Estimating burst probability of water pipelines with a competing hazard model

H. Shin, K. Kobayashi, J. Koo and M. Do

ABSTRACT

Because pipeline systems represent more than 80% of the total asset value of water-distribution systems, their management is an important issue for water utilities. A pipeline deteriorates over time after installation and, along with the deterioration, pipe bursts can occur as various types, and the choice of a maintenance and repair strategy will depend on the burst types. It is therefore important to forecast the occurrence probability of each burst type. This paper addresses a competing deterioration-hazard model that allows modelling of deterioration by multiple types of failure and focuses on the bursts which occur in the pipe body or connection. The Weibull hazard model is used to address the lifetime of each pipeline, measured from when it was buried, and the model takes into account the competing nature of various types of failure by using a competing hazard model. The competing deterioration-hazard model allows us to determine the probability of deterioration in the pipe body and connection. The model is estimated by Bayesian inference using a Markov chain Monte Carlo method. The applicability of the method to data for an existing pipeline system is examined.

Key words | asset management, Bayesian inference, competing deterioration-hazard model, pipe burst, pipeline systems

H. Shin (corresponding author)
K. Kobayashi
Department of Urban Management,
Kyoto University,
Kyotodaigaku-katsura, Nisigyou-ku,
Kyoto 615-8540,
Japan
E-mail: *flyhigh21c@gmail.com*

J. Koo
Department of Environmental Engineering,
University of Seoul,
Seoul,
Korea

M. Do
Department of Urban Engineering,
Hanbat National University,
Daejeon,
Korea

INTRODUCTION

Water-supply pipelines, which form important components of the infrastructure of cities, require huge annual maintenance budgets. Consequently, the establishment of optimal regimes for maintaining water pipeline systems has become a major issue for water-utility managers throughout the world. In the management of infrastructure assets, optimal maintenance strategies are frequently based on lifecycle-cost analysis, which is dependent on the deterioration model (Kobayashi *et al.* 2010).

In the field of water-supply systems, many studies have been conducted to assess the condition of pipeline systems and to predict their deterioration process. Shamir & Howard (1979) and Marvin (1996) assumed that breaks in pipelines increase exponentially with their age, and they obtained break-prediction models by using regression analysis. Clark *et al.* (1982) reported a method for estimating the expected failure time of pipelines, whereas Shinstine (1999) examined the relationship between pipeline breaks and the diameters of pipes. Because water pipeline systems are usually buried underground, the monitoring and inspection of such systems is difficult and it is hard to accumulate adequate observational data for use as a basis for deterioration forecasting analysis. Because of the difficulties in observing deteriorations of pipelines directly, we decided to predict the deterioration of pipelines by examining failures caused by the deterioration process.

Marks (1985), Constantine & Darroch (1995), and Park (2004) used proportional hazards models, based on the failure-prediction model proposed by Cox (1972), to predict the risk of a pipeline break. Many probabilistic models that use various probability functions have been developed with the average annual number of pipe breaks on the pipeline

systems as an indicator of the structural state and the times to failure between pipe breaks considered as random variables (Le Gat & Eisenbeis 2000; Mailhot *et al.* 2000, 2003; Pelletier *et al.* 2003). These models have overcome challenges that observation data typically show properties, right censored observations (Eisenbeis *et al.* 1999; Mailhot *et al.* 2000), left truncation (Mailhot *et al.* 2000) and selective survival bias (Scheidegger *et al.* 2013). By setting the deterioration state as a binary condition, 'failure' or 'normal operation', it is possible to predict the service life of a pipeline by using a conventional hazard model. There are numerous reports of studies in which this type of deterioration prediction method has been applied to other types of system. Aoki *et al.* (2005) proposed a method in which a Weibull hazard model is used to predict the lifetime of tunnel lighting equipment. Tanaka *et al.* (2010) similarly used Weibull hazard models to predict the deterioration of pipelines.

In general, the major cause of interruption of water pipeline systems is deterioration of the pipes. In the conventional models for the prediction of pipeline deterioration prediction, the type of pipe failure is not classified, and all failures are considered as a single type of failure. In a real pipeline system, however, pipe failures caused by deterioration appear in various forms. We therefore classified pipeline failures as 'B-burst', which occur in the pipe body, or 'C-burst', which occur in pipe-connection parts. The lifetime of a given part is defined as the period from its installation to burst, and in this study it is assumed that the burst is regarded as major damage and the damaged pipeline is replaced immediately. The Weibull deterioration-hazard model is used to address the lifetime of each pipeline, and takes into account the nature of the competition between several types of failure by using a competing deterioration-hazard model. The deterioration of the pipeline is predicted by developing a competing deterioration-hazard model that considers competition between C-burst and B-burst. The proposed competing deterioration-hazard model allows us to determine the probability density of bursts in the pipe body and connection.

The competing hazard model assumes that competing causes of failure are independent of one another and that the incidence of each cause of failure can be analyzed from lifetime data. Such methods have been used in many fields, including medicine, economics, and engineering. The competing

hazard model is widely used in accelerated lifetime testing (ALT) to estimate the lifetime distribution of components. Nelson (1990) discussed an analysis of typical competing hazard models for constant stress ALT data. Kim & Bai (2002) reported a competing hazard model that considered only two competing causes of failure by using ALT data.

