Proceedings of the Institution of Civil Engineers
Water Management 161
August 2008 Issue WM4
Pages 207–214
doi: 10.1680/wama.2008.161.4.207

Paper 700067
Received 13/10/2007
Accepted 06/08/2008

Keywords:
failures/pipes & pipelines/risk & probability analysis/statistical analysis/water supply

Azam Dehghan
PhD Student, Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Victoria, Australia

Kerry McManus
Associate Professor, Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Victoria, Australia

Emad Gad
Associate Professor, Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Victoria, Australia

Institution of Civil Engineers

Statistical analysis of structural failures of water pipes

A. Dehghan BEng, K. J. McManus PhD and E. F. Gad PhD

Statistical analysis and prediction of failure rates of water distribution pipes are usually performed using parametric lifetime models. In this paper, a new probabilistic measure for the failure rate, called the 'likelihood of number of failures', is defined and formulated for cases where the pipe lifetimes follow parametric models. The resulting theoretical failure rates are time-invariant and, therefore, the parametric models would be useful only if the failure rates of water distribution pipes are stationary random processes. This paper then examines the stationarity of pipe failure rates in practice. For the water pipes in the western district of Melbourne (Australia), the failure rates are empirically calculated using a 4-year failure history, and it is observed that the distribution of empirical failure rates varies with time. In order to explain these variations, the pattern of rainfall in the region is compared with the pattern of failure rate variations, and in 70% of the times the two patterns are observed to be consistent. Two approaches are proposed to tackle the time-varying nature of pipe failure rate processes: regular updating of the parameters of lifetime models or developing a non-parametric technique for modelling of pipe failure rates.

NOTATION

CICL	cast iron, cement-lined
ENOF(nT)	expected number of failures during the nth time interval
$f_{\mathrm{TFF}}(t)$	probability distribution function of the time to the first failure (TFF)
IFT	inter-failure time
LNF	likelihood of number of failures
NOF$_k$(nT)	event of occurrence of k failures during the nth time interval
nT	an arbitrary time interval (most recent time interval in a failure prediction application)
$P_k(nT)$	probability of occurrence of NOF$_k$(nT)—a LNF value
$P_k^{\mathrm{EMP}}(S_{\mathrm{i}})$	empirical estimate of the LNF value P_k during the time period S_{i}
pdf	probability distribution function (for continuous random variables)
pmf	probability mass function (for discrete random variables)
S_{i}	time period during which the LNF values are empirically estimated
TFF	time to the first failure occurring after the time $(n-1)T$
t_{f}	time passed from most recent failure
α	scale parameter of a Weibull distribution
β	probability value associated with a confidence interval
δ	half-width of a confidence interval
η	shape parameter of a Weibull distribution

1. INTRODUCTION

During recent decades, substantial research effort has been conducted on degradation analysis of water pipelines. Such an analysis is generally performed by modelling past behaviour of pipe breakages and projecting it into the future. Different types of modelling techniques have been developed to analyse the pipe breakages, their reliability and remaining life.[1–6]

One type of failure analysis is descriptive analysis; this consists of calculating descriptive statistics to provide insight regarding breakage patterns and trends. There are few case studies of this kind of analysis reported in the literature. Descriptive analysis can only be performed in cities or areas with comprehensive databases on the characteristics of their pipes and on pipe breakages. Some cities often cited for participating in such studies are Winnipeg, Manitoba, Canada,[7–9] New York,[2,10,11] suburban Paris and Bordeaux, France,[12] three municipalities of Quebec, Chicoutimi, Gatineau and Saint-Georges[13] and Boston.[14]

Since data of this kind are not available in most of water distribution systems, statistical analysis is used to predict the failure behaviour of pipelines. In statistical analysis, a commonly used approach in failure/reliability analysis of deteriorating water pipes is the use of lifetime models such as Weibull, lognormal[3,6] or Herz[15] distributions (e.g. Deb[16]).

