Water pipe condition assessment: a hierarchical beta process approach for sparse incident data

Zhidong Li · Bang Zhang · Yang Wang · Fang Chen ·
Ronnie Taib · Vicky Whiffin · Yi Wang

Received: 17 November 2012 / Accepted: 20 May 2013 / Published online: 29 June 2013
© The Author(s) 2013

Abstract Prediction of water pipe condition through statistical modelling is an important element for the risk management strategy of water distribution systems. In this work a hierarchical nonparametric model has been used to enhance the performance of pipe condition assessment. The main aims of this work are three-fold: (1) For sparse incident data, develop an efficient approximate inference algorithm based on hierarchical beta process. (2) Apply the hierarchical beta process based method to water pipe condition assessment. (3) Interpret the outcomes in financial terms usable by the water utilities. The experimental results show superior performance of the proposed method compared to current best practice methods, leading to substantial savings on reactive repairs and maintenance, as well as improved prioritization for capital expenditure.

Editors: Kiri Wagstaff and Cynthia Rudin.

Z. Li · B. Zhang · Y. Wang (✉) · F. Chen · R. Taib · Y. Wang
National ICT Australia, Eveleigh, Sydney, Australia
e-mail: Yang.Wang@nicta.com.au

Z. Li
e-mail: Zhidong.Li@nicta.com.au

B. Zhang
e-mail: Bang.Zhang@nicta.com.au

F. Chen
e-mail: Fang.Chen@nicta.com.au

R. Taib
e-mail: Ronnie.Taib@nicta.com.au

Y. Wang
e-mail: Yi.Wang@nicta.com.au

Z. Li · B. Zhang · F. Chen · Y. Wang
University of New South Wales, Kensington, Sydney, Australia

V. Whiffin
Sydney Water, Parramatta, Sydney, Australia
e-mail: Vicky.Whiffin@sydneywater.com.au

Keywords Hierarchical beta process · Bayesian nonparametric · Water main · Condition assessment · Sparse data

1 Introduction

Water supply networks constitute one of the most crucial and valuable urban assets. The combination of growing populations and aging pipe networks is putting the onus on water utilities to develop advanced risk management strategies in order to maintain their distribution systems in a financially viable way. Failures of reticulation water mains (small diameter pipes) usually bear low consequences hence renewal (i.e. replacement) decisions can be made on the basis of failure history. However, failures of critical water mains (diameter above 300 mm) typically bring severe consequences due to service interruptions and negative economic and social impacts, such as flooding and traffic disruption, as illustrated in Fig. 1. The financial and social costs of reactive repairs in such scenarios amount to more than one billion dollars annually in Australia alone, so utilities are strongly focusing on preventative repairs and renewal which carry much lower costs.

Water utilities develop risk based approaches in order (1) to avoid critical main failures by prioritizing renewal of specific assets; but also (2) to avoid replacing any pipe that is still in working condition. The service life or condition of a pipe can vary dramatically depending on localized conditions and various factors including the pipe material, age, construction methods, and soil conditions. So, the maintenance process starts by prioritizing high risk pipes, then proceeding to physically inspect these assets to confirm their actual condition, before finally deciding on whether to renew. The risk management process intends to balance the escalating costs associated with these steps, roughly increasing by one order of magnitude from shortlisting, to inspection, to preventative renewal, to reactive renewal when a pipe was not identified correctly. Clearly, the prioritization step can bring significant savings to water utilities.

The prioritization step requires a good understanding of the failure risk, and in particular a good estimate of the likelihood of failure, from which potential failure costs can usually be derived easily, based on the area serviced. The mechanisms of water pipe failure have been studied for decades, and various physical and mechanical models, involving pipe wall thickness (Ferguson et al. 1996), material deterioration according to environmental conditions and quality of manufacturing (Rajani and Kleiner 2001), and hydraulic characteristics (Misiunas 2005), have been developed to estimate the remaining pipe life. However, non-intrusive technologies for pipe condition assessment are still very limited and not cost effective because input parameters such as pipe wall thickness may rely on an inspection step. That latter step, called condition assessment, often requires excavation, the use of specialized analysis equipment, and public disruption.

Fig. 1 Scenes of critical water main failure captured on site

On the other hand, statistically based methods avoid these issues for water main failure prediction, by using historical infrastructure failure data. The probability of a pipe failure itself is a random variable dependent on a set of pipe physical attributes (e.g. age) as well as related environmental conditions (e.g. soil type). Various parametric or semi-parametric models such as the Cox model, Markov model, and Weibull model (Ibrahim et al. 2001) have been developed for water pipe failure analysis. However, parametric models are usually limited by their fixed model structure based on a priori assumptions on the data behaviour and their inability to adaptively adjust the model to the complexity of the problem.

To address these limitations, we propose the use of Bayesian nonparametric learning to predict water pipe condition. Historical water pipe data can be incorporated and the model can grow to accommodate future data as necessary. This novel modelling approach for pipe condition prediction has the potential to work effectively across many different pipe types and local conditions worldwide. Compared to traditional statistical modelling, nonparametric modelling aims to avoid assumptions on the structure of the model at the onset. Nonparametric learning has been applied successfully in various industries, for instance, to predict remission times for leukemia patients, time between explosions in coal mines and weather forecasts (Ibrahim et al. 2001). While the general framework of nonparametric learning can be found in the literature of survival analysis and topic modelling (Chen et al. 2012), to our knowledge the flexibility of using Bayesian nonparametric methods for pipe condition prediction has not been investigated.

