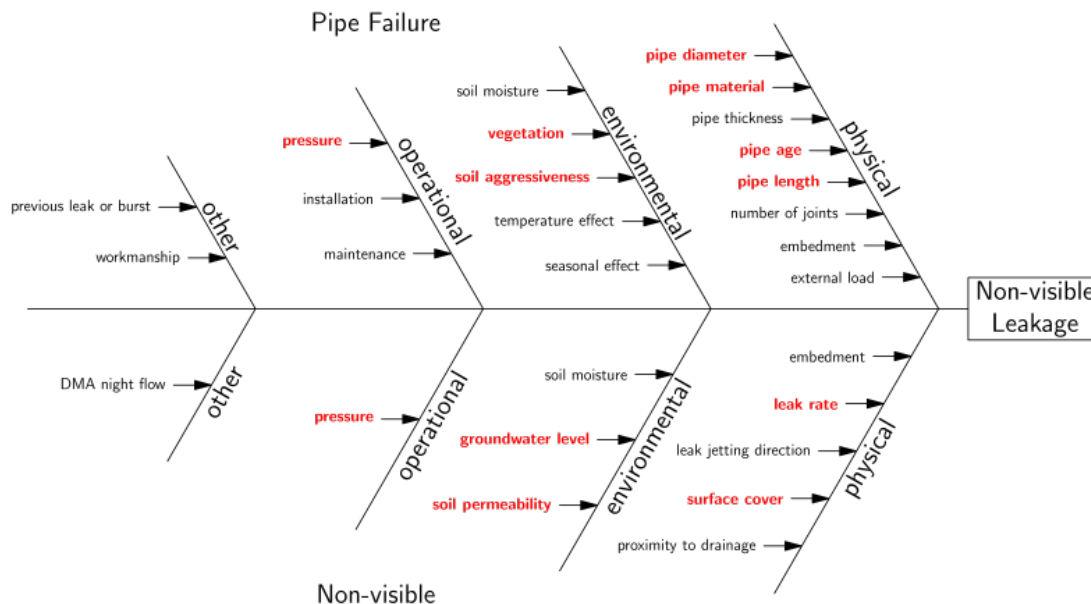


Figure 1 Potential Influential Factors Fish Bone Diagram

The leak events and influential covariates are joined to pipe data based on their spatial relationship in ArcGIS. Then the attribute table is exported and transformed to CSV format which can be read in statistical package R

## 2.2 Statistical Analysis

The statistic models used in this model are generalised linear models (GLMs) which have a general expression

$$h(\mu) = \eta = \mathbf{X}\boldsymbol{\beta} \quad (1)$$

Where  $\mu$  is a statistic of a dependent variable,  $\mathbf{X}$  is vector of covariates,  $\boldsymbol{\beta}$  is vector of regression parameters,  $\eta$  is linear predictor and  $h(\mu)$  is the link function (Dobson & Barnett 2008).

A Poisson GLM (log-linear model) is aligned with the Poisson distribution for the counts of observed values. It is used to model the mean of the dependent variable. The link function is a natural logarithm.

$$\log(\mu) = \beta_0 + \sum \beta_i x_i \quad (2)$$

Where  $\mu$  is the mean number of breaks per pipe in this case.

A logistic GLM predicts the probability of discrete outcomes from explanatory inputs that may be discrete, continuous, dichotomous and/or a combination of any of them. The distribution of observed values can be either Bernoulli, binomial, categorical or multinomial. The link function is logit function.

$$\text{logit}(p) = \beta_0 + \sum \beta_i x_i \quad (3)$$

$$\text{Or } p = \frac{\exp(\beta_0 + \sum \beta_i x_i)}{1 + \exp(\beta_0 + \sum \beta_i x_i)} \quad (4)$$

Where  $p$  is the probability of a break occurring in Leak Occurrence Model and probability of Leak is nonvisible in Nonvisible Model.

The logistic regression model is useful for assessing the risk of pipe failure, since in many cases, the occurrence of pipe failure is considered more important than the number of faults found from an asset management point of view.