Because the pipeline systems are underground, system administrators face difficulties due to insufficient amounts of observation data. Thus, the insufficiency of data interrupts the practical application of the statistical model. In order to overcome this problem, in this study, the competing deterioration-hazard model is estimated by a Bayesian technique based on the Metropolis–Hasting method (M-H method), a Markov chain Monte Carlo method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult.

METHOD

Competing deterioration-hazard model

Competing hazards appear in cases in which two or more events can occur. The main idea of a competing hazard model is that the occurrence of an event of interest has to be taken into account while considering the occurrence of competing events. In pipeline systems, the case in which a pipeline is replaced because of a B-burst can be considered. However, there might be the events that could lead to replacement of the pipeline, such as C-bursts. It is therefore possible to assume that a C-burst is a competing event, if we assume that a B-burst is the event of interest, because a B-burst interrupts the occurrence of a C-burst. To introduce the competing hazard among the pipe failure types, the major damage which requires pipe replacement is considered. In this study we focus on the B-burst and C-burst.

Herein, we classify the state of a pipeline as being one of two condition levels: a 'healthy condition' and a 'burst' resulting from B-burst or C-burst. It is assumed that the burst denotes a state in which major damage is found and replacement is required immediately. On the other hand, the healthy condition reflects not only a normal operation but a condition where no major damage is found. In addition, it is also assumed that the records of past repaired

incidents, leakages or breaks are not considered as burst because these incidents would not be major damage.

The pipeline system must be discretized into pipe segments (Mailhot *et al.* 2000). A pipe segment is defined as the same as a pipeline is a series of pipes with relatively homogeneous characteristics such as pipe diameter and type of material and installation period.

Each pipeline is represented by i $(i = 1, \ldots, n)$, and the elapsed time from laying pipeline i to the present is expressed as τ. In addition, we assume that more than one type of pipe burst j $(j = 1, \ldots, J)$ is possible for pipeline i. The life span of pipeline until burst in pipe i is expressed by the random variable ζ_i and this is subject to the probability-density function $f_j(\zeta_i)$ and the distribution function $F_j(\zeta_i)$ for each type of burst type j. Here, the domain of life span ζ_i is $[0, \infty)$. In addition, the probability that pipe burst will not occur until time τ is defined as $[\tilde{F}(\tau)]$ and is known as the survival probability. This can be expressed as follows:

$$\tilde{F}(\tau) = 1 - F(\tau) \tag{1}$$

when competing hazards exist, the conditional probability that pipe burst does not occur in pipe i until an arbitrary time t_i and that pipe burst occurs by burst type j during the time span $[\tau, \tau + \Delta\tau]$ can be represented by the following equation:

$$\lambda_j(\tau) = \lim_{\Delta\tau \to 0} \frac{\Pr(\tau < \zeta \le \tau + \Delta\tau, TF = j|\zeta > \tau)}{\Delta\tau} = \frac{f_j(\tau)}{\tilde{F}(\tau)} \tag{2}$$

where $\lambda_j(\tau)$ is the hazard function for each burst type j and TF denotes type of burst. Note that in this competing hazard model, to obtain the hazard function, the density function $f_j(\tau)$ should be divided by $\tilde{F}(\tau)$ rather than by $\tilde{F}_j(\tau)$. For example, dividing $f_c(\tau)$ by $\tilde{F}_c(\tau)$ gives the conditional probability that a C-burst will not occur before an arbitrary time τ and that a C-burst will occur at τ. However, in this case the probability that a pipe burst will occur through C-burst would be overestimated, because the occurrence of a pipe B-burst, which is a competing risk of pipe C-burst, is not considered. It is therefore reasonable that the hazard function for C-burst has to be defined as the probability of a C-burst occurring when no pipe burst occurs until the arbitrary time τ.

The overall survival function $\tilde{F}(\tau)$ can be defined as follows:

$$\tilde{F}(\tau) = \exp\left\{ - \int_0^\tau \lambda(u) \, du \right\} \tag{3}$$

where $\lambda(\tau)$ is overall hazard function and is defined by $\lambda(\tau) = \sum_{j=1}^J \lambda_j(\tau)$. The overall survival function $\tilde{F}(\tau)$ is the probability that any burst type does not occur; it can therefore be represented by the joint probability of the partial survival distribution function for each burst type $\tilde{F}_j(\tau)$, $(j = 1, \ldots, J)$, as follows:

$$\tilde{F}(\tau) = \prod_{j=1}^J \tilde{F}_j(\tau) \tag{4}$$

The partial survival function $\tilde{F}_j(\tau)$ can also be defined as follows:

$$\tilde{F}_j(\tau) = \exp\left\{ - \int_0^\tau \lambda_j(u) \, du \right\} \tag{5}$$

Accordingly, from Equation (2), the partial density function for each type of burst $f_j(\tau)$ can be expressed as follows:

$$\begin{aligned} f_j(\tau) &= \lambda_j(t_i)\tilde{F}(\tau) \\ &= \lambda_j(\tau) \prod_{j=1}^J \exp\left\{ - \int_0^\tau \lambda_j(u) \, du \right\} \end{aligned} \tag{6}$$

