Spatial Modelling of Pipeline Failure Rates including a Co-Variate for the Impact of Water Pipe Failures on Adjoining Pipes

Ashleigh Ballantyne

Michael Small[1], Melinda Hodkiewicz[2]
1. School of Mathematics and Statistics
2. School of Mechanical and Chemical Engineering
The University of Western Australia

David Burton
CEED Client: Water Corporation

Abstract

The current approach to reticulation main replacement in Perth does not use the failures of specific pipes to assess the likelihood of nearby pipes failing, and considers any one pipe failure independent of any other. Past investigations have only considered the physical and temporal characteristics of the pipes and their environments, without considering the effect of a failure on its operational environment, even if the failures occurred on the same, or connected pipes. This project investigates whether there is a relationship between the failure of a pipe, and the occurrence of nearby subsequent failures within a given time frame. This is done through calculating the distance to the closest subsequent failure for every pipe failure across differing cohorts of the pipe population, and determining how likely it would be to observe such results if the failures were randomly spatially located. By incorporating this information into their risk analysis methodology, the Water Corporation hopes to improve their decision-making around the prioritisation of areas of the pipe network for renewal, and to achieve the best risk mitigation in the most cost-effective manner. This analysis identified several pipe subgroups with identifiably strong spatial relationships between their constituent failures.

1. Introduction

The current approach to reticulation main replacement in Perth is to replace only short sections of pipe at a time, using a risk-based model to target those pipes that are assessed to be most likely to fail. Historically, pipe replacement has only been conducted pre-emptively on Water Corporation assets where the consequences of failure were unacceptable. For example, if a trunk water main beneath a railway line or main road were to fail, it would not only cause major public inconvenience, but would also constitute a severe public safety hazard. By contrast, reticulation pipe, failures of which have a minimal impact upon customers, remain in service until they are no longer viable.

The Water Corporation supplies water to population centres across most of the state of Western Australia, the largest of which is the Perth Metropolitan Area. The number of in-service pipes in the area is increasing with the population, and with most pipes aged at over

50 years old, the total number of yearly pipe failures is expected to increase. Pipes installed especially in the decades of 1940-1960 are beginning to fail more frequently than their peers. Statistical models for determining a pipe's risk of failure are currently used only for distribution main pipes. The decision to replace reticulation pipe is still based on the number of bursts per year; in the Perth Metropolitan Area, pipes that burst three or more times per year are considered for replacement (McEwan, 2017).

This approach to pipe replacement is becoming less economically viable as the cost of repairing and/or replacing failed pipes is increasing alongside the growing number of active pipes. As such, the Water Corporation is moving towards a focus on pre-emptive replacement, based on each pipe's statistical risk of failure over a given time horizon. Aiming to repair, refurbish, or replace at-risk pipes before they fail or cause damage at surface level is a far more efficient use of funds in terms of managing risk and reducing water loss. An analysis on cost savings based on a sample of data from the Water Corporation's leak detection and repair program found that the average cost of repairing a pipe proactively is almost half that of repairing a pipe reactively (McEwan, 2017).

A great deal of work has been conducted into building statistical models of pipe failure in water distribution utilities by previous CEED Scholars. This investigation builds directly from the previous work of CEED Scholar Diyang Qi's Modelling Non-Visible Leaks to Improve Targeted Detection, and takes reference from the results of other studies by past CEED scholars. The results of their models indicated that:

- Longer and older pipes have a greater chance of leakage (Wei, 2012; Qi, 2015).
- Pipe failure is material sensitive; Concrete/Cement based pipe has the highest failure rate, followed by Metal pipe. Plastic pipe has the lowest failure rate (Qi, 2015).
- Failures in cast iron pipes show a strong positive correlation with differential soil movement (Pratt, 2011) and heightened operational pressure (Wei, 2012).
- Failures in reinforced concrete pipes are unaffected by soil corrosivity, but express a positive correlation with differential soil movement (Pratt, 2011).
- Smaller diameter pipes have a higher chance of leakage than larger diameter pipes.
- A higher soil aggressiveness rating and total dissolved salt level increases the probability of leakage (Pratt, 2011; Pan, 2012; Wei, 2012; Qi, 2015).
- Leakage from reinforced concrete pipe and pipe installed under a natural soil surface is most likely to be nonvisible (i.e. not observable at surface level) (Qi, 2015).