## 3. Results and Discussion

The model fitting processes are conducted in statistic package R. The model selection approach is stepwise regression. The model starts with all covariates and then goes through the iterative process of eliminating covariate with highest p-value and refitting until the remaining covariates are all statistically significant at 10% significance level.

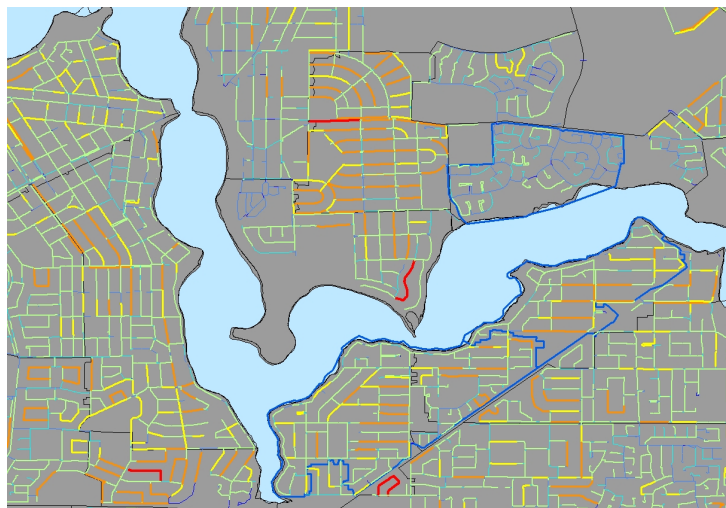
Potential Influential Covariates		Leak Occurrence Model		Nonvisible Model
		Mean Number of Breaks	Probability of Break Occuring	Probability of Leak is nonvisible
(Intercept)		-10.35	-10.04	-2.071
LENGTH		0.005165	0.00718	
Material	AC			
	ACL	1.173	1.569	-15.4
	CI	-0.7992	-0.8214	0.3204
	CU	-0.6198	-0.4001	-1.638

	DI	-2.005	-1.916	20.21
	GRP	-11.46	-9.977	NA
	GS	-0.7627	-0.8112	-15.19
	GWI	-14.29	-13.38	NA
	HDPE	-10.75	-9.711	NA
	MDPE	0.3253	1.104	0.6733
	P	0.4848	1.229	0.2119
	RC	-0.301	-0.3322	0.9619
	S	-1.621	-1.419	-15.23
	unknown	13.64	4.204	NA
	VIC	-14.16	-13.18	NA
Age		0.06797	0.06553	
Decade	1900s			
	1910s	-0.4676	-0.5305	
	1920s	-1.21	-1.324	
	1930s	0.3266	0.0742	
	1940s	1.383	1.178	
	1950s	2.009	1.736	
	1960s	2.102	1.796	
	1970s	2.904	2.441	
	1980s	3.064	2.55	
	1990s	3.731	3.294	
	2000s	-7.253	2.781	
	2010s	5.766	5.374	
Diameter	<100mm			
	100-149mm	-0.00053	-0.01261	
	150-199mm	-0.4426	-0.529	
	200-300mm	-0.777	-0.648	
	>300mm	-13.14	-12.61	
Pressure	<30m			
	30-55m		-0.2	
	60-70m		-0.01147	
	>70m		0.1952	
Soil Aggressiveness Ranking	1			
	2	0.5795	0.5285	-0.3795
	3	0.7152	0.6636	-0.9307
	4	0.8513	0.813	-0.00063
	5	-12.85	-11.97	NA
	0	-1.351	-1.977	NA
TDS_LEVEL	0 - 500mg/L			
	500 - 1000mg/L	0.07879	0.03125	
	1000 - 1500mg/L	0.3168	0.197	
	1500 - 3000mg/L	0.6542	0.5164	
	3000 - 7000mg/L	0.5858	0.3925	
Depth to Water Table		-0.01017	-0.00609	
Soil Permeability	1			
	2	0.6018	0.5863	-1.091
	3	0.181	0.009111	-15.15
	4	0.2725	0.4381	-14.52