Pipe burst depends largely on the duration of use of the pipeline. The hazard function should therefore consider the elapsed time. In this study, the Weibull hazard model, which is suitable for addressing this process, is applied with the assumption that the probability of pipe burst increases with time, as follows:

$$\lambda_j(\tau) = \gamma_j m_j \tau^{m_j - 1} \tag{7}$$

where m_j is the acceleration parameter that represents the time dependency of the hazard function and γ_j is the parameter expressing the arrival rate of pipe burst. It is assumed that γ_j depends on the characteristics of the

pipeline, and that it can be expressed as follows:

$$\gamma_j = \exp\left(\boldsymbol{x}_i \boldsymbol{\beta}'_j\right) \tag{8}$$

where $\boldsymbol{x}_i = (x_i^1, \cdots, x_i^k)$ is the characteristic vector that represents the observed value for pipeline i and $\boldsymbol{\beta}_j = (\beta_j^1, \cdots, \beta_j^k)$ represents the unknown parameter vectors. In addition, k is the total number of covariates and the sign' denotes transposition. By using the Weibull hazard model, the probability-density function $f_j(\tau)$ and survival function $\tilde{F}_j(\tau)$ can be expressed as follows:

$$f_j(\tau) = \gamma_j m_j \tau^{m_j - 1} \exp\left(-\gamma_j \tau^{m_j}\right) \tag{9}$$

and

$$\tilde{F}_j(\tau) = \exp\left(-\gamma_j \tau^{m_j}\right) \tag{10}$$

Estimation method

Let us discuss the estimation method for the competing deterioration-hazard model based on inspection data. The time at which the pipe was buried is set as $t = 0$ and t_i denotes the observed duration of use of pipeline i ($i = 1, \ldots, n$). If a pipe burst occurs and the life span of the pipeline ends, its duration of use is equal to the life span, $\zeta_i = t_i$. On the other hand, it is assumed that a pipeline for which no burst has been reported until the inspection time still survives. In other words, if a pipeline's life span has not ended, this exceeds its duration of use, $\zeta_i > t_i$. Then, let us introduce the dummy variable ε_i, which denotes whether pipe burst has occurred or not:

$$\varepsilon_i = \begin{cases} 0 & \text{survival} \\ 1 & \text{burst} \end{cases} \tag{11}$$

In addition, the reported pipe burst type, in this study, the burst type is classified as either a C-burst ($j = c$) or a B-burst ($j = b$), can also be represented by the dummy variable d_i:

$$d_i = \begin{cases} 0 & \text{B} - \text{burst} \\ 1 & \text{C} - \text{burst} \end{cases} \tag{12}$$

The observation information for pipeline i can be represented as follows: $\xi_i = (d_i, \varepsilon_i, t_i, \boldsymbol{x}_i)$. Here, we define the unknown parameter vector for the competing deterioration-hazard model as $\theta = (\boldsymbol{\beta}, \boldsymbol{m})$. The parameters, $\boldsymbol{\beta}$ and $\boldsymbol{m}$ denote $\boldsymbol{\beta} = (\boldsymbol{\beta}_c, \boldsymbol{\beta}_b)$ and $\boldsymbol{m} = (m_c, m_b)$, respectively. If we suppose that there is observed information for pipeline i, $\xi_i = (d_i, \varepsilon_i, t_i, \boldsymbol{x}_i)$, the conditional probability that the observed information occurs in pipeline i can be represented by the following equation:

$$
\begin{aligned}
&\ell\left(d_i, \varepsilon_i, t_i | \boldsymbol{x}_i, \theta\right) \\
&= \left[\left\{\lambda_c\left(t_i | \boldsymbol{x}_i, \theta\right) \tilde{F}_c\left(t_i | \boldsymbol{x}_i, \theta\right) \tilde{F}_b(t_i | \boldsymbol{x}_i, \theta)\right\}^{d_i}\right. \\
&\quad \left. \cdot\left\{\lambda_b\left(t_i | \boldsymbol{x}_i, \theta\right) \tilde{F}_b\left(t_i | \boldsymbol{x}_i, \theta\right) \tilde{F}_c\left(t_i | \boldsymbol{x}_i, \theta\right)\right\}^{1 - d_i}\right]^{\varepsilon_i} \\
&\quad \cdot\left[\tilde{F}_c\left(t_i | \boldsymbol{x}_i, \theta\right) \tilde{F}_b\left(t_i | \boldsymbol{x}_i, \theta\right)\right]^{1 - \varepsilon_i}
\end{aligned} \tag{13}
$$

This assumes that the pipe burst of each of the n pipelines is mutually independent from that of other parts of the pipeline system. The simultaneous probability density of the pipe deterioration can therefore be expressed by the following likelihood function:

$$L(\theta | \xi) = \prod_{i=1}^{n} \ell\left(d_i, \varepsilon_i, t_i | \boldsymbol{x}_i, \theta\right) \tag{14}$$

where ξ represents $\xi = (\xi_1, \cdots, \xi_n)$. The unknown parameter θ can be estimated by the maximum likelihood estimation (MLE) method, which provides an estimate of parameter θ that maximizes the likelihood function.