Our work particularly investigated the hierarchical beta process (HBP) for the prioritization step above. The method can be used to predict the failure rate of each individual pipe more accurately by capturing specific failure patterns of different water-pipe groups. Experimental results show that nonparametric modelling outperforms previous parametric modelling for pipe condition assessment. The main aims of this work are: (1) For sparse incident data, develop an efficient approximate inference algorithm based on hierarchical beta process. (2) Apply the hierarchical beta process based method to water pipe condition assessment. (3) Interpret the outcomes in financial terms usable by the water utilities.

The paper is organized as follows. Section 2 reviews related work. Section 3 details the collected data. Section 4 presents our method with formulation and parameter inference. Experimental results are demonstrated and discussed in Sect. 5. Section 6 presents real-life application of our method, followed by experts' commentaries in Sect. 7. Section 8 describes lessons learnt from this work. The conclusion is drawn in Sect. 9.

2 Related work

Various parametric or semi-parametric models have been developed for water pipe failure analysis in the mechanical, civil, structural, and environmental engineering fields. Among the early work, time-dependent models were proposed to predict pipe breakage (Shamir and Howard 1979; Kettler and Goulter 1985). Marks et al. (1987) and Andreou et al. (1987) have proposed to use Cox's proportional hazards (Cox 1972) to model the probability of failure of water mains within a certain phase. Other statistical models, such as the entropy based methods (Awumah et al. 1990), Markov model based methods (Kleiner and Rajani 2001; Micevski et al. 2002), fuzzy set theory based method (Tchórzewska-Cieślak 2011), hierarchical fuzzy expert system (Fares and Zayed 2010), Weibull model and other variants (Ibrahim et al. 2001), have also been considered. Recently Tian et al. (2011) calculated the expectation of failure probability for each pipe with Cox model. For real world applications, Pelletier et al. (2003) applied survival analysis to three municipalities using the annual number of water pipe breaks. Rogers and Grigg (2009) developed prediction models according

to industry needs by investigating two actual cases with utility professionals. Rogers (2011) also applied their models to the water distribution system in Denver, USA. These results significantly improved condition assessment, however, water utilities need to further refine prediction accuracy, as a single false negative, resulting in an unpredicted critical failure, can cost from hundreds of thousands to a few million dollars.

Nonparametric methods have now become quite popular and well accepted in practice as they offer a more flexible modelling strategy that requires fewer assumptions about model structure. Of particular interest is the beta process (BP) which provides a Bayesian nonparametric prior for statistical models that involves binary prediction results. The prior could be in a very different shape to produce different model structures. For example, by tuning the parameters in the beta distribution, we can either generate the non-informative priors or power-law behaviors. A Bernoulli process parameterized by a value drawn from a beta process provides a binary indicator over the latent space. The beta process (or beta-Bernoulli process) was first studied for survival analysis by Hjort (1990). Then Thibaux and Jordan (2007) proposed the realization of the beta process itself, and connected it to the Indian Buffet process (Griffiths and Ghahramani 2011). The beta process has found wide applications in the context of dictionary learning (Zhou et al. 2009), image interpolation (Zhou et al. 2011), visual tracking (Wang et al. 2013), and factor analysis (Paisley and Carin 2009). Moreover, since there exist conjugate priors for the beta-Bernoulli process, the conditional posterior distribution is analytical, which enables a simple posterior inference combined with Gibbs Sampling. Analogous to the hierarchical Dirichlet process (Teh et al. 2006), the hierarchical beta process (Thibaux and Jordan 2007) can be adopted to model the grouped data. The hierarchical structure reflects the dependency amongst the different groups. A dependent hierarchical beta process is further developed by Zhou et al. (2011) considering covariate dependence.

From the modelling point of view, the hierarchical beta process was proposed by Thibaux and Jordan (2007), which is the most related work to this paper. Comparing with the work of Thibaux and Jordan (2007), this work extends the hierarchical beta process by developing an efficient inference algorithm for sparse incident data, and it also considers both mean and concentration parameters rather than using fixed concentration during the inference process.

There are a few recent works in machine learning that deal with failure predictions, such as the Kalman filter based preventive maintenance in reliable systems (Yang 2003), the multiple instance learning based computer system failure predication (Murray et al. 2006), and the proportional hazard model based clinical trial analysis (Klein and Goel 1992). Similarly, our method can be used to predict water pipe failures with the potential to be extended to other areas.

3 Data collection

We accessed the data from two different geographical regions, A and B from the greater Sydney area. Here region A is a highly urbanised commercial area, and region B is a lower density suburban area. For each region, two datasets are available: one contains the attributes of all water pipes in the region, and the other gathers failure records. The pipe attributes include identification number, laid date, length, material, diameter size, location, protective coating, surrounding soil type, etc. Failure records span from 1999 to 2010 and include report date, type of failure, failure location, number of customers affected and repair costs. In these regions the oldest pipes were laid before 1900, with an average age of 68 years for region A and 40 years for region B. The 12 year observation period was relatively short

compared to the life cycle of water pipes, so that most (about 99 %) pipes do not fail or fail just once during the observation period, making the data sparse. The soil types were similar for all the water mains belonging to the same region.

The data covers 14,765 water pipes and 1,696 failure records for region A; respectively 17,877 and 6,672 for region B. As mentioned, we only focused on the condition assessment of critical water mains (diameter above 300 mm). Most of them are manufactured in cast iron and are cement lined (CICL). There are 1315 critical water mains in region A, and 805 in region B. For the critical water mains, there are 69 failures in region A and 56 in region B.