In this context, survival analysis has been commonly applied in order to develop parametric lifetime models. The analysis of survival data is a traditional statistical theme. Cox,[17] however, introduced the proportional hazards model (PHM) in order to estimate the effects of different covariates on the time to failure of a system. Kaara[18] and Andreou[19] introduced the use of a proportional hazards model for analysis of the failures in water distribution networks and many researchers applied hazard models completely or partially to model the failure process (e.g. Le Gat[20]). The class of failure analysis models that uses the hazard function is semi-parametric. The reason is that its hazard function is the product of an unspecified baseline hazard function and a parametric function relating the hazard function and the covariates.

Dehghan and McManus[21] have proposed a neural network model for survival analysis of water pipes. Like Weibull and lognormal models, the neural model is also parametric, and its parameters (the synapsis weights) are trained using a failure history.

All parametric lifetime models share the underlying assumption that the random processes of failures in water mains are stationary random processes. A random process is an ensemble of consecutive random variables that corresponds with possible outcomes of a random event. For example, the number of pipe failures occurring during one month is a random variable, and the ensemble of such numbers corresponding with consecutive months is a random process.

In engineering applications, usually a random process is referred to as stationary if the mean and variance of the process are time-invariant; otherwise it is referred to as a non-stationary process.[22,23] More precisely, by definition, a random process is wide-sense-stationary if its mean and second-order statistical properties (its correlation function) are time-invariant. If the distribution functions of all the random variables that constitute the random process are identical, then the random process is referred to as strict-sense stationary.[23]

Using a lifetime model with time-invariant parameters for the water pipes implicitly assumes that the random process of time-intervals between consecutive failures of the pipes is a stationary process in strict sense. The current paper aims to illustrate that the random processes of water pipe failures (failure rates or inter-failure times) are non-stationary random processes, and demonstrates the deficiencies of parametric techniques for the analysis of such failure processes through mathematical and empirical analyses.

Demonstration of non-stationary nature of pipe failure processes can be performed using any of the various failure-related quantities that have been analysed in a probabilistic modelling context in the literature. Two more popular examples of such quantities are inter-failure times (IFTs) and failure rates. Since the authors of this paper have chosen to focus on failure rates (as this was the main quantity required by the industry partner, City West Water, Melbourne, Victoria, Australia, to be analysed and predicted in this research project), a new probability-based equivalent definition of failure rate is introduced in section 2 and its characteristics are also discussed. Theoretical failure rates for general parametric models and two-parameter Weibull models are derived in section 3. This is followed by explanation of empirical calculation of the probabilistic failure rates using a failure database, as presented in section 4. Comparative results for theoretical and empirical failure rates in a case study are presented in section 5. Section 6 concludes the paper.

2. A PROBABILISTIC DEFINITION OF FAILURE RATES

Existing failure analysis methods for water pipes usually quantify the past behaviour of pipe failures in terms of either failure rates[24] or IFTs and project them into the future. In order to monitor the pattern of pipe failures and demonstrate that the failure process is non-stationary, the current paper suggests a new set of measures to be studied instead of failure rates or inter-failure times, namely the probabilities of certain numbers of failures occurring during specific time intervals.

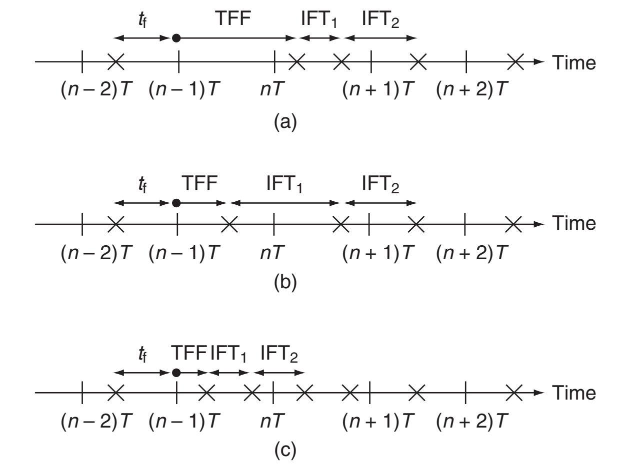

Fig. 1. Demonstration of inter-failure times in three instances of occurrence of the events: (a) $NOF_0(nT)$; (b) $NOF_1(nT)$; and (c) $NOF_2(nT)$

The failure history is divided into equal time intervals. The length of time intervals, denoted by T in this paper, should be chosen carefully. Very long time intervals result in a rough analysis in which variations of the failure process during the long time intervals are neglected. On the other hand, the length of time intervals should be long enough to include a fair number of failures on average. The nth time interval is the interval within $[(n-1)T, nT]$.