	5	-0.2923	-0.9165	19.79
	6	0.5671	0.7598	0.08231
	7	-0.1844	-0.1229	0.7116
	0	-12.97	-12.44	NA
Land Cover	driveway			
	footpath			-0.5469
	road			-1.105
	sand			-0.08816
	tree			-0.1385
	verge			-0.09026

**Table 1 Estimated Coefficients of Significant Covariates in GLM Analysis**

- **Length:** pipe length is positively correlated to both mean number of breaks and probability of break occurring. It is common sense that the longer the pipe the greater the chance of having a break under the same condition.
- **Material:** the models show AC, RC CI and S, ranking from high risk to low risk, are the four pipe materials that have strong correlation to failure. This finding agrees with exploratory data that these materials are the top ranked for number of breaks per kilometre pipes. In addition, RC pipe is more likely to have non-visible leaks than others.
- **Size:** the larger diameter pipes will experience less breakage than the small diameter pipes. Exploratory data analysis shows the pipe size has some inter-correlations with material and installation decades.
- **Installation Decade:** pipe installed in early 20 century is not sensitive to this variable. The 1940s to 1960s groups are most likely to fail and the likelihood reduces as time approaching to 21 century. The installation decade not only reflects the age of pipe both also involves non-age factors such as the material quality, material popularity and installation technology in a particular time period.
- **Age:** it is as expected that the older pipes are more likely to fail. The model stands with wear-out effect of pipe system.
- **Soil Aggressiveness:** the soil aggressiveness ranking was developed by WC AMB in 2007. Soil aggressiveness was ranked from 1 to 5 in reference to the corrosivity level of soil to either ferrous or cement based pipes. A zero ranking is assigned to the plastic pipe. A higher ranking indicates high corrosion potential.



**Figure 2 Probability of Leak Occurrence Heat Map Zoom In**

- **TDS\_level:** total dissolved salts (measured in mg/L) also relates to pipe corrosion. The model result shows the similar outcome as soil aggressiveness.
- **Soil Permeability:** the hydraulic conductivity of the top soil level is graded from 1 to 7 from slow to rapid and 0 for surface water. Non-visible leaks are likely to occur in more permeable soil.

Figure 2 is an example output of this modelling project. The probability of break occurring for each reticulation pipe is predicted by the logistic leak occurrence model. Seven classes with different colours are used to represent the probabilities. This heat map can help the targeting of ALD zone. In un-monitored, the heat map can help to target high risk areas instead of the entire suburb.

## 4. Conclusions and Future Work

This project has achieved its proposed objectives. A better understanding of the Perth water reticulation system and the influential factors of pipe failure is gained from exploratory data analysis. A new geodatabase of ALD data is built and predictive models are successfully developed. The outcomes of this project have been positive. WC could obtain economic benefits from the improvement of ALD planning and smarter renewal decisions with assistance of the models. To validate the models and improve the workability of the predictive models, the following activities are recommended:

1. Review models annually by updating the break data, pipe data and inspection data.
2. Use Urban Monitor data to classify land cover for Non-visible model mapping.
3. Restructure the pipe data into small length segments for visualisation of Non-visible model.

## 5. Acknowledgements

The author would like to express his appreciation for the support from supervisors, Melinda Hodkiewicz, Nazim Khan and Gopalan Nair and Client Mentors, Jon Langford and Joe Standring. Additional thanks to David Low and Mike Canci for their assistance in GIS applications and information sharing.

## 6. References

- Dobson, AJ & Barnett, AG 2008, An introduction to generalised linear models, 3rd Edition, Chapman and Hall, Boca Raton.
- Langford, J 2013, Operating Implementation Business Case (OIBC) Part B-Long Form.
- Langford, J & Standring, J 2014, Water Loss Management -- Overall Strategy, Water Corporation.
- Yamijala, S, Guikema, SD & Brumbelow, K 2009, 'Statistical models for the analysis of water distribution system pipe break data', Reliability Engineering & System Safety, vol. 94, no. 2, pp. 282-293.