The Water Corporation suspects that when one water pipe section fails, its failure may have been affected by the conditions of and around the surrounding pipe network. It is reasonable to assume that pipes within the same area would share many characteristics. Factors such as the soil the pipes were placed in, the weather and traffic conditions they are subjected to, and the time at which they were installed are all likely to be very similar between two adjacent pipes, and functionally identical on different parts of the same pipe. As such, it is also reasonable to assume that failures within close proximity in time and distance may be related.

This analysis seeks to determine how pipe failures are related in space and time. The main hypothesis to be tested asks whether, discounting all other variables, failures of water distribution pipes in the Perth Metropolitan Area occurring closer together than can be accounted for by random chance. If a relationship can be found between the incidence of a pipe failure and other nearby pipe failures, then renewal efforts can be directed towards both pipes with a high failure history, and towards any nearby pipe sections that they may affect.

2. Process

An analytical model must be developed in order to determine whether there is a spatial relationship between failures. This analysis focuses on building a statistical procedure that can help determine the relationship between spatially correlated failures, based on features such as pipe age, material, and environment. Specifically for this analysis, we need to determine whether failures are likely to happen close together in space and time, and whether one failure is likely to affect the chances of a nearby failure occurring within a given period. This can be done through investigating how the average distance between one pipe failure and the nearest subsequent failure changes with differing attributes such as location, material, and age.

2.1 Data Preparation

For this analysis Water Pipe, Leaks and Bursts, and Locality Boundary records from the Water Corporation's SAP and GIS corporate databases were combined to build the Pipe and Failure data frames for analysis with the R statistical package. Pipe records were removed from the data if they were not listed as active (i.e. listed as "Dead", "Not in Use", "Proposed", or "Removed"), or if the recorded diameter was invalid (i.e. ≤ 0 mm). Failure records were removed from the data if they occurred on a removed pipe, or if the installation date of the pipe was recorded as after the failure date (negative age at occurrence of failure). These account for 3400 of 249,972 pipe records (1.4%) and 241 of 12430 failure records (1.9%) being removed from the data, leaving 246,572 pipe records and 12,189 failure records.

2.2 Distance Analysis

To observe the spatial relationship between failures, two averages must be calculated. The distance from each pipe failure to the closest failure that occurred afterwards (closest subsequent failure, CSF) is recorded and averaged. The same mean CSF value is then calculated when the date values for every work order are randomised. This randomised mean is recalculated 199 times, each time with the date values randomised again, for a total of 200 trials. This gives us the *baseline* or true mean, and a sample of *shuffle* means. The difference between the baseline and the distribution of the shuffle means tells us how different reality is from a random set of possible failures. That is, how much closer together are we observing subsequent failures occur, than if there were no spatial relationship between failures? If the baseline is significantly lower than the spread of shuffle means, then failures are occurring closer together in reality than they would do if the failures were randomly spatially located.

Because pipe failures must occur on pipes, and these pipes are not uniformly distributed across space, randomising the failure date vector is the best way to achieve a sample of independent, random failures. By keeping all other factors the same, we are considering only the effect of a failure's spatial location on all other failures. The tests are thus standardised for factors such as population and pipe density. This random sampling must be conducted for a reasonably large number of repetitions, in order to garner a representative, unbiased average.