The event of occurrence of k failures during the nth time interval is denoted by $NOF_k(nT)$. There is a direct relationship between the IFTs and the number of failures occurring during each time interval. In order to show this relationship, three instances of occurrence of the events $NOF_0(nT)$, $NOF_1(nT)$ and $NOF_2(nT)$ are illustrated in Fig. 1, where TFF denotes the time to the first failure occurring after the time $(n-1)T$, and the next consecutive IFTs are denoted by IFT_1 and IFT_2, respectively. The time passed from most recent failure is also denoted by t_f.

As Fig. 1 shows, the event of occurrence of no failure during the nth time interval, $NOF_0(nT)$, is equivalent to

$$\mathrm{NOF}_0(nT) \equiv \{\mathrm{TFF} > T\} \tag{1}$$

and similarly, the events $NOF_1(nT)$ and $NOF_2(nT)$ are equivalent to

$$\mathrm{NOF}_1(nT) \equiv \{(\mathrm{TFF} \leqslant T) \wedge (\mathrm{TFF} + \mathrm{IFT}_1 > T)\} \tag{2}$$

$$\mathrm{NOF}_2(nT) \equiv \{(\mathrm{TFF} \leqslant T) \wedge (\mathrm{TFF} + \mathrm{IFT}_1 \leqslant T) \wedge (\mathrm{TFF} + \mathrm{IFT}_1 + \mathrm{IFT}_2 > T)\} \tag{3}$$

The probability of occurrence of k failures during the nth time interval is denoted by $P_k(nT)$ and is equal to $\Pr\{\mathrm{NOF}_k(nT)\}$. The variable n implies possible variations of such probabilities with time, which will be discussed further in this paper.

Definition. Each of the probability values in the set $\{P_k(nT)|k = 0, 1, \ldots, M\}$ is called a 'likelihood of number of failures' (LNF value), where M is the maximum number of failures that can occur within a time interval.

Calculation of the theoretical and empirical LNF values will be discussed in the next sections of this paper. It is important to note that in the probabilistic approach to define and evaluate the water pipe failure rates as introduced in this paper, unlike the deterministic approach, the analysis does not merely return a certain number of failures (or failure rate as commonly accepted in infrastructure system analysis context). Instead, the focus is on failure probabilities, resulting in more accurate and reliable analysis and failure prediction (compared with traditional approaches) for developing the maintenance strategies, as explained in the following paragraphs.

Having the LNF values for the nth time interval, the most likely expected number of failures, denoted by the symbol ENOF(nT), in that time interval can be directly calculated as the statistical mean of the number of failures given by

$$\mathrm{ENOF}(nT) = \sum_{k=0}^{M} kP_k(nT) \tag{4}$$

This value is equivalent to the failure rate as commonly computed in infrastructure system analysis. By using the LNF values, however, a confidence interval can also be calculated for the above failure rate. A confidence interval quantifies the existing uncertainty in the calculated failure rate, and it is particularly useful if the future failure rates are calculated by equation (4). For example, the statement 'with a probability of 90%, 18–22 failures will occur in each month in future' is more meaningful and more useful for planning, compared with the statement '20 failures will occur monthly'.