During the analysis, we found two factors that may significantly influence the results of pipe failure prediction. The first factor was that the failure rate of pipes does not always increase with age. Besides the known features in the dataset and public resources, many other hidden factors impact the condition of water mains, such as manufacture condition. For example, the records show that some old pipes failed less than younger pipes. The second factor was the sparseness of critical water mains breaks already noted above. To enhance the performance of pipe failure prediction, these two factors need to be carefully considered when forming our Bayesian nonparametric method based solution.

4 Method

4.1 Hierarchical beta process for pipe condition assessment

A beta process, $B \sim \text{BP}(c, B_0)$, is a positive random measure on a space Ω , where c , the concentration function, is a positive function over Ω , and B_0 , the base measure, is a fixed measure on Ω . If B_0 is discrete, $B_0 = \sum_k q_k \delta_{w_k}$, then B has atoms at same locations $B = \sum_k p_k \delta_{w_k}$, where $p_k \sim \text{Beta}(c(w_k)q_k, c(w_k)(1 - q_k))$, and each $q_k \in [0, 1]$. An observation data X could be modelled by a Bernoulli process with the measure B , $X \sim \text{BeP}(B)$, where $X = \sum_k z_k \delta_{w_k}$, and each z_k is a Bernoulli variable, $z_k \sim \text{Ber}(p_k)$. Furthermore, when there exists a set of categories, and each data belongs to one of them, hierarchical beta process could be used to model the data. Within each category, the atoms and the associated atom usage are modelled by a beta process. Meanwhile a beta process prior is shared by all the categories. More details could be found in Thibaux and Jordan (2007). For a water distribution system, denote $\pi_{k,i}$ as the probability of failure for a pipe in the k -th group. Consider hierarchical construction for pipe condition assessment,

$$\begin{aligned} q_k &\sim \text{Beta}(c_0 q_0, c_0(1 - q_0)), \quad \text{where } k = 1, 2, \dots, K, \\ \pi_{k,i} &\sim \text{Beta}(c_k q_k, c_k(1 - q_k)), \quad \text{where } i = 1, \dots, n_k, \\ z_{k,i,j} &\sim \text{Ber}(\pi_{k,i}), \quad \text{where } j = 1, \dots, m_{k,i}. \end{aligned} \quad (1)$$

Here q_k and c_k are the mean and concentration parameters for the k -th group, q_0 and c_0 are hyper parameters for the hierarchical beta process, $z_{k,i} = \{z_{k,i,j} \mid j = 1, \dots, m_{k,i}\}$ is the history of pipe failure, $z_{k,i,j} = 1$ means the pipe failed in j -th year, otherwise $z_{k,i,j} = 0$.

For hierarchical beta process, a set of $\{q_k\}$ are used to describe failure rates of different groups of pipes. For each pipe group, with fixed concentration parameter c_k , the posterior

distribution becomes:

$$\begin{aligned}
 p(q_k, \pi_{k,1:n_k} \mid z_{k,1:n_k}) &\propto p(q_k, \pi_{k,1:n_k}, z_{k,1:n_k}) \\
 &= p(q_k) \prod_i \left[p(\pi_{k,i} \mid q_k) \prod_j p(z_{k,i,j} \mid \pi_{k,i}) \right] \\
 &\propto (1 - q_k)^{c_0(1-q_0)-1} q_k^{c_0 q_0 - 1} \prod_i \left[\frac{\pi_{k,i}^{c_k q_k - 1 + \sum_j z_{k,i,j}} (1 - \pi_{k,i})^{c_k(1-q_k) - 1 + m_{k,i} - \sum_j z_{k,i,j}}}{\Gamma(c_k q_k) \Gamma(c_k(1 - q_k))} \right],
 \end{aligned} \quad (2)$$

where $\pi_{k,1:n_k} = \{\pi_{k,i} \mid i = 1, \dots, n_k\}$ and $z_{k,1:n_k} = \{z_{k,i} \mid i = 1, \dots, n_k\}$. Marginalizing out all of $\pi_{k,i}$,

$$\begin{aligned}
 p(q_k \mid z_{k,1:n_k}) &\propto p(q_k, z_{k,1:n_k}) = p(q_k) \prod_i \left[\int p(\pi_{k,i} \mid q_k) p(z_{k,i} \mid \pi_{k,i}) d\pi_{k,i} \right] \\
 &\propto q_k^{c_0 q_0 - 1} (1 - q_k)^{c_0(1-q_0) - 1} \prod_i \frac{\Gamma(c_k q_k + \sum_j z_{k,i,j}) \Gamma(c_k(1 - q_k) + m_{k,i} - \sum_j z_{k,i,j})}{\Gamma(c_k q_k) \Gamma(c_k(1 - q_k))}.
 \end{aligned} \quad (3)$$

It can be known for each $\pi_{k,i}$,

$$\begin{aligned}
 p(\pi_{k,i} \mid q_k, z_{k,1:n_k}) &= \frac{p(q_k, \pi_{k,i} \mid z_{k,1:n_k})}{p(q_k \mid z_{k,1:n_k})} \\
 &= \frac{p(q_k) p(\pi_{k,i} \mid q_k) p(z_{k,i} \mid \pi_{k,i}) \prod_{l \neq i} \left[\int p(\pi_{k,l} \mid q_k) p(z_{k,l} \mid \pi_{k,l}) d\pi_{k,l} \right]}{p(q_k) \prod_l \left[\int p(\pi_{k,l} \mid q_k) p(z_{k,l} \mid \pi_{k,l}) d\pi_{k,l} \right]} \\
 &\propto \pi_{k,i}^{c_k q_k - 1 + \sum_j z_{k,i,j}} (1 - \pi_{k,i})^{c_k(1-q_k) - 1 + m_{k,i} - \sum_j z_{k,i,j}}.
 \end{aligned} \quad (4)$$

