Predicting Water Pipe Failures

Li Chik
Department of Civil Engineering
Monash University

Supervised by: Prof Jayantha Kodikara
Dr David Albrecht
Outline

1. Introduction
2. Data Analysis
3. Methodology
4. Results
5. Conclusion
Consequence of Water Pipe Failures

- Disruption to the water supply
- Disruptions to road network and/or other infrastructure
- Flooding
- Reputation and confidence
What is the problem?

- Optimisation problem:
  - Replace too early → service life of the asset is not fully utilised
  - Replace too late → pipe failure
Aim

Estimate the probability of pipe failures to reduce the number of unexpected disruptions in the water distribution network.
Data

Cast Iron Pipe Properties:
- Number of assets: 8404
- Total pipe length (km): 544
- Number of failures: 2476
- Construction period: 1860 to 1929
- Observation period: 1994 to 2012

Map showing locations of Melbourne and Rotorua in Australia.
Pipe Failure Rate

Pipe Failure rate vs Construction Year

Pipe Failure rate vs Pipe Diameter (mm)

Pipe Failure Rate vs Square Root Length

Pipe Failure Rate vs Pressure

MONASH University
Pipeline Asset and Risk Management System (PARMS)

- A statically model current used by the utilities
- Poisson generalised linear model:
  - Generalisation of the linear regression model
  - Poisson distributed response variable

\[
P(X = n) = \frac{(\lambda)^n e^{-\lambda}}{n!}
\]

\[E(x) = \lambda \quad Var(x) = \lambda\]
Pipeline Asset and Risk Management System (PARMS)

For $i = 1: \text{Num Construction Year}$

For $j = 1: \text{Num Failure Year}$

Log link function

$\ln(\mu_{ij})$

$\ln(\text{Length}_{ij})$ (offset term)

$\alpha_0$

$\alpha_1i$

$\alpha_2j$

$\beta$

$\text{Construciton Year}_i$

$\text{Age}_{ij}$

$\text{Failure Year}_j$

$Z_{ij} \sim \text{Poi}(\mu_{ij})$

For $j = 1: \text{Num Failure Year}$

For $i = 1: \text{Num Construction Year}$
Pipeline Asset and Risk Management System (PARMS)

\[ \ln(L_{\text{length}_i}) \]

\[ \alpha_0 \]

\[ \alpha_1 \]

\[ \ln(\lambda_i) \]

\[ \alpha_0 \]

\[ \text{Age}_i \]

\[ \beta \]

\[ Z_i \sim \text{Poi}(\lambda_i) \]

\[ Z_i \]

For \( i=1:\text{Num Pipes} \)
Beta Distribution

\[ \text{Beta}(cq, c(1 - q)) = \frac{\Gamma(cq + c(1 - q))}{\Gamma(cq)\Gamma(c(1 - q))} \left( x^{cq-1} (1 - x)^{c(1-q)-1} \right) \]

**Expected Value** = \( q \)

**Variance** = \( \frac{q(1-q)}{c+1} \)

\( 0 \leq x \leq 1 \)
Hierarchical Beta Process (HBP)

Cast Iron (material group)
$q_1 \sim Beta(c_0q_0, c_0(1 - q_0))$

Pipe i (probability of failure)
$\pi_{k,i} \sim Beta(c_1q_1, c_1(1 - q_1))$

Year j
$z_{k,i,j} \sim Ber(\pi_{1,i})$

What we observed from the data

Estimated from a similar dataset,
$q_0=$average number of failures per pipe,
$c_0=$Total number of observation for all pipe

$c_0 = 471575$
$q_0 = 0.0061$
$c_1 = 1$

What we observed from the data
WinBUGS Model

- Trained data from 1994 to 2012
- Bayesian network model in WinBUGS
Model Fitting

Estimated Failure vs Observed Failure

**PARMS**
- Large variation
- Limited covariates used
- Between pipe variation

**HBP**
- Good fit
- Non-parametric model
- No covariates used
# Expected Failure Probability

<table>
<thead>
<tr>
<th></th>
<th>Expected Failure Probability</th>
<th>0-0.05</th>
<th>0.05-0.1</th>
<th>0.1-0.15</th>
<th>0.15-0.2</th>
<th>0.2-0.25</th>
<th>0.25-0.3</th>
<th>0.3-0.35</th>
<th>0.35-0.4</th>
<th>0.4-0.45</th>
<th>&gt;0.45</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PARMS</strong></td>
<td>Number of Pipes</td>
<td>7236</td>
<td>594</td>
<td>308</td>
<td>109</td>
<td>55</td>
<td>31</td>
<td>30</td>
<td>16</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Number of Failures (2013)</td>
<td>79</td>
<td>25</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Failure Percentage</td>
<td>1.1%</td>
<td>4.2%</td>
<td>5.2%</td>
<td>7.3%</td>
<td>3.6%</td>
<td>9.7%</td>
<td>13.3%</td>
<td>25.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>HBP</strong></td>
<td>(Pipe Material)</td>
<td>7144</td>
<td>845</td>
<td>247</td>
<td>109</td>
<td>41</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Number of Pipes</td>
<td>7144</td>
<td>845</td>
<td>247</td>
<td>109</td>
<td>41</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Number of Failures (2013)</td>
<td>62</td>
<td>33</td>
<td>16</td>
<td>15</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Failure Percentage</td>
<td>0.9%</td>
<td>3.9%</td>
<td>6.5%</td>
<td>13.8%</td>
<td>19.5%</td>
<td>20.0%</td>
<td>25.0%</td>
<td>75.0%</td>
<td>50.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
Model Predictions to 2013

- **PARMS**
  - Cannot differentiate which pipes are more likely to fail
  - No failures detected for about first 1.5% length
  - 15% failures detection at 10% length

- **HBP**
  - Better failure detection compare to PARMS
  - 30% failure detection at 10% length
  - Still large room for improvement
Conclusion

- The fitting ability of HBP is very high compared to PARMS.
- The HBP is able to rank the pipe with better accuracy compared to the PARMS model.
- A large difference is still present between the actual ground truth and the prediction made by HBP. Therefore, it might be possible to try and improve the model to close the difference.
- HBP lacks time component.
Acknowledgement

- Smart Water Fund