Assume that using the LNF values, a measure for the expected number of failures, ENOF, is calculated using equation (4). The interval [ENOF $-\ \delta$, ENOF $+\ \delta$] is the β-confidence interval corresponding to this failure rate, if

$$\Pr(x \in [\mathrm{ENOF} - \delta, \mathrm{ENOF} + \delta]) = \beta \tag{5}$$

Calculation of the half-width of the confidence interval, δ, is straightforward by histogram analysis of the LNF values. The LNF values for immediate right and left neighbours of ENOF are added to the LNF value for the ENOF value. If the result is equal to β, the interval between these two neighbourhoods is the β-confidence interval. Otherwise, this interval should be extended (symmetric) until the area under the LNF curve equals β.

It is emphasised that possibility of computation of confidence intervals is not exclusive to LNF values. Indeed, wherever a quantity is modelled and estimated in a probabilistic modelling context, a confidence interval can be calculated for it and this is an important merit of probabilistic modelling techniques (over deterministic ones) that does not appear to have been commonly discussed in the literature.

3. DERIVATION OF THEORETICAL LNF VALUES FROM LIFETIME MODELS

Lifetime distribution models form the core components of probabilistic approaches used in the analysis of water pipe failures. Such models usually contain parametric functions with constant coefficients that are typically estimated using linear regression on a given history of failure records. If such a model is available, it provides a probability density function for IFTs and the lifetime, which is the time to the first failure (TFF), denoted by $f_{\mathrm{IFT}}(t)$ and $f_{\mathrm{TFF}}(t)$, respectively. There is a direct relationship between these two density functions, as explained below.

The sum t_f + TFF is an IFT and therefore it is a random variable with the IFT density function. Since in probabilistic lifetime modelling, the consecutive failure times are assumed independent from each other, the random variables t_f and TFF are independent and the density function of their sum equals the convolution of their individual density functions

$$f_{\mathrm{IFT}}(t) = f_{t_f}(t)^* f_{\mathrm{TFF}}(t) = \int_0^t f_{t_f}(\tau) f_{\mathrm{TFF}}(t-\tau)\,\mathrm{d}\tau \tag{6}$$

On the other hand, since $T - t_f$ is also a time to the first failure, the density of t_f can be expressed as $f_{t_f}(t) = f_{\mathrm{TFF}}(T-t)$ and, by substituting into equation (6), the following integral equation for the density functions $f_{\mathrm{TFF}}(t)$ and $f_{\mathrm{IFT}}(t)$ is derived

$$f_{\mathrm{IFT}}(t) = \int_0^t f_{\mathrm{TFF}}(T-\tau) f_{\mathrm{TFF}}(t-\tau)\,\mathrm{d}\tau \tag{7}$$

From equation (1), the LNF value $P_0(nT)$ is given by

$$P_0(nT) = \Pr\{\mathrm{NOF}_0(nT)\} = \Pr\{\mathrm{TFF} > T\} = \int_T^\infty f_{\mathrm{TFF}}(t)\,\mathrm{d}t \tag{8}$$

The event of occurrence of only one failure during the nth time interval, $\mathrm{NOF}_1(nT)$, is expressed in equation (2) and the LNF value $P_1(nT)$ is derived as follows

$$\begin{aligned} P_1(nT) &= \Pr\{\mathrm{NOF}_1(nT)\} \\ &= \Pr\{(\mathrm{TFF} \leqslant T) \wedge (\mathrm{TFF} + \mathrm{IFT}_1 > T)\} \\ &= \int_0^T \int_{T-t_1}^\infty f_{\mathrm{TFF}}(t_1) f_{\mathrm{IFT}}(t_2)\,\mathrm{d}t_2\,\mathrm{d}t_1 \end{aligned} \tag{9}$$

Similarly, from equation (3), the LNF value $P_2(nT)$ is derived as

$$\begin{aligned} P_2(nT) &= \Pr\{\mathrm{NOF}_2(nT)\} \\ &= \Pr\{(\mathrm{TFF} \leqslant T) \wedge (\mathrm{TFF} + \mathrm{IFT}_1 \leqslant T) \\ &\qquad \wedge (\mathrm{TFF} + \mathrm{IFT}_1 + \mathrm{IFT}_2 > T)\} \\ &= \int_0^T \int_0^{T-t_1} \int_{T-t_1-t_2}^\infty f_{\mathrm{TFF}}(t_1) f_{\mathrm{IFT}}(t_2) f_{\mathrm{IFT}}(t_3)\,\mathrm{d}t_3\,\mathrm{d}t_2\,\mathrm{d}t_1 \end{aligned} \tag{10}$$