Bayesian estimation method for competing deterioration-hazard model

In the MLE method, huge amounts of data are required to secure precision but it is not always possible to accumulate a great number of data, especially in the pipeline systems. Pipeline system administrators face difficulties of insufficient amounts of observation data because pipeline systems are underground. The Bayesian estimation method can provide estimation results by fusing prior information, such as human experience and expert knowledge, with an insufficient amount of observation data (Kobayashi et al. 2012). In

addition, the Bayesian estimation method is easy in comparison with the MLE method because in the Bayesian estimation method it is not required to derive the Jacobian and Hessian matrices. Furthermore, in the high dimensional nonlinear equation problem, the equation has multiple local optimum solutions. In this case, a poor choice of starting point in the MLE method can cause converging to a local optimum that is not the global optimum, or failure to converge entirely. The competing deterioration-hazard model is a high dimensional nonlinear expression of parameter θ and the optimization problem may have a large number of solutions including complex valued solutions. Thus, in this case, using the Bayesian estimation method instead of the maximum likelihood estimation method can solve the high dimensional nonlinear multinomial expression. In this section, we present a methodology for estimating the unknown parameter vector θ of the competing deterioration-hazard model by means of a Bayesian estimation method using observed data.

The Bayesian approach permits the estimation of θ on the basis of the inspection data ξ and prior information regarding θ. By using the M-H method, the estimation is carried out by sampling a large number of values of θ from its posterior distribution, which can be expressed as follows:

$$\pi(\theta|\xi) \propto L(\theta|\xi)\pi(\theta) \tag{15}$$

where $\pi(\theta|\xi)$ is the posterior probability density function of θ, $L(\theta|\xi)$ is the likelihood function, and $\pi(\theta)$ is the prior probability density function of θ. The newly obtained data are denoted by $\xi = (\xi_1, \cdots \xi_n)$. By substituting the Weibull hazard model (9) and (10) into Equation (14), the likelihood function can be expressed as follows:

$$L(\theta|\xi) = (m_c)^{d\cdot\varepsilon}(m_b)^{(1-d)\cdot\varepsilon} \exp\left\{\sum_{i=1}^{n}(d_i\varepsilon_i x_i\beta_c' + d_i\varepsilon_i(m_c-1)\right.$$
$$\ln t_i + (1-d_i)\varepsilon_i x_i\beta_b' + (1-d_i)\varepsilon_i(m_b-1)$$
$$\left. \ln t_i - \exp(x_i\beta_c')t_i^{m_c} - \exp(x_i\beta_b')t_i^{m_b} \right\} \tag{16}$$

where $\bar{d} = \sum_{i=1}^{n}\bar{d}_i$ and $\bar{\varepsilon} = \sum_{i=1}^{n}\bar{\varepsilon}_i$.

In this study we assume that the prior probability density function of parameter, m and β, follow a gamma distribution and a conjugate multidimensional normal distribution, respectively, $m \sim g(m_0, k_0)$, $\beta \sim \mathcal{N}_K$ (μ_0, Σ_0). With this assumption, the probability density function of the gamma distribution function $g(m_0, k_0)$ and the K-dimensional normal distribution $\mathcal{N}_K(\mu_o, \Sigma_o)$ can be further expressed as follows:

$$f(m|m_0, k_0) = \frac{1}{\Gamma(m_0)}k_0^{m_0}m^{m_0-1}e^{-k_0 m} \tag{17}$$

and

$$g(\beta|\mu_o, \Sigma_o) = \frac{1}{\sqrt{(2\pi)^K|\Sigma_o|}}$$
$$\cdot \exp\left\{-\frac{1}{2}(\beta-\mu_o)\Sigma_o^{-1}(\beta-\mu_o)'\right\} \tag{18}$$

where $\Gamma(m_0)$ denotes the gamma function and μ_o and Σ_o represent the prior expectation vector and the prior variance-covariance matrix of $\mathcal{N}_K(\mu_o, \Sigma_o)$, respectively. On the basis of Equation (15), the posterior probability density function $\pi(\theta|\xi)$ is defined as follows:

$$\pi(\theta|\xi) \propto L(\theta|\xi)f(m|m_0, k_0)g(\beta|\mu_o, \Sigma_o)$$
$$\propto (m_c)^{d\cdot\varepsilon}(m_b)^{(1-d)\cdot\varepsilon}m^{m_0-1}\exp\left\{\sum_{i=1}^{n}(d_i\varepsilon_i x_i\beta_c' + d_i\varepsilon_i(m_c-1)\right.$$
$$\ln t_i + (1-d_i)\varepsilon_i x_i\beta_b' + (1-d_i)\varepsilon_i(m_b-1)\ln t_i - \exp(x_i\beta_c')t_i^{m_c}$$
$$\left. -\exp(x_i\beta_b')t_i^{m_b} - k_0 m - \frac{1}{2}(\beta-\mu_o)\Sigma_o^{-1}(\beta-\mu_o)'\right)\right\} \tag{19}$$

The M-H method is used to perform sampling from an empirical distribution that is similar to $\pi(\theta|\xi)$ and accordingly obtains samples from the original distribution (Kobayashi & Kaito 2012). Furthermore, a random walk is used to improve the efficiency of sampling. The M-H method is described below.

Step 1. Initial establishment

The initial value of parameters $\theta = (\beta, m)$, the number of iterations for parameter sampling $\bar{N}$, and the burn-in period $\underline{N}$ are established. In addition the stride of the random walk is set.

Step 2. Sample extraction for estimation of the parameter θ

When the number of simulations is $n+1$, the parameter estimation θ^{n+1} is generated as described in Steps 2-1 to 2-3.