By placing limits on the search for the CSF, different questions can be answered. If we wish to look only at propagating failures along connected pipes, we can set a 50m distance boundary, such that the analysis only finds the CSF within a 50m radius of the origin, or returns null if no match was found. If we wish to investigate failures close together in time, we can set a 28 day time boundary, such that the CSF only includes failures up to 28 days after the occurrence of the origin failure. A distance boundary of 50m, and time boundaries of 7, 28, 50, 100, 250, and 500 days have been used to investigate the relationship between failures that have occurred very close by to each other over time.

By taking the standard difference between the distribution of shuffle means and the baseline mean, we can determine the strength of the observed relationship between failure locations. The standard difference is calculated as a z-score measured in the number of standard deviations from the average of the shuffle means, calculated as $z = (b - \mu)/\sigma$, where b is the baseline mean, μ is the average of the shuffle means, and σ is the standard deviation of the shuffle means. A significance level of 0.05 is used, which approximates 2 standard deviations. If a baseline mean is more than 2 standard deviations from the average of the shuffle means, then there is evidence that the spatial location of failures is not independent. A final ranking is assigned given the derived z-scores. If a pipe subgroup has an average z-score greater than 2 over all time boundaries, it is identified as “Needs Attention”. Z-score averages between 1 and 2 are listed as “Normal”, between 0 and 1 as “Good”, and less than 0 as “Very Good”.

3. Results and Discussion

The initial data exploration resulted in some interesting observations regarding the density and spread of pipe failures in the Perth Metropolitan Area. A band of suburbs in the Stirling area was identified, with conspicuously high failure counts. A similar band was found leading through major population centres down towards the Armadale area. By ranking suburbs based on failure counts, we can identify areas where the spatial density of failures is conspicuously high. Figure 1a highlights the top 25 suburbs by failure count, with the top 5 suburbs in a darker shade. Figure 1b identifies the regions (with boundaries as defined by the Australian Bureau of Statistics (ABS) 2016 SA3 borders) with the highest failure counts, with darker shades highlighting the higher failure count regions, and paler shades for regions with lower failure counts. The Stirling region (Figure 1b, highlighted in the darkest shade) in particular recorded 2003 leaks and bursts between January 2010 and January 2017, more than double that of the next SA3 region of Joondalup, with 892 failures.

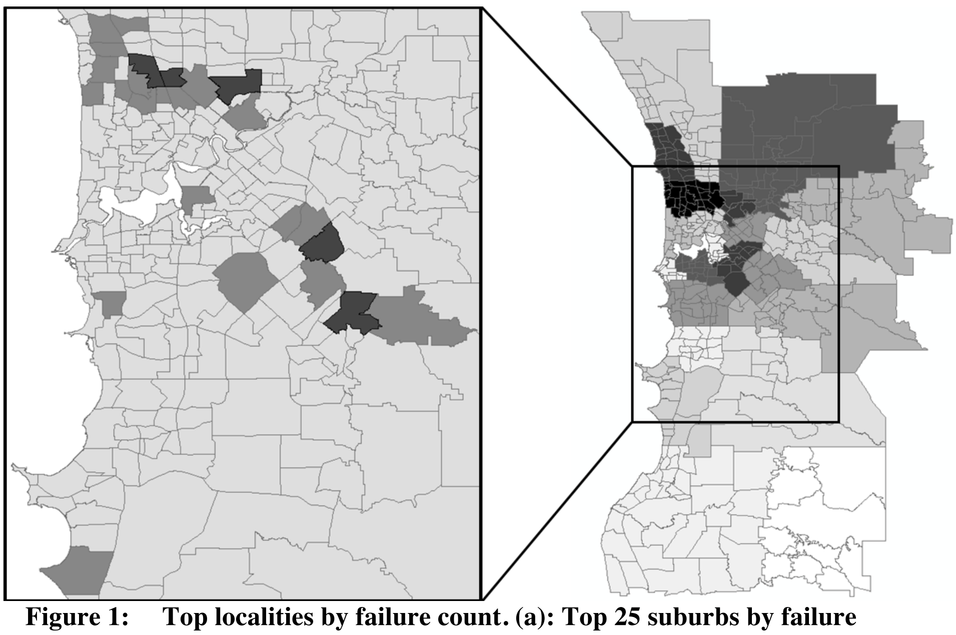

Figure 1: Top localities by failure count. (a): Top 25 suburbs by failure count, with the top 5 in a darker shade. (b): Top SA3 regions by failure count, with darker shades indicating more failures.