The above derivation can be generalised to every k number of failures, for which the probability $P_k(nT)$ is given by

$$P_k(nT) = \Pr\{(\text{TFF} \leqslant T) \wedge (\text{TFF} + \text{IFT}_1 \leqslant T) \wedge \cdots \wedge \left(\text{TFF} + \sum_{i=1}^{k-1} \text{IFT}_i \leqslant T\right) \wedge \left(\text{TFF} + \sum_{i=1}^{k} \text{IFT}_i > T\right)\} = \int_0^T \int_0^{T-t_1} \cdots \int^{T-\sum_{i=1}^{k-1} t_i} \int_{T-\sum_{i=1}^{k} t_i}^{\infty} \times f_{\text{TFF}}(t_1) f_{\text{IFT}}(t_2) \cdots f_{\text{IFT}}(t_{k+1})\, \mathrm{d}t_{k+1} \cdots \mathrm{d}t_1 \tag{11}$$

Equations (8) to (11) show that the LNF values are time-invariant (independent of the absolute time nT) as they merely depend on the number of failures k and the time-invariant joint probability density functions of the IFTs. In order to clarify this point, the LNF values $P_0(nT)$ and $P_1(nT)$ are derived for a two-parameter Weibull lifetime model, which has been repeatedly applied for failure analysis of many types of units, with the following probability density function[25]

$$f_{\text{IFT}}(t) = \frac{\eta}{T\alpha}\left(\frac{t/T}{\alpha}\right)^{\eta-1} \mathrm{e}^{-[(t/T)/\alpha]^{\eta}} \tag{12}$$

where η and α are the shape and scale parameters. The following formula for $P_k(nT)$ is derived

$$P_k(nT) = \int_0^T \int_0^{T-t_1} \cdots \int_0^{T-\sum_{i=1}^{k-1} t_i} \times \int_{T-\sum_{i=1}^{k} t_i}^{\infty} \left(\frac{\eta}{T\alpha}\right)^{k+1} f_{\text{TFF}}(t_1) \left(\frac{t_2 \cdots t_{k+1}}{T_k \alpha^k}\right)^{\eta-1} \times \mathrm{e}^{-\left(\frac{t_2/T}{\alpha}\right)^{\eta} - \cdots - \left(\frac{t_{k+1}/T}{\alpha}\right)^{\eta}} \mathrm{d}t_{k+1} \cdots \mathrm{d}t_1 \tag{13}$$

The above LNF values are time-invariant and this property is not specific to the Weibull model. Indeed, as long as the distribution has constant parameters that are not updated with time, the derived LNF values are independent of time (they do not depend on either n or T) and merely depend on k. When time-invariant lifetime distributions such as Weibull distribution in equation (12) are utilised to model the failure process, the random process formed by the consecutive IFTs is implicitly assumed to be strict-sense stationary (with time-invariant probability density function). The derivations made in this section show that in such cases, failure rate (number of failures occurring during a specified time interval) is a strict-sense stationary process, too. More precisely, it would be a discrete random process with time-invariant probability mass function (pmf) and the LNF values defined in this paper would be its pmf.

4. EMPIRICAL ESTIMATION OF LIKELIHOODS OF NUMBER OF FAILURES

Having a dataset including water pipe failures over a long period, the LNF values can be empirically estimated using a histogram technique. The failure history is divided into some time units referred to as time periods. The duration of the time periods should be short enough to assume that LNF values remain almost constant during the time intervals within each time period. On the other hand, the time periods should be long enough to provide reasonable empirical estimates for LNF values during each period, by using the histogram technique. For instance, in the analysis presented in the current paper, each time period is 3 months long and each time interval is 1 day long. If a time period S_i includes the intervals within n_1T and n_2T, that is $S_\text{i} = [n_1T, n_2T]$, then the following empirical LNF values are given by the histogram technique

$$P_k^{\text{EMP}}(n_1T) = \cdots = P_k^{\text{EMP}}(n_2T) = P_k^{\text{EMP}}(S_\text{i}) = \frac{\text{Number of NOF}_k \text{ events occurred during } S_\text{i}}{n_2 - n_1} \tag{14}$$

where P_k^{EMP} is the empirical estimate of P_k.

For each time period S_i, the expected number of failures denoted by ENOF(S_i) is the statistical mean of the number of failures occurred during a time interval within S_i

$$\text{ENOF}(S_\text{i}) = \sum_{k=0}^{M} \{kP_k^{\text{EMP}}(S_\text{i})\} = \frac{\text{Number of failures occurred during } S_\text{i}}{n_2 - n_1} \tag{15}$$

5. CASE STUDY

This study uses a failure history of water pipes that distribute the drinking water in the western suburbs of Melbourne. These pipes belong to City West Water (CWW), which is a water retailer company. Melbourne Water in its *Water Main Renewal Study*[26] noted that the western region of Melbourne was experiencing a disproportionately high rate of failures. It reported a burst rate three times that of Melbourne's other two water supply systems. For example, between 1972 and 1990 the annual average water main failure rate throughout CWW's current licence area was approximately 1 failure/km/annum, while this value was 0·3–0·5 failures/km/annum for the other two regions.[27] In 1995/96 water industry benchmarks revealed that CWW had the highest water main break rate in Australia.[26]

Given this background, since 1999, CWW recognised the need for the failure analysis of its water mains. Accordingly, an investigation for cast iron (CI) pipes and associated failures with a view to formulating a strategy for cost-effective asset management in both the short and longer term was conducted.[28] That study resulted in some parametric models for failure prediction of those water mains. The reason for choosing CI was that CI pipes comprise more than half of CWW's water mains and contribute disproportionately to the number of failures and customer service key performance indicators (KPIs).[28]

This study is conducted on a dataset, consisting of breakages of CI pipes that have occurred during 1997–2000 in the CWW and no records of previous pipe breaks were available for this study. This failure database includes 6381 pipe breakage records. Each breakage record contains the following fields: pipe identification (ID), construction date of the pipe, pipe diameter (mm), pipe material, pipe length, failure date, type of failure and the pipe location in AMG (Northing and Easting)