Step 2-1

The stride of the random walk ν is assumed to follow a normal distribution with a mean of 0 and a variance of $(\sigma)^2$. The new candidate value θ' is then calculated as follows:

$$\theta' = \theta^n + \nu \qquad (20)$$

Step 2-2

The acceptance probability is calculated as follows:

$$\alpha_{n+1} = \min\left\{\frac{\pi(\theta'|\xi)}{\pi(\theta^{(n)}|\xi)},\ 1\right\} \qquad (21)$$

Step 2-3

The uniform distribution $u_n \sim U(0,1)$ is generated, and then the sample is determined by applying the following condition:

$$\theta^{(n+1)} = \begin{cases} \theta', & \text{if} \quad u_n \leq \alpha_n \\ \theta^{(n)}, & \text{if} \quad u_n > \alpha_n \end{cases} \qquad (22)$$

If the acceptance probability is greater than u_n, the candidate value is accepted; otherwise, the original value is retained.

Step 3. Final judgment of the algorithm

Step 2 is repeated until the number of samplings reaches N.

The samples are then accumulated except for those that were generated during the burn-in period. If the number of samples N is sufficiently large, the parameters estimated by using the above algorithm will converge on the estimated value of the posterior distribution. Geweke test statistics (Geweke 1992) are used to test whether the sampling process of the M-H method reaches a steady state and the number of samplings N is appropriate or not.

EMPIRICAL STUDY

Overview of the empirical study

To analyze the deterioration of a real pipeline, we focused on the water distribution system of S city in South Korea. The pipe material, ductile cast iron pipe (DCIP), is regarded as the target for this study. The whole data of DCIPs comprise approximately 26,500 pipelines, 850 km in length, with an average age of around 13 years. Inspection data were obtained from historical records for pipe bursts in S city during the 9-year period 2001–2009. During this period, 1,405 cases of pipe replacement caused by B- and C-burst were recorded. Here, in this study, it is assumed that the replaced pipelines had major damage and its condition state is classified as burst. On the other hand, the historical records of past repair are not considered as burst because a repair is not associated with major damage. Table 1 shows the basic information of the data used in this study.

The inspection data contain information on whether or not pipe burst occurred and the type of burst for each damaged pipeline. In this study, the type of burst is classified as either a B-burst in a pipe body or a C-burst in pipe connections. Accidents that occurred in other subcomponents, such as valves, rubber packings, and so on, are neglected. On the other hand, the pipe diameter and length are used as characteristic information that affects pipe burst. On the basis of this information, the duration of survival before burst of a pipe is expressed by using the Weibull hazard model, and the competing deterioration-hazard model is used to consider the competition between C-bursts and B-bursts in the pipeline. The model is then estimated by using the Bayesian estimation method.

Estimation results

The Weibull hazard model used for the Bayesian estimation is specified as follows:

$$\lambda_j(t_i) = \exp\left(\beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2}\right)m_j t_i^{m_j-1}$$
$$(i = 1, \cdots, n;\ j = c,\ b) \qquad (23)$$

Table 1 | Features of target pipelines

Features	Value		
Material	Ductile cast iron		
Years laid (average age)	From 1957 to 2009 (13 years)		
Diameter/mm	75–900		
Number of pipes	26,577		
Total length/km	848.1		
Number of bursts	1,405	C-burst	833
		B-burst	572

The unknown parameter β_{j0} is a constant term, β_{j1} and β_{j2} represent the pipe diameter and pipe length, respectively. In this study, other characteristic variables that reflect the influence of outer and inner rust, soil unit weight, top traffic volume, and so on were neglected, either because of their small impacts or because data were unavailable. The unknown parameters can be expressed as follows:

$$\theta = (\beta_{c0}, \ \beta_{c1}, \ \beta_{c2}, \ \beta_{b0}, \ \beta_{b1}, \ \beta_{b2}, \ m_c, \ m_b) \tag{24}$$

We assume that the prior probability density function of the unknown parameters, m and β follow $m \sim g(m_0, k_0)$, $\beta \sim \mathcal{N}_K(\mu_0, \Sigma_0)$. Unfortunately, in this empirical study, because of the absence of detailed substantive knowledge, it is difficult to obtain information about the expectations and the variance of unknown parameters. However, if the number of observed data is large enough, the influence of prior distribution can be ignored. Thus, a non-informative prior distribution is applied for the Bayesian estimation. The non-informative prior distribution can be obtained by setting the variance of the prior distribution to be sufficiently large, as follows:

$$m \sim g\left(1, \ k_m^{-1}\right) \tag{25}$$

$$\beta \sim \mathcal{N}_K\left(\mathbf{O}, k_\beta \mathbf{I}\right) \tag{26}$$

where k_m and k_β are sufficiently large integers. $\mathbf{O}$ and $\mathbf{I}$ are a zero vector and a unit matrix, respectively, and N is normal distribution.

In order to improve the precision of estimation, the Bayesian updating rule (Kobayashi et al. 2012) is used. We created three different data groups (D_{2000}, D_{5000}, D_{10000}) which are extracted based on original data set. Here, the subscript numbers denote the number of extracted data. The estimation is performed in the order of the small size of the data and the estimation results (the mean, variance and covariance) are used as prior information of the next estimation using the Bayesian updating rule.