Given these clear wide-area effects, a natural step for the Distance Analysis was to investigate how failures were spatially related in each SA3 region. Figure 2 displays the results of a Distance Analysis by SA3 region, where the y-axis values indicate the strength of the spatial relationship, and the x-axis indicates the period after the primary failure from which we search for subsequent failures. A higher number indicates that failures are occurring significantly closer together than they would do if they were independent, and not spatially related. The regions are ranked based on their average level of significance over the tests. The Stirling and South Perth regions show the strongest correlation between nearby failures, while Canning, Rockingham, and Wanneroo show a negative correlation, indicating that failures in these areas have a tendency to occur further apart than expected. However, these results are insignificant, and are thus indicative of a weak trend, rather than a true spatial relationship.

These results lead us to ask, what makes these regions different? Why do failures occur closer together in some regions, and farther apart in others? It is not a factor of the number of failures; South Perth recorded 278 failures over the 7-year period, the 4th lowest region by failure count. Neither is it a factor of the number of pipes in operation; Wanneroo has the highest number of in-service pipe sections, at 26,697, considerably above that of Stirling (20,154) and Joondalup (18,538), which have the next highest numbers of active sections.

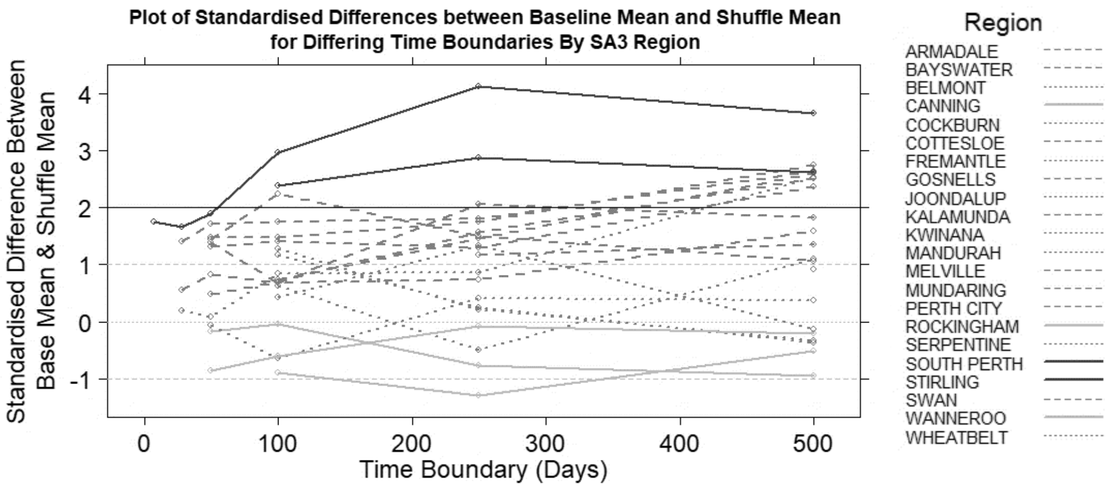

Figure 2: The strength of the spatial correlation between failures in each SA3 region. A higher number indicates failures are occurring closer together than random, while a value of 0 indicates no difference. Values greater than 2 are significant.