Hence given q_k ,

$$\pi_{k,i} \mid q_k, z_{k,i} \sim \text{Beta} \left(c_k q_k + \sum_j z_{k,i,j}, c_k(1 - q_k) + m_{k,i} - \sum_j z_{k,i,j} \right). \quad (5)$$

The posterior for each $\pi_{k,i}$ could be estimated after marginalizing q_k .

$$\begin{aligned}
 p(\pi_{k,i} \mid z_{k,1:n_k}) &= \int p(q_k, \pi_{k,i} \mid z_{k,1:n_k}) dq_k \\
 &= \int p(q_k \mid z_{k,1:n_k}) p(\pi_{k,i} \mid q_k, z_{k,i}) dq_k.
 \end{aligned} \quad (6)$$

The inference algorithm is shown in Fig. 2. Following the work of Thibaux and Jordan (2007), $p(q_k \mid z_{k,1:n_k})$ in (6) is approximated by a set of samples based on (3), and the concentration parameters are fixed during the inference process. However, both mean and concentration are important parameters that impact the underlying distribution (seen Fig. 3 for example). Even with the same mean, the concentration parameter could drastically vary the distribution. When the concentration value is large, the distribution highly concentrates to its mean value. When the concentration value is small, the distribution spreads over a big range and sometimes even becomes bimodal. To overcome this problem, the next section

Algorithm 1.

Input: pipe failure history $\{z_{k,i,j}\}$, $k = 1, \dots, K$, $i = 1, \dots, n_k$, $j = 1, \dots, m_{k,i}$; hyper parameters c_0 and q_0 ; concentration parameters $\{c_k\}$.

Output: The estimation of failure probability $\pi_{k,i}$ for each pipe.

for $t = 1, 2, \dots, T$,

 sample $q_k^{(t)}$ from $p(q_k | z_{k,1:n_k})$;

$\pi_{k,i}^{(t)} = \frac{c_k q_k^{(t)} + \sum_j z_{k,i,j}}{c_k + m_{k,i}}$;

end for

calculate $\bar{\pi}_{k,i} = \frac{1}{T} \sum_{t=1}^T \pi_{k,i}^{(t)}$;

return $\bar{\pi}_{k,i}$.

Fig. 2 Inference algorithm for hierarchical beta process with fixed concentration parameter

Fig. 3 Beta distributions with different mean and concentration parameter values

considers both mean and concentration parameters during the inference, which is important for more effective pipe condition assessment.

4.2 Approximate inference for sparse incident data

For each pipe group, considering both mean and concentration, the posterior distribution becomes

$$\begin{aligned}
 p(q_k, c_k, \pi_{k,1:n_k} | z_{k,1:n_k}) &\propto p(q_k, c_k, \pi_{k,1:n_k}, z_{k,1:n_k}) \\
 &= p(q_k) p(c_k) \prod_i \left[p(\pi_{k,i} | q_k, c_k) \prod_j p(z_{k,i,j} | \pi_{k,i}) \right].
 \end{aligned} \quad (7)$$

Assume non-informative prior for c_k and marginalize out $\pi_{k,1:n_k}$,

$$\begin{aligned}
 p(q_k, c_k | z_{k,1:n_k}) &\propto q_k^{c_0 q_0 - 1} (1 - q_k)^{c_0(1 - q_0) - 1} \\
 &\times \prod_i \frac{\Gamma(c_k) \Gamma(c_k q_k + \sum_j z_{k,i,j}) \Gamma(c_k(1 - q_k) + m_{k,i} - \sum_j z_{k,i,j})}{\Gamma(c_k + m_{k,i}) \Gamma(c_k q_k) \Gamma(c_k(1 - q_k))}.
 \end{aligned} \quad (8)$$

It can be seen that generally the joint distribution of q_k and c_k is computationally expensive to estimate. On the other hand, incident data like pipe failure records are usually quite sparse within the observation period. Most items do not fail or fail just once during the observation period (i.e. $\sum_j z_{k,i,j} \leq 1$). It can be seen that under such situation the computational complexity could be greatly reduced for the inference algorithm. Denote $s_{k,i} = \sum_j z_{k,i,j}$. According to the property of Gamma function, for sparse incident data,

$$p(q_k, c_k | z_{k,1:n_k}) \propto q_k^{c_0 q_0 - 1} (1 - q_k)^{c_0(1-q_0) - 1} \prod_i \left[\frac{(c_k q_k)^{s_{k,i}} \prod_{l=0}^{m_{k,i} - s_{k,i} - 1} (c_k(1 - q_k) + l)}{\prod_{l=0}^{m_{k,i} - 1} (c_k + l)} \right], \quad (9)$$

Using Taylor expansion, the posterior distribution could be approximated by

$$p(q_k, c_k | z_{k,1:n_k}) \propto q_k^{c_0 q_0 - 1} (1 - q_k)^{c_0(1-q_0) - 1} \times \prod_i \left[\left(\frac{c_k q_k}{c_k + m_{k,i} - 1} \right)^{s_{k,i}} (1 - q_k)^{\sum_{l=0}^{m_{k,i} - 1 - s_{k,i}} \frac{c_k}{l + c_k}} \right]. \quad (10)$$

Hence the posterior could be approximated by a beta distribution as well

$$p(q_k | c_k, z_{k,1:n_k}) \approx \text{Beta} \left(c_0 q_0 + \sum_i s_{k,i}, c_0(1 - q_0) + \sum_i \sum_{l=0}^{m_{k,i} - 1 - s_{k,i}} \frac{c_k}{l + c_k} \right). \quad (11)$$

The computational complexity can be further reduced when there is sufficient number of items in the group. Usually $n_k \gg 1$ for each group, it can be known that $\text{Var}(q_k | c_k, z_{k,1:n_k}) \ll 1$ since $\text{Var}(q_k | c_k, z_{k,1:n_k}) \leq \frac{1}{4(c_0 + \sum_i s_{k,i} + n_k)}$. Hence

$$q_k \approx \frac{c_0 q_0 + \sum_i s_{k,i}}{c_0 + \sum_i s_{k,i} + \sum_i \sum_{l=0}^{m_{k,i} - 1 - s_{k,i}} \frac{c_k}{l + c_k}}. \quad (12)$$

Given both q_k and c_k ,

$$\pi_{k,i} | q_k, c_k, z_{k,i} \sim \text{Beta} \left(c_k q_k + \sum_j z_{k,i,j}, c_k(1 - q_k) + m_{k,i} - \sum_j z_{k,i,j} \right). \quad (13)$$

For sparse incident data, both mean and concentration parameters are considered during the process of inference. The approximate inference algorithm is presented in Fig. 4.

5 Empirical results

In this experiment our HBP method was tested and compared with popular survival analysis methods, namely the Cox and Weibull models. The Cox model is usually used to model only the first time failure for each pipe, while the Weibull model can deal with multiple failures. The critical water mains are categorized into different groups according to their coating, region, and laid year. For fair comparison, the other explanatory factors act multiplicatively (Andreou et al. 1987) on the hazard rate in the Cox model or the priors in the Weibull model and the proposed method. Figure 5 shows the results of predicting pipe failures in two regions by different models. The test curves exhibit the average performance for the

Algorithm 2.

Input: pipe failure history $\{z_{k,i,j}\}, k = 1, \dots, K, i = 1, \dots, n_k, j = 1, \dots, m_{k,i}$; hyper parameters c_0 and q_0 ;

Output: The estimation of failure probability $\pi_{k,i}$ for each pipe.

for $t = 1, 2, \dots, T$,

sample $c_k^{(t)}$ and $q_k^{(t)}$ from (10) with $q_k = \frac{c_0 q_0 + \sum_j s_{k,i}}{c_0 + \sum_i s_{k,i} + \sum_i \sum_{l=0}^{m_{k,i}-1-s_{k,i}} \frac{c_k}{l+c_k}}$;

$$\pi_{k,i}^{(t)} = \frac{c_k^{(t)} q_k^{(t)} + \sum_j z_{k,i,j}}{c_k^{(t)} + m_{k,i}};$$

end for

calculate $\bar{\pi}_{k,i} = \frac{1}{T} \sum_{t=1}^T \pi_{k,i}^{(t)}$;

return $\bar{\pi}_{k,i}$.

Fig. 4 Approximate inference for hierarchical beta process with sparse incident data

Fig. 5 Results of pipe failure prediction for region A (left) and region B (right) by different models

Table 1 The performance diagnostics of different models

Metrics	region A			region B		
	HBP	Weibull	Cox	HBP	Weibull	Cox
sensitivity	0.58	0.54	0.53	0.65	0.59	0.55
specificity	0.59	0.54	0.53	0.66	0.60	0.54
F measure	0.48	0.44	0.43	0.48	0.43	0.44
AUC	0.57	0.50	0.47	0.61	0.58	0.50

last three years (2008–2010). To evaluate the prediction of pipe failures for a given year, all failure records available before that year are used as training data. The x -axis represents the cumulative percentage of inspected water pipes, and the y -axis represents the percentage of detected pipe failures.

Table 1 shows performance diagnostics including sensitivity, specificity, F measure, and AUC by different methods (the first three are calculated based on the point of the ROC curve where the false positive rate is approximately equal to the false negative rate). Table 1 indicates that there is more than 10 % improvement in AUC compared to previous

Table 2 The performance diagnostics (mean $\pm$ std.) using randomly selected subsets

Metrics	region A			region B		
	HBP	Weibull	Cox	HBP	Weibull	Cox
sensitivity	0.57 $\pm$ 0.03	0.55 $\pm$ 0.05	0.53 $\pm$ 0.07	0.67 $\pm$ 0.02	0.59 $\pm$ 0.04	0.57 $\pm$ 0.07
specificity	0.58 $\pm$ 0.03	0.56 $\pm$ 0.05	0.54 $\pm$ 0.07	0.68 $\pm$ 0.02	0.59 $\pm$ 0.04	0.57 $\pm$ 0.07
F measure	0.47 $\pm$ 0.04	0.45 $\pm$ 0.05	0.44 $\pm$ 0.07	0.51 $\pm$ 0.03	0.45 $\pm$ 0.04	0.46 $\pm$ 0.07
AUC	0.62 $\pm$ 0.03	0.54 $\pm$ 0.05	0.51 $\pm$ 0.08	0.65 $\pm$ 0.03	0.62 $\pm$ 0.02	0.52 $\pm$ 0.11

Table 3 Comparison between the proposed method and other methods using t -test

Metrics	region A		region B	
	HBP vs. Weibull	HBP vs. Cox	HBP vs. Weibull	HBP vs. Cox
sensitivity	$t = 2.40, p < 0.05$	$t = 2.12, p < 0.05$	$t = 5.39, p < 0.05$	$t = 3.85, p < 0.05$
specificity	$t = 2.36, p < 0.05$	$t = 2.13, p < 0.05$	$t = 5.37, p < 0.05$	$t = 3.86, p < 0.05$
F measure	$t = 1.53, p = 0.08$	$t = 1.08, p = 0.15$	$t = 2.10, p < 0.05$	$t = 1.98, p < 0.05$
AUC	$t = 4.81, p < 0.05$	$t = 5.11, p < 0.05$	$t = 2.23, p < 0.05$	$t = 4.79, p < 0.05$

methods, leading to a significant economic saving on pipe condition assessment and system maintenance (details of the impact are shown in the next section).

In the proposed algorithm, a set of samples is used to approximate the posterior distribution of mean and concentration parameters. The sampling based approximation will converge to the underlying distribution when the number of samples (or iterations) becomes large enough (about 5000 in our experiment). A simulation experiment with randomly selected subsets of the data (80 % of the pipes, repeating 10 times) has also been performed. Table 2 shows the corresponding performance diagnostics by different methods using randomly selected subsets. It can be seen that the proposed approach overall outperforms the other methods for pipe failure prediction.

Statistical tests (t -test at 5 % level of significance) have been performed to compare the proposed method with the other methods (see Table 3). Especially, one-sided paired t -test (Japkowicz and Shah 2011) is used to reveal the significance of the difference between two methods. It can be seen that the improvement is statistically significant for sensitivity, specificity, and AUC.

In addition, Table 4a shows the results of nonparametric test, i.e. Friedman test (Japkowicz and Shah 2011), for the comparison of the three methods on the two regions altogether. The corresponding post hoc tests, i.e. Nemenyi tests, have also been performed for pairwise comparison (see Table 4b). Consistent with the results of t -test, the results based on nonparametric test and followed post hoc tests also indicate that the improvement is statistically significant for the performance metrics.

Moreover, simulated data is used to evaluate the sensitivity of the approximation algorithm. In the experiment, the simulated data consists of 10 years' failure records for 1000 different items. The items are categorized into 20 groups and the overall failure rate is about 1 % per year. The failure probability of each item is estimated by the proposed algorithm. The experiment has been repeated three times. Comparing with the ground truth, the average error of failure probability estimation for each of the three simulations is 7.89×10^{-4} , 9.51×10^{-4} , and 9.95×10^{-4} , respectively. It can be seen that the proposed algorithm effectively estimates the failure probability in the simulation experiment.

Table 4 Comparison between the proposed method and other methods using (a) Friedman test and followed (b) Nemenyi tests

(a)		(b)		
Metrics	Friedman test	Metrics	HBP vs. Weibull	HBP vs. Cox
sensitivity	$\chi_F^2 = 12.9, p < 0.05$	sensitivity	$q = 2.85, p < 0.05$	$q = 3.32, p < 0.05$
specificity	$\chi_F^2 = 12.9, p < 0.05$	specificity	$q = 2.85, p < 0.05$	$q = 3.32, p < 0.05$
F measure	$\chi_F^2 = 7.6, p < 0.05$	F measure	$q = 2.21, p < 0.05$	$q = 2.53, p < 0.05$
AUC	$\chi_F^2 = 22.4, p < 0.05$	AUC	$q = 3.16, p < 0.05$	$q = 4.90, p < 0.05$

Fig. 6 Results of pipe failure prediction for region B by BP and HBP with and without fixed concentration parameter

From the prediction results shown in Fig. 5 and Table 1, it can be seen that the HBP model achieves better performance in most cases. Compared to parametric modelling, non-parametric modelling is not limited by a fixed model structure so it is able to adjust the model complexity adaptively to the accumulation of failure data. That flexibility enables the HBP model based method to provide a robust performance of pipe failure prediction in different regions.

For the hierarchical beta process, a set of mean and concentration parameters are used to describe the failure rate of different groups of pipes. Originally, the concentration parameters are fixed in the hierarchical beta process method. Compared to previous work that only estimates the means, both mean and concentration parameters are jointly estimated in this work, which appears to be important for more effective pipe condition assessment. Figure 6 shows prediction results with and without fixed concentration parameters. It can be seen that simultaneous estimation of mean and concentration assists the proposed method by improving the accuracy of failure prediction results.

Figure 7 shows the estimation of failure probabilities by the proposed method for pipes of different laid years in region B. With a fixed model structure, parametric methods like the Weibull model usually assume that the failure rate increases with pipe age, which clearly doesn't apply consistently. Such a fixed model structure can be overcome by nonparametric methods, leading to more accurate prediction results. Figure 7b shows the corresponding error between the predicted failure probability and the actual failure rate for recent three years. It can be seen that the estimation is consistent with the observed data. Moreover, sometimes there are no failures recorded in certain pipe groups, which does not necessarily mean that the water mains have never failed before, as the records could be missed or inaccurate. Compared to the beta process, the HBP allows dependencies across pipe groups,

Fig. 7 (a) Estimated failure probabilities by the proposed method for water mains laid from 1950s to 1970s in region B. (b) Error between the predicted failure probability and the actual failure rate

Fig. 8 Ranking of pipe failure probabilities in region A (*left*) and B (*right*)

leading to smoothed estimation results for each group. The hierarchical construction also helps improve the accuracy of pipe failure prediction (see also Fig. 6).

6 Impact on real-life application

The estimated failure probability provided by our method allows a direct ranking of all pipes, hence addresses the needs of the prioritization step. Using the pipes' location, we can draw risk maps based on failure probabilities, as in Fig. 8, which can be used to identify larger pipe sets requiring attention.

Let's now examine the practical implications of the improved prediction provided by our method (see Fig. 5) over the existing models currently used in the risk management process. The budget and resources allocable for pipe condition assessment are usually limited, so that each year only a small fraction of the critical water mains can be physically inspected, typically around 1 % of the whole network length. It is hence crucial that the top ranking pipes in the priority list should actually present a need for renewal, otherwise the inspection

Fig. 9 Detection results for first 1 % of all critical water mains in region B

costs, will be spent to no avail. Figure 9 compares the various methods assuming only 1 % of mains can be inspected. The rate of pipe failure prediction (both mean and variance) when inspecting top 1 % of the pipes is shown in Fig. 9c. The HBP outperforms the other methods by predicting almost 25 % of the failures, which when extrapolated to the whole urban network (with capital expenditure of about one million dollars on condition assessment per year), represents a saving evaluated to several hundred thousand dollars per year, over its Weibull counterpart.

Incidentally, this improvement in prediction accuracy also decreases the number of false negatives for critical mains about to fail, hence avoiding a number of disastrous critical main failures. As the financial and community cost of one critical main break ranges from hundreds of thousands to a few million dollars, this generates estimated savings in excess of a million dollars per year, again calculated over its Weibull counterpart.

The deployment will be through a software tool or system, using historical records in the existing format used by the utility, and providing a ranking list of pipes and their likelihood of failure. This tool would be integrated into a risk management system used by the utility, where the risk of each pipe is measured by considering both the probability of pipe failure and the corresponding economic and social cost of the failure. Our partner utility company is already using such a comprehensive risk management system, which includes a tool for failure prediction based on the Weibull model. To improve the effectiveness of risk management for preventative repairs, this work aims to upgrade the failure prediction model currently used in the system with the proposed method.

7 Expert commentaries

In this work, the involvement of domain experts includes the following: (1) Provide pipe failure records, define various terms, and explain related features; (2) Identify useful features based on their domain experience; (3) Provide feedback on our methods and results; (4) Advise how to represent results so that they can be readily understood by the industry; (5) Evaluate the outcomes based on the industry needs; (6) Guide in-field investigation of water pipes and supply pipe samples.

The following expert commentary is provided by Mr. Dammika Vitanage, a Program Manager for Treatment and Infrastructure at Sydney Water.

All water utilities with buried water pipes are faced with the issue of finding pipes that are at high-risk of failure before they fail and result in significant service disruption to the

community. In order to achieve this it is necessary to have an accurate method for remotely identifying high-risk pipes that is sufficiently robust to cope with the network variables of age, pipe material, environmental conditions and different degrees of urbanization.

This is very interesting work which has the potential to have a world-wide impact in pipe condition assessment. The work presented in this paper has been developed based on a relatively small data set to test the machine learning approach. We have been sufficiently impressed by the results to fund further development of this work and have introduced the approach to our international colleagues, who have also established a collaboration project with the research team.

In order to implement the new algorithm we require it to be further tested in different locations and get a feel for its ability to cope with the extreme events such as major rain events in reactive soils which can have a significant impact on pipe stresses. Australia is a good testing ground for this kind of work.

Results from this work suggest that the nonparametric modelling approach has potential to more accurately identify high risk pipes over the existing parametric models that are currently available to the water industry. This will allow us to target our condition assessment activity so that more poor-condition pipes are found with the same level of condition assessment activity. Further it could be possible to avoid replacement of those pipes that still have a safe level of remaining service life, providing more effective use of our assets. The information can be used to target effective monitoring plans and collect the right information for future prediction.

8 Lessons

The evaluation criteria for domain experts on specific application can be very different from the standard performance measurements used in machine learning domain. For instance, machine learning experts usually use cutting-off point or AUC area of ROC curve to measure an approach's performance. However, in this work, only a small portion of the performance curve (see Fig. 9) is of interest, because the amount of pipes that can be inspected in practice is only a small percentage of the whole network. Thus, the failure predication accuracy with 1 % of the pipes inspected is the main criteria to measure different approaches' performances.

Intermediate experimental statistics are suggestive of the model design. In some sense, it helps to avoid a biased assumption and suggest appropriate modelling. For instance, from the statistics of the pipe failures in different year/age (see Fig. 7), we can see the assumption that failure probability increases monotonically with the age of pipe is not always true.

The visualization and presentation of the results need adapt to the domain experts' habits and technical background. To domain experts, instead of the whole performance curve, only a small portion of the curve (see Fig. 9) is of interest. The risk map in Figure 8, showing the location of each pipe, is another example of such adaptation, where risk levels of individual pipes are depicted by different colours.

Categorizing the data helps to identify subsets of the whole dataset which possess very different failure patterns. It makes the modelling process simpler and more accurate for a system affected by a large variety of factors, especially when part of them are not ascertainable or measurable. Usually water main systems also exhibit that characteristic because they are deployed across large scale areas with very different environmental factors, such as traffic load, soil corrosivity, etc. In our experiment, separately modelling the failure patterns of region A and region B provides more accurate results than modelling the two regions together.

9 Conclusion

This work proposes the use of Bayesian nonparametric learning to predict water pipe condition from existing infrastructure data. Compared to traditional statistical modelling approaches, nonparametric modelling is not limited by a fixed model structure and it is able to adjust the model complexity adaptively according to the complexity of failure data. This work proposes the use of Bayesian nonparametric learning to predict water pipe condition from existing infrastructure data. Especially, an approximate inference algorithm has also been developed for hierarchical beta process with sparse incident data. Our experimental results exhibit that nonparametric modelling outperforms previous parametric modelling for pipe condition assessment. In practical terms, this represents major financial savings through more targeted condition assessment inspections, as well as a reduction in the number of extremely costly critical mains failures. In our future work, heterogeneous types of information (cross-region and sometimes incomplete) will be fused in nonparametric modelling for condition prediction and decision support.