To conduct the M-H method, the number of iterations required to reach a steady state (the burn-in period) was set to $\underline{N} = 10,000$ and the number of iterations for parameter sampling was set to $\bar{N} = 20,000$. The 10,000 burn-in samples were omitted and the remaining 10,000 parameter samples were used to carry out the estimation.

Table 2 shows the results of the Bayesian estimation of competing deterioration-hazard models for each of the databases D_{2000}, D_{5000}, D_{10000} and original data set. The estimations obtained by the M-H method show the

Table 2 │ Results of estimation of parameters for the competing deterioration-hazard model

Parameters	D_{2000}	D_{5000}	D_{10000}	Original data set
β_{c0}	−11.681 (−12.874, −10.617)	−9.955 (−10.717, −9.293)	−9.504 (−9.963, −9.055)	−9.594 (−9.941, −9.277)
	0.053	0.086	0.036	0.069
β_{c1}	−3.284 (−5.175, −1.165)	−2.929 (−4.369, −1.534)	−0.763 (−1.615, 0.047)	−0.994 (−1.595, −0.406)
	0.134	0.035	0.060	0.011
β_{c2}	6.344 (1.065, 10.737)	5.194 (2.239, 7.885)	2.755 (1.412, 3.800)	2.657 (1.941, 3.284)
	0.042	0.129	0.124	0.038
β_{b0}	−12.461 (−14.264, −10.842)	−10.332 (−11.358, −9.374)	−10.166 (−10.790, −9.644)	−10.094 (−10.429, −9.768)
	0.139	0.200	0.067	0.226
β_{b1}	−2.489 (−5.115, 0.287)	−3.100 (−5.468, −0.822)	−1.228 (−2.422, −0.189)	−1.884 (−2.863, −1.022)
	0.391	0.354	0.197	0.112
β_{b2}	4.510 (−0.797, 7.570)	8.126 (5.808, 10.720)	3.396 (1.966, 4.409)	3.106 (2.442,3.764)
	0.303	0.036	0.145	0.001
m_c	2.954 (2.600, 3.398)	2.379 (2.178, 2.601)	2.199 (2.051, 2.338)	2.256 (2.161, 2.360)
	0.090	0.120	0.032	0.070
m_b	3.001 (2.528, 3.537)	2.370 (2.078, 2.630)	2.306 (2.147, 2.492)	2.338 (2.237, 2.449)
	0.083	0.105	0.003	0.188

Values in (·) show 95% credible intervals and values shown in italic type in each row are the Geweke statistical test.

probability distribution of the parameters. In Table 2 the values estimated by the Bayesian estimation method are the sample average of parameters, and the values in parentheses refer to 95% credible intervals. All the credible intervals of estimated parameters do not contain zero. Because all the 95% confidence intervals do not contain zero, the estimated values will be significant at the 5% level (Wu & Hamada 2009). As shown in Table 2, as the amount of observation data increases the credible intervals become narrower. The absolute value of the Geweke test statistics shown in italic type are all less than 1.96, so the convergent hypothesis cannot be dismissed at a significance level of 5%.

Figure 1 shows the posterior densities of model parameters for the D_{10000} database. As Figure 1 suggests, the estimation was conducted with high confidence because the shapes of distribution for most of the posterior parameters show normal distribution.

With the estimation results for the competing deterioration-hazard model, it is possible to formulate the survival probability for each type of pipe burst: C-burst or B-burst. Figures 2 and 3 show the survival probability of DCIP for each of the databases to B-burst and C-burst, respectively. The survival probability curves of the Bayesian mean estimates are shown. In addition, Figures 2 and 3 show that as the amount of observation data increases and Bayesian updating is conducted, the survival probability curves approach the survival probability curves obtained from the

original database. As shown in Figures 2 and 3, the survival probability curve obtained from the D_{10000} database shows almost the same path compared to the survival probability curve obtained from the original data set. It means that the Bayesian updating rule improves the efficiency of model estimation and data acquisition.

Figures 2 and 3 also show that the survival probabilities for both C-burst and B-burst decrease over time and that the survival probability for C-burst decreases more rapidly than that for B-burst. In other words, in a ductile cast-iron pipe, bursts in pipe connections (C-bursts) occur at a higher rate than bursts in the pipe body (B-bursts).

Figure 4 compares the survival probabilities of C-burst and B-burst obtained from the Bayesian mean estimates of competing deterioration-hazard model and the conventional Weibull deterioration-hazard model. Figure 4 also shows that the competing deterioration-hazard model predicts a higher survival probability than does the conventional Weibull deterioration-hazard model. It is noteworthy that the reason why the competing deterioration-hazard model predicts a slower deterioration is that this model considers the occurrence of a competing event when the probability of the event of interest is sought.