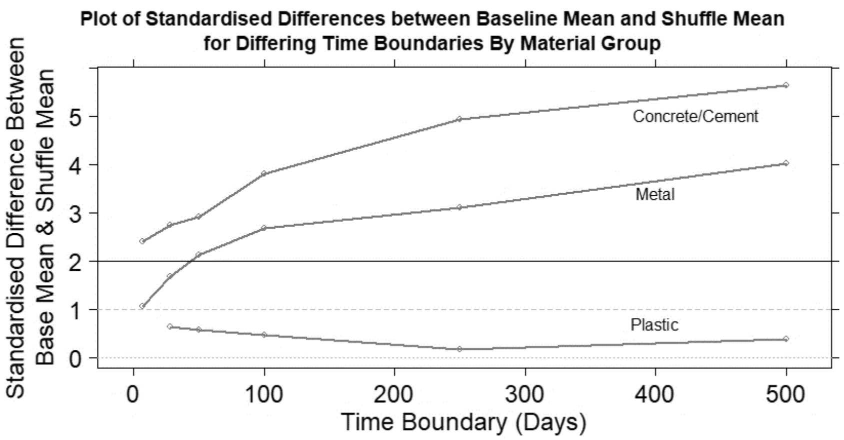

Figure 3: The strength of spatial correlation between failures in each material group subset.

By comparing the proportions of material types across these regions, we may start to define why these areas are different; Canning, Rockingham, and Wanneroo have all had a large number of plastic pipe (e.g. PVC, polyethylene, etc.) installations in the past three decades, resulting in substantial proportions of their pipe populations being plastic-based. Stirling and South Perth have also had plastic pipe installations, but in far fewer numbers; in fact, they have had fewer pipe installations overall in the past three decades. A Distance Analysis by material group (Figure 3) shows a significant difference between the behaviours of failures occurring on pipes of different material types. The evidence indicates that failures on plastic pipes are independent, while Metal and Concrete/Cement based pipe failures are not.

4. Conclusions and Future Work

Pipe failures have been found to be positively spatially correlated when separated by ABS SA3 region. The SA3 regions of Stirling and South Perth show significant evidence for short-range failure propagation, as do the subgroups of Metal and Concrete/Cement based pipes when the pipes are separated by Material Group. When a plastic pipe fails, overall, the data suggests that another failure within 50m is no more likely than random occurrence. Certain subgroups indicate that failures are occurring further apart than they would if the failures were independently spatially located, however none of these results are significant, and serve only as a weak indication of behaviour rather than a confirmation of a relationship.

Further work remains to be done in developing the Ranking process, and determining final recommendations for renewal prioritisation based on the behaviours of each of the pipe subgroups. Given the results, the project recommendations are for the regions of Stirling, South Perth, and Armadale to have a more in-depth physical or statistical investigation, and in general that Concrete, Cement, and Metal pipes be considered more closely than Plastic pipes. Validation testing and optimisation using January-June 2017 failure data will be carried out, to ensure that the final rankings can be used to help predict future failures. Future work that may arise from this project could be an application of the methodology to sewer pipe failures, or other events that occur in networks.

5. Acknowledgements

This project was made possible by David Burton, Ian Gibb, and Ashley Price at the Water Corporation, and by Professors Michael Small and Melinda Hodkiewicz at the University of Western Australia. The author would like to thank all of those involved, with an additional thanks to the AIPM team at the John Tonkin Water Centre, who were always willing to help.

6. References

McEwan, D. (2017) Re: Information for Project & Project Brief. [E-mail]. Sent 13 January.

Pan, J. (2012) Validation of Cast Iron Pipe Modelling, *CEED Seminar Proceedings 2012*, pp. 67-72.

Pratt, C. (2011) Factors Influencing Pipe Failures in the WA Environment, *CEED Seminar Proceedings 2011*, pp. 43-48.

Qi, D. (2015) Modelling Non-Visible Leaks to Improve Targeted Detection, *Bachelor's Thesis*, The University of Western Australia.

Wei, R. (2012) Replacement Strategy for Cast Iron Pipes in Western Australia, *CEED Seminar Proceedings 2012*, pp. 103-108.