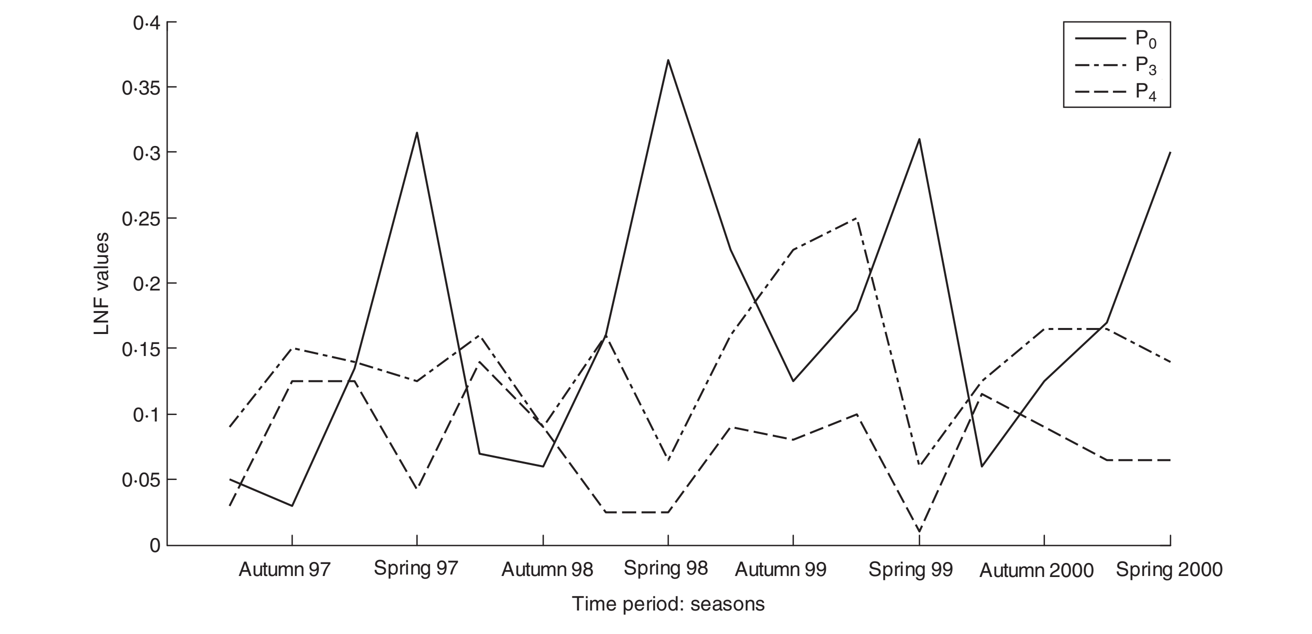

Fig. 2. Empirical LNF values P_0, P_3, P_4 for the 16 consecutive seasons during 1997–2000

coordinates. The two types of pipes in this database are CI pipes and CI cement-lined (CICL) pipes. The pipes in the failure database have a range of ages, with construction dates varying between 1857 and 1975 and with a range of diameters (80 mm, 100 mm, 125 mm, 150 mm and 175 mm). For each breakage record, the postcode of the failed pipe is found by matching the AMG coordinates of the pipe location with the postcode map provided by Australia Post.

In order to consider the effect of the material, size and geographical location of the pipes in the lifetime models, a separate subset of the failures in the history are analysed in this case study: the breakages of the CICL pipes with diameter of 100 mm, located in an area covered by a single postcode 3021: 1450 failures have occurred for 100 mm CICL pipes in the postcode area 3021 during 1997–2000. Simulations have shown that this number of failure records over the course of 4 years (16 seasons) is sufficient for the purpose of the analysis presented in this paper. In this analysis, each time period is 3 months (one season) long and failures are counted on a day-by-day basis, that is each time interval is 1 day long. For each season, a set of LNF values are empirically calculated using equation (14).

Figure 2 shows the empirical values of $P_k(S_i)$ for $k = 0$, 3, 4 and their variations over 16 consecutive seasons (4 years: 1997–2000). Such a variation is not predicted by existing lifetime models. The main reason is that in contrast to what is assumed by probabilistic lifetime models, the random process of water pipe failures is non-stationary. This is mainly because of environmental factors that affect the rate of failures and IFTs are not incorporated into the lifetime models employed for failure analysis of water pipes. In order to clarify this argument, a two-parameter Weibull model is fitted to the water pipe failure records. Having the parameters η and α, the IFT density function in equation (7) is substituted using the Weibull density function given in equation (12), and the TFF density function is numerically calculated. Then, numerical calculation of the integrals in equation (13) results in the following constant LNF values

$$P_0 = 0{\cdot}2843;\ P_3 = 0{\cdot}1361;\ P_4 = 0{\cdot}0629 \tag{16}$$

In addition to the size, material and geographical location of the pipes, other factors affect the pipe failure process. Some examples include construction details, external and internal loads, and corrosion. In most breakages with clear mechanical causes, corrosion has an accelerating role by weakening the fabric of the pipe. Although such factors are not considered in this case study, during the 4-year extent of the failure history, most of these factors could be assumed to vary slightly from one season to the next. One factor that is not steady over the consecutive seasons is soil movement. This is a particularly critical factor in a region with expansive soil that is subject to swelling and shrinkage that varies in proportion to the amount of moisture present in the soil. As water is initially introduced into the soil (by rainfall or watering), it expands, and after drying out, it contracts, often leaving small fissures or cracks. Excessive drying and wetting of the soil progressively deteriorates the structures over years and the resulting soil movement can exert enough pressure (as large as 718·2 KPa[29]) to crack pipelines.

Substantial areas of the state of Victoria are covered by expansive clay soils. The expansive soil map of Victoria (Mann[30]) shows the expansion of low to extremely expansive soils. The region under the present study is located in the area with expansive soil and pipe fractures are likely to occur over time owing to soil movements that mainly depend on rainfall. While the soil type is almost time-invariant, rainfall is a non-stationary process. Rainfall profile can therefore be considered as an influencing factor contributing to the non-stationary nature of the failure process.

Figure 3(a) shows the rainfall records for the 16 seasons during 1997–2000. The variations of rainfall are significant through the seasons, as the average rainfall is 110 mm and the standard deviation is 42 mm. Dry seasons are expected to be associated with high rates of breakages owing to soil shrinkage. In addition, large

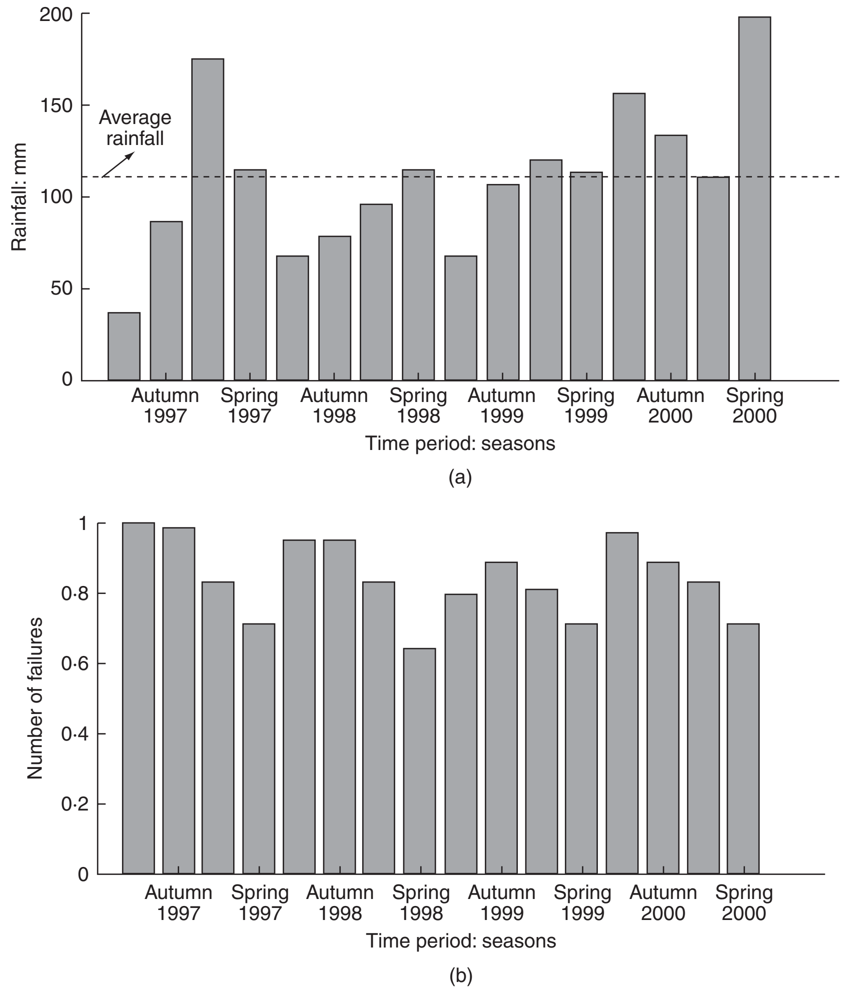

Fig. 3. Comparison of rainfall in each season with its corresponding empirical average number of failures: (a) rainfall; (b) ENOF values

Season (S_i)	ENOF	σ_{ENOF}	Season (S_i)	ENOF	σ_{ENOF}
Summer 1997	1·00	0·0223	Summer 1999	0·78	0·0327
Autumn 1