Acknowledgements This study is a joint work of National ICT Australia and Sydney Water Corporation. The authors extend their appreciation to Dammika Vitanage of Sydney Water Corporation for his expert commentaries.

References

- Andreou, S. A., Marks, D. H., & Clark, R. M. (1987). A new methodology for modelling break failure patterns in deteriorating water distribution systems: theory. *Advances in Water Resources*, 10, 11–20.
- Awumah, K., Goulter, I., & Bhatt, S. K. (1990). Assessment of reliability in water distribution networks using entropy based measures. *Stochastic Hydrology and Hydraulics*, 4, 309–320.
- Chen, X., Zhou, M., & Carin, L. (2012). The contextual focused topic model. In *18th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 96–104).
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B Methodological*, 187–220.
- Fares, H., & Zayed, T. (2010). Hierarchical fuzzy expert system for risk of failure of water mains. *Journal of Pipeline Systems Engineering and Practice*, 1, 53–62.
- Ferguson, P., Heathcote, M., Moore, G., & Russell, D. (1996). Condition assessment of water mains using remote field technology. *Journal - American Water Works Association*, 23, 6–8.
- Griffiths, T. L., & Ghahramani, Z. (2011). The Indian Buffet process: an introduction and review. *Journal of Machine Learning Research*, 12, 1185–1224.
- Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18(3), 1259–1294.
- Ibrahim, J., Chen, M.-H., & Sinha, D. (2001). *Bayesian survival analysis*. Berlin: Springer.
- Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms, a classification perspective*. Cambridge: Cambridge University Press.
- Kettler, A. J., & Goulter, I. C. (1985). An analysis of pipe breakage in urban water distribution networks. *Canadian Journal of Civil Engineering*, 12, 286–293.
- Klein, J. P., & Goel, P. K. (1992). *Survival analysis: state of the art*. Berlin: Springer.
- Kleiner, Y., & Rajani, B. B. (2001). *Comprehensive review of structural deterioration of water mains: statistical models* (Technical report NRCC-42586). National Research Council Canada.
- Marks, H. D., Andreou, S., Jeffrey, L., Park, C., & Zaslavski, A. (1987). *Statistical models for water main failures*. US Environmental Protection Agency (Co-operative Agreement CR8-1-0558) M.I.T. Office of Sponsored Projects No. 94211. Boston, MA.
- Micevski, T., Kuczera, G., & Coombes, P. (2002). Markov model for storm water pipe deterioration. *Journal of Infrastructure Systems*, 8, 49–56.
- Misiunas, D. (2005). *Failure monitoring and asset condition assessment in water supply systems*. Doctoral dissertation, Lund University, Sweden.
- Murray, J. F., Hughes, G. F., & Kreutz-Delgado, K. (2006). Machine learning methods for predicting failures in hard drives: a multiple-instance application. *Journal of Machine Learning Research*, 6, 783–816.

- Paisley, J., & Carin, L. (2009). Nonparametric factor analysis with beta process priors. In *26th Annual international conference on machine learning* (pp. 777–784).
- Pelletier, G., Mailhot, A., & Villeneuve, J. P. (2003). Modeling water pipe breaks-three case studies. *Journal of Water Resources Planning and Management*, 129(2), 115–123.
- Rajani, B. B., & Kleiner, Y. (2001). *Comprehensive review of structural deterioration of water mains: physically based models* (Technical report NRCC-43722). National Research Council Canada.
- Rogers, P. D. (2011). Prioritizing water main renewals: case study of the Denver water system. *Journal of Pipeline Systems Engineering and Practice*, 2, 73–81.
- Rogers, P. D., & Grigg, N. S. (2009). Failure assessment modeling to prioritize water pipe renewal: two case studies. *Journal of Infrastructure Systems*, 15, 162–171.
- Shamir, U., & Howard, C. D. D. (1979). An analytic approach to scheduling pipe replacement. *Journal - American Water Works Association*, 71(5), 248–258.
- Tchórzewska-Cieślak, B. (2011). Fuzzy failure risk analysis in drinking water technical system. *Reliability: Theory and Applications*, 2, 138–148.
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101, 1566–1581.
- Thibaux, R., & Jordan, M. I. (2007). Hierarchical beta processes and the Indian Buffet process. In *International conference on artificial intelligence and statistics* (pp. 564–571).
- Tian, C. H., Xiao, J., Huang, J., & Albertao, F. (2011). Pipe failure prediction. In *IEEE international conference on service operations, logistics, and informatics*.
- Wang, Y., Li, Z., Wang, Y., & Chen, F. (2013). A Bayesian non-parametric viewpoint to visual tracking. In *Workshop on the applications of computer vision* (pp. 482–488).
- Yang, S. K. (2003). A condition-based failure-prediction and processing-scheme for preventive maintenance. *IEEE Transactions on Reliability*, 52, 373–383.
- Zhou, M., Chen, H., Paisley, J., Lu, R., Sapiro, G., & Carin, L. (2009). Non-parametric Bayesian dictionary learning for sparse image representations. In *Advances in neural information processing systems*.
- Zhou, M., Yang, H., Sapiro, G., Dunson, D., & Carin, L. (2011). Dependent hierarchical beta process for image interpolation and de-noising. In *International conference on artificial intelligence and statistics* (pp. 883–891).