A methodology to estimate economically optimal replacement time interval of water distribution pipes

Suwan Park*, Jung Wook Kim*, Agbenowosi Newland** and Hwandon Jun***

*School of Civil & Env. Eng., Pusan National University, Busan, South Korea
(E-mail: wjddnrq@naver.com, swanpark@pusan.ac.kr)

**Booz Allen Hamilton, 8283 Greensboro Drive, McLean VA, 22102 USA
(E-mail: agbenowosi_newland@bah.com)

***BK21 Global Leaders in Construction Engineering, Korea University, Seoul, South Korea (E-mail: hwandonjun@gmail.com)

Abstract In this paper a systematic methodology is developed to estimate economically optimal replacement time interval of a water distribution pipe using the proportional hazards modeling approach. To maximize the full potential of the proportional hazards model (PHM) the pipes in a case study area are redefined so that the pipes have consistent internal and external characteristics along their newly defined lengths. The survival functions of the PHMs are constructed for the six ordered survival time groups of cast iron 6 inch pipes in the area and subsequently used to estimate the intervals of future failure times of a pipe. The recorded and predicted future failure time intervals are used to model the trend of pipe failure using the General Pipe Break Prediction Model (GPBM). The equivalence relationship between the GPBM and the threshold break rate of a pipe is used to estimate the economically optimal replacement time interval of a pipe.

Keywords Optimal replacement; proportional hazards model; threshold break rate; water distribution pipe

Introduction
The PHM (Cox, 1972) represents one of the most advanced statistical models for estimating water main failure patterns. The model is used to assess the hazard rate of an item/assuming that additive changes in the value of a numeric variable, which is called as a covariate, cause corresponding additive changes in the log of the hazard as

\[ h_i(t) = h_0(t) \exp(\beta' x_i) \]  

where \( h_0(t) \) is an arbitrary and unspecified baseline hazard function, \( x \) is a vector of covariates that influence failure, and \( \beta \) is a vector of coefficients of the covariates.

The corresponding survival function is expressed as

\[ S_i(t) = \exp \left( - \int_0^t h_0(\tau) \exp(\beta' x_i) d\tau \right) = [S_0(t)]^{\exp(\beta' x_i)} \]  

where \( S_0(t) \) is the baseline survival function. The PHM enables one to estimate relative effects of the failure-causing factors on survival or failure of pipes by separately modeling the time-dependent ageing of pipes, which is modeled as the baseline hazard function, and the effects of the factors on pipe failure, which is usually assumed as time-independent. Numerous case studies of applications of the PHM can be found in the literature (Andreou et al., 1987; Lei, 1997; Li, 1992; Park, 2004).
To maximize the full potential of the PHM, it is essential that individual pipes are defined by their internal and external characteristics. For this purpose, the 6 inch cast iron (CI) pipes in the case study area water main break database are redefined to obtain the ‘Individual Pipe Failure Database (IPFD)’ so that the redefined individual pipes have consistent internal and external characteristics along their redefined lengths. The survival functions of the PHMs are used to predict future failure time intervals of a pipe in the IPFD and, then, the historical and predicted failure times are used in fitting the GPBM (Park, 2002). The interval of economically optimal replacement time of a pipe is estimated using the equivalence relationship between the threshold break rate (Loganathan et al., 2002) of a pipe and the fitted GPBM. An example of the procedure of estimating the interval of economically optimal replacement time is provided using an individual pipe selected from the IPFD.

Method for defining individual pipes

The criteria that are used to define individual pipe in the case study area are based on the pipe break database, GRID database, and the soil database available for the area. The pipe break database contains information on the installation time, material-joint type, length and times of pipe failure. It also includes locations of break in terms of the proportion of the length in relation to the starting point of a pipe. The GRID database is a proprietary local grid reference system that contains information on internal pressure and population in a geographical grid. The height and width of a grid are 400 and 500 m, respectively. There are four pressure types in the GRID database that are Type I(8.8 ~ 12.3 kg/cm²), II(7.0 ~ 8.8 kg/cm²), III(3.5 ~ 7.0 kg/cm²), and IV(2.1 ~ 3.5 kg/cm²). The soil database contains information on the level of land development along a pipe. Either urban land or non-urban land type is used for the level of land development, and the locations where the level of land development change are contained in the database in terms of the proportion of the length in relation to the starting point of a pipe.

Using the three databases the criteria to redefine individual pipes are determined as the level of land development (LD), pipe material-joint type (CATEGORY), and the geographical location of a grid (GRID). Figure 1 shows the procedure of defining an example pipe No. ‘6194’ that is 2447 m long as individual pipes using the three criteria schematically. Actual computation for this procedure is done using MATLAB.

In Figure 1 the numbers above each of the three bars show the proportion of the location at which the criteria change in relation to the starting location of the pipe, and
the numbers under the bottom line represent the length of each newly defined individual pipe. Since the criteria of a newly defined pipe must not change along its length, this pipe results in 12 newly defined individual pipes as shown in Figure 1. By using this method of defining individual pipes the 6 inch CI pipes in the original water main break database are redefined as a total of 9,642 individual pipes, and this newly defined database is called as the 'Individual Pipe Failure Database (IPFD)'. The total length of the 6 inch CI pipes in the IPFD is about 2,205,060 m and the total number of breaks is 6,607. The IPFD also contains information on the properties of the pipes and the time and locations of the breaks that have occurred from January, 1903 to December, 1997.

**Estimation of the survival functions of the STGs**

To account for the changing risk sets along consecutive pipe failures the methodology of constructing PHMs between pipe breaks (Park, 2004) are used for this study. This methodology extends the capability of the general PHM (Cox, 1972) by which the hazard rate of only one failure is modeled. For this purpose the pipes are assumed to regain a new life after a repair, and the PHMs are constructed for the six ordered survival time groups of the pipes. A survival time group (STG) represents a set of pipes that have the same minimum total number of failures. The PHMs are constructed so that the proportional hazards modeling of STG I results in the PHM for the 1st break and the modeling of STG N results in the PHM between the \((N - 1)\)th and the Nth break, where \(N = \text{II}, \ldots, \text{VI}\). The survival times used for STG I are the elapsed months since installation until the 1st break, and the survival times used for STG N are the elapsed months since the \((N-1)\)th break until the Nth break.

The criteria to define individual pipes can also be used as the covariates in the PHM. The covariates considered in the modeling procedures are the degree of land development (DL), internal pressure type (PT), length of pipe (L), and the population in a grid (C). The degree of land development is considered for a covariate in this paper since detailed and specific properties of soil surrounding a pipe are usually hard to obtain, and the interpretation of the relative effects of detailed soil properties on pipe failure may not be able to find practical applications. The pipe material-joint types are also considered as the covariates for which spun CI rigid joint and spun CI flexible joint type are coded as SR and SF, respectively. SR and SF are defined as ‘1’ for the corresponding material-joint type of a pipe; otherwise they are defined as ‘0’. DL is defined as ‘1’ for the urban land type and ‘0’ for the non-urban land type. L and C are defined as ‘1’ if the length and the population of a grid for a pipe are greater than the mean of the corresponding logarithmic values of the pipes in a STG; otherwise they are defined as ‘0’.

The covariates to be included in the models are determined by analyzing the adjusted survival curves, the proportionality assumption of the hazard rate of the PHM, and the statistical significance of the estimated regression coefficient of each covariate. The final covariates determined to be significant in the PHMs and the corresponding survival functions are SR, SF, DL, L, and C. The internal pipe pressure (PT) was not selected as an important covariate in all of the models. Since the pressure data used in the analyses are for a grid and not for a specific pipe, it is conjectured that more accurate data are needed to assess the effects of the pressure on the pipe failure.

The SAS system is used to obtain the coefficients of the covariates, and the baseline survival function for each STG is obtained using the ‘baseline’ statement in SAS. The survival function for each STG is constructed using the estimated baseline survival
function and the estimated regression coefficients of the covariates as

\[
\text{STG I} : \hat{S}_I(t) = \{\exp(e^{-21.337 \cdot t^{2.246}})\}^{\exp(0.603 \cdot SR + 1.834 \cdot SF + 0.582 \cdot DL + 0.963 \cdot L_i + 0.218 \cdot C_i)}
\] (3)

\[
\text{STG II} : \hat{S}_II(t) = \{\exp(e^{-4.664 \cdot t^{0.686}})\}^{\exp(0.293 \cdot SR + 0.082 \cdot SF + 0.289 \cdot DL + 1.730 \cdot L_i + 0.397 \cdot C_i - 0.487 \cdot DL \cdot C_i)}
\] (4)

\[
\text{STG III} : \hat{S}_III(t) = \{\exp(e^{-3.361 \cdot t^{0.600}})\}^{\exp(0.256 \cdot SR + 0.06 \cdot SF + 0.854 \cdot L_i - 0.169 \cdot C_i)}
\] (5)

\[
\text{STG IV} : \hat{S}_IV(t) = \{\exp(e^{-2.917 \cdot t^{0.609}})\}^{\exp(0.739 \cdot L_i)}
\] (6)

\[
\text{STG V} : \hat{S}_V(t) = \{\exp(e^{-2.417 \cdot t^{0.548}})\}^{\exp(-0.878 \cdot SR - 1.0186 \cdot SF + 0.647 \cdot DL + 0.324 \cdot L_i)}
\] (7)

\[
\text{STG VI} : \hat{S}_VI(t) = \{\exp(e^{-2.642 \cdot t^{0.627}})\}^{\exp(0.441 \cdot L_i)}
\] (8)

The baseline hazard function for each STG can be derived using Equation (9).

\[
\hat{h}_0(t) = -\frac{d}{dt}\ln(\hat{S}_0(t))
\] (9)

The PHM for each STG as in the form of Equation (1) can be constructed accordingly using Equation (9) and the corresponding exponential functions of the covariates in Equations (3)–(9).

**Estimation of the economically optimal replacement time interval**

The methodology of estimating economically optimal replacement time interval of an individual pipe developed in this paper is based on the equivalence relationship between the threshold break rate (Loganathan et al., 2002) and the GPBM (Park, 2002), which is expressed as

\[
\frac{d}{dt}\left((1 - WF) \cdot Bl + Al \cdot t + WF \cdot \exp(Ae \cdot t + Be)\right) = \frac{\ln((1 + R)/(1 + i))}{\ln(1 + C/F)}
\] (10)

where \(t\) is the break times in elapsed years since installation, \(R\) is the interest rate (l/yr), \(i\) is the inflation rate (l/yr), \(C\) is the repair cost, \(F\) is the replacement cost, and \(WF, Bl, Al, Be,\) and \(Ae\) are the coefficients to be estimated. The left and right hand side of Equation (10) represent the time derivative of the GPBM and the threshold break rate (Brkth), respectively.

The threshold break rate represents the break rate of a pipe at which the present worth of the total costs of repair and replacement is at minimum. The GPBM is capable of fitting pipe break trends of exponential, linear, and in between of exponential and linear using the weighting factor, \(WF\). The regression coefficients and the weighting factor of the GPBM are estimated using the least squares method. The economically optimal replacement time as elapsed years from installation of an individual pipe is obtained by solving Equation (10) with respect to time as

\[
t^* - \frac{\ln\left(\frac{Brkth \cdot WF \cdot Al}{Ae \cdot WF}ight)}{Ae} - Be.
\] (11)

The GPBM is constructed using historically available break times of an individual pipe. Since one or two break times can result in a linear model at most, at least three recorded break times are needed to construct a meaningful GPBM. In the mean time, the survival functions derived from the constructed PHMs provide the survival probability of a pipe at a given time. Therefore, the survival functions of the STGs are used to estimate
the future failure time intervals of a pipe. Then, recorded break times and predicted failure time intervals of a pipe are subsequently used to the model the breakage pattern with the GPBM. The economically optimal replacement time interval of a pipe is estimated using Equation (11). An example of this methodology is illustrated below.

The pipe used for an example is selected from the IPFD. It is 446 m long, 6 inch spun CI flexible joint type, installed in January of 1971, and has failed in 237 and 260 months (or 19.75 and 21.667 years) since installation. The replacement cost per unit length of this pipe is US $304 and the repair cost is US $3,120. The interest rate and inflation rate used for the calculation of the threshold break rate are 0.045/year and 0.01/year, respectively. The covariate values of this pipe are 0, 1, 1, 1, and 1 for $SR$, $SF$, $LD$, $L$, and $C$ respectively. Since the pipe has only two recorded breaks, future break times of this pipe are required to be estimated to model the failure pattern. The future 3rd, 4th, 5th, and 6th break times as elapsed months between each pair of successive breaks are obtained by solving Equations (5)–(8) with respect to time as

\[
\hat{t}_3 = \left[ \frac{p^{1/CT_{III}}}{\exp(-3.361)} \right]^{1/0.600} = 70 \text{ and } 81
\]  
(12)

\[
\hat{t}_4 = \left[ \frac{p^{1/CT_{IV}}}{\exp(-2.917)} \right]^{1/0.600} = 44 \text{ and } 49
\]  
(13)

\[
\hat{t}_5 = \left[ \frac{p^{1/CT_{V}}}{\exp(-2.417)} \right]^{1/0.548} = 21 \text{ and } 25
\]  
(14)

\[
\hat{t}_6 = \left[ \frac{p^{1/CT_{VI}}}{\exp(-2.642)} \right]^{1/0.627} = 24 \text{ and } 26
\]  
(15)

where $p$ represents the chosen lower or upper bound survival probability, which is selected as 0.48 or 0.52 for the example pipe, and $CT$s are the covariate terms of the PHM for each STG shown in Equations (5)–(8). Equations (12) and (13) represent the elapsed months from the 2nd to the 3rd break, and from the 3rd to the 4th break, respectively, using 0.48 and 0.52 for $p$. Equations (14) and (15) are defined in a similar manner. Therefore, expressing as elapsed years since installation the predicted 3rd, 4th, 5th, and 6th break time intervals are estimated as [27.511,28.441], [31.128, 32.482], [32.943, 34.581], and [34.898, 36.775], respectively.

A GPBM is constructed using the recorded break times, 19.75 and 21.667 years, and the lower bounds of the future break time intervals that are 27.511, 31.128, 32.943, and 34.898 years. Another GPBM is also constructed using the recorded break times and the upper bounds of the future break time intervals that are 28.441, 32.482, 34.581, and 36.775 years. The curves 'A' and 'B' in Figure 2 represent the fitted GPBM of the two cases and are expressed mathematically as

\[
A(t) = (1 - 0.71)(0.2975t - 4.8256) + 0.71\exp(0.1043t - 1.8218)
\]  
(16)

\[
B(t) = (1 - 0.72)(0.2628t - 4.1077) + 0.72\exp(0.0919t - 1.5653)
\]  
(17)

Using the fitted GPBM, $A(t)$ and Equation (11) the lower bound of the optimal replacement time in years from installation is obtained as \[
\hat{t}_A = \ln\left(\frac{1.5012 + 0.71\cdot0.2975 - 0.2975}{0.1043\cdot0.71} \right) + 1.8218 = 45.75
\]  
(18)
In addition, the upper bound of the optimal replacement time in years from installation is obtained using the fitted GPBM, \( B(t) \) and Equation (11) as

\[
\begin{align*}
\ln t^* &= \ln \left( \frac{1.5012 + 0.72 \cdot 0.2628 - 0.2628}{0.0919} \right) + 1.5653 = 50.45
\end{align*}
\]

Therefore, the economically optimal replacement time interval of the pipe is estimated as from year 2017 to 2021.

**Conclusions**

By analyzing the constructed survival functions of the PHMs, it was found that the failure times of all of the six STGs follow the Weibull distribution. Using the estimated baseline survival functions the baseline median survival time of the STG I, II, III, IV, V and VI are estimated as 639, 527, 147, 66, 42, and 38 months, respectively, which can be interpreted as that the time interval between consecutive breaks generally decreases as the number of break increases. In addition, it is found that the hazard rate for the 1st break increased as time progresses and the hazard rate for the 2nd and more break showed decreasing hazard rates. The reasons for this difference in the patterns of hazard rates are conjectured as such that the hazard rate of the 1st break increases due to deterioration, and that the hazard rate for the 2nd and more breaks are relatively high in the early periods due to imperfect repair.

Regarding the 1st failure the hazard rate of spun CI rigid joint type pipes was estimated as about 1.8 times of pit CI pipes, and in the case of spun cast flexible joint type pipes it was estimated as about 6.3 times of pit CI pipes given that all other values of the covariates remain constant. For the 2nd or more failures the relative effects of pipe material-joint types on failure was not conclusive due to the 95% confidence interval of the hazard ratios. The degree of land development affects pipe failure for STGs I, II, and V, and the average hazard ratio was obtained as 1.8 which can be interpreted as the hazard of failure for urban-land type is about 1.8 times more than that of non-urban-land type.

The lower and upper bound survival probabilities used in estimating the future break time intervals for the example pipe is chosen arbitrarily to obtain a reasonable range of optimal replacement time interval of 5 years. Since the failure patterns of water distribution pipes vary considerably, it is conjectured that a different range of the lower and upper bound survival probabilities should be used to obtain a reasonable range of optimal replacement time interval.
References