CONCLUSIONS

A pipe deterioration model is important for the asset management of pipeline systems. Pipe failures caused by
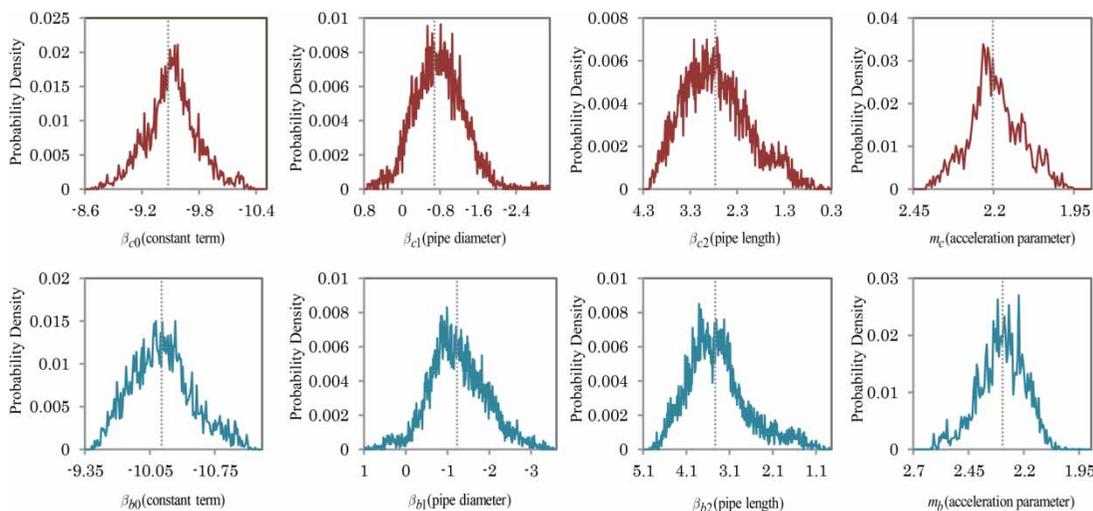

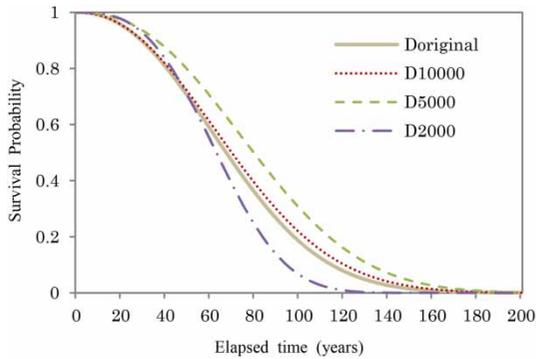**Figure 1** │ Posterior distribution of model parameters of D_{10000}.

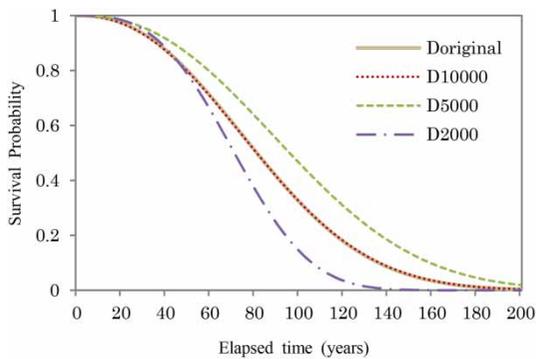**Figure 2** │ Survival probability: C-burst.

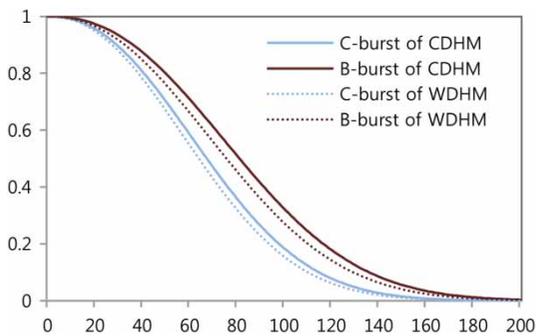**Figure 3** │ Survival probability: B-burst.

Figure 4 │ Survival probability: comparison of the competing deterioration-hazard model (CDHM) and the Weibull deterioration-hazard model (WDHM).

deterioration appear in various forms. Thus, a deterioration forecasting model which considers failure types enables the establishment of an efficient rehabilitation strategy. We have developed a competing deterioration-hazard model that considers competition among several types of burst in pipeline systems and the proposed model allows us to determine

the probability of burst for each type of burst. The competing deterioration-hazard model is estimated using Bayesian estimation method.

The empirical study was carried out by using an inspection data set of a real pipeline system. In the empirical study, because of the absence of detailed substantive knowledge, the competing deterioration-hazard model is estimated with non-informative prior distribution and the Bayesian updating rule is used to improve the precision of estimation. The results show that the more estimation results are updated, the more precise estimation results can be obtained. The estimation results obtained from the D_{10000} database show almost the same results obtained from the original data set. This result indicates that the Bayesian updating rule improves the efficiency of model estimation and data acquisition. In addition, in this study, although we used non-informative prior distribution because of the absence of detailed substantive knowledge, if we can accumulate prior information, the proposed method would be a good way forward.

According to the results of the occurrence probability prediction of C- and B-burst obtained by competing deterioration-hazard model, more care is necessary for the pipe connection because the probability of pipe burst in a pipe connection (C-burst) is higher than that in pipe bodies (B-burst). In addition, the results show that the conventional Weibull deterioration-hazard model, which does not consider competing properties, overestimates pipe burst rates. The bias which arises between the competing deterioration-hazard model and the conventional Weibull deterioration-hazard model comes from the feature of competing hazard model that considers the occurrence of a competing event when the probability of the event of interest is sought. Even though the prediction accuracy of the competing deterioration-hazard model is slightly high in comparison with the conventional Weibull deterioration-hazard model, it is noteworthy that there is the potential for improvement. The proposed competing deterioration-hazard model can be improved further if we could overcome some problems such as left truncated data and high percentage of right censored data, which were not considered in this study. In addition, it is required that much more observed data set and empirical studies are accumulated.

In this study, we classified the pipe burst type into B-burst which occurred in the pipe body and C-burst

which occurred in the pipe connection. As shown in the results, we were able to see that the C-burst and B-burst had different deterioration rates. Because the choice of a maintenance and repair method will depend on the type of burst, the competing deterioration-hazard model enables us to establish an optimum maintenance strategy for the pipeline system. In addition, we believe that our new model can be extended to other items of infrastructure and will contribute to advancing asset management.

Our proposed model has not discussed the following points, which are considered for a future extension of our study:

(1) In this paper, most of the failures that typically affect real pipeline systems (i.e. pipe breaks, leakages, etc.) are disregarded. To establish optimal maintenance strategy, it is important to consider the repairs due to breaks or leakages.

(2) The limited and missing information, left-truncated or survival selection, which are often embedded in observed data, have not been mentioned.

(3) Supposedly, considering competing hazards would be more relevant if more than two competing hazards exist. This could be explored by using synthetic data and considering different amounts of competing failure types.

REFERENCES

Aoki, K., Yamamoto, K. & Kobayashi, K. 2005 The estimation of the hazard model for deterioration predictions. *J. Civil Eng. JSCE* **791** (VI-67), 111–124 (in Japanese).

Clark, R. M., Stafford, C. L. & Goodrich, J. A. 1982 Water distribution systems: A spatial and cost evaluation. *J. Water Resour. Plann. Manage. Div. ASCE* **108** (3), 243–256.

Constantine, G. & Darroch, J. 1995 Predicting underground pipeline failure. *J. Aust. Water Assoc.* **2** (2), 9–10.

Cox, D. R. 1972 Regression models and life tables. *J. R. Stat. Soc.* **34** (B), 187–220.

Eisenbeis, P., Røstum, J. & Le Gat, Y. 1999 Statistical models for assessing the technical state of water networks e some European experiences. In: *Proceedings of the AWWA Annual Conference*, Chicago, USA.

Geweke, J. 1992 Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting* (J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds). Clarendon Press, Oxford, pp. 169–193.

Kim, C. M. & Bai, D. S. 2002 Analyses of accelerated life test data under two failure modes. *Int. J. Reliab. Qual. Saf. Eng.* **9** (2), 111–126.

Kobayashi, K. & Kaito, K. 2012 A mixed prediction model of ground subsidence for civil infrastructures on soft ground. *J. Appl. Math.* **2012**, 1–20.

Kobayashi, K., Do, M. & Han, D. 2010 Estimation of Markovian transition probabilities for pavement deterioration forecasting. *KSCE J. Civ. Eng.* **14** (3), 343–351.

Kobayashi, K., Kaito, K. & Lethanh, N. 2012 A Bayesian estimation method to improve deterioration prediction for infrastructure system with Markov chain model. *Int. J. Archit. Eng. Constr.* **1** (1), 1–13.

Le Gat, Y. & Eisenbeis, P. 2000 Using maintenance records to forecast failures in water networks. *Urban Water* **2** (3), 173–181.

Mailhot, A., Pelletier, G., Noel, J. & Villeneuve, J. 2000 Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: methodology and application. *Water Resour. Res.* **36** (10), 3053–3062.

Mailhot, A., Poulin, A. & Villeneuve, J. P. 2003 Optimal replacement of water pipes. *Water Resour. Res.* **39** (5), 1136–1150.

Marks, H. D. 1985 Predicting urban water distribution maintenance strategies: A case study of New Haven, Connecticut, Coop. Agreement R810558-01-0, Massachusetts Institute of Technology, Sponsored Project 94211, USEPA.

Marvin, K. 1996 *Predicting the Burst Life of an Individual Main.* Rep. No. 114, Urban Water Research Association of Australia, Melbourne.

Nelson, W. 1990 *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses.* Wiley, Hoboken.

Park, S. 2004 Identifying the hazard characteristics pipes in water distribution systems by using the proportional hazards model: 1. Theory. *KSCE J. Civ. Eng.* **8** (6), 663–668.

Pelletier, G., Mailhot, A. & Villeneuve, J. P. 2003 Modeling water pipe breaks – three case studies. *J. Water Resour. Plan. Manage.* **129**, 115–123.

Scheidegger, A., Scholten, L., Maurer, M. & Reichert, P. 2013 Extension of pipe failure models to consider the absence of data from replaced pipes. *Water Res.* **47**, 3696–3705.

Shamir, U. & Howard, C. D. D. 1979 An analytic approach to scheduling pipe replacement. *J. Am. Water Works Assoc.* **71** (5), 248–258.

Shinstine, D. S. 1999 Reliability Analysis Using the Minimum Cut-Set Method for Three Water Distribution Systems. MS Thesis, University of Arizona, Tucson, AZ.

Tanaka, T., Nam, L. T., Kaito, K. & Kobayashi, K. 2010 Probabilistic analysis of underground pipelines for optimal renewal time. *J. Water Supply Res. Technol. AQUA* **59** (6–7), 445–451.

Wu, C. F. J. & Hamada, M. S. 2009 *Experiments: Planning, Analysis, and Optimization,* 2nd edn. Wiley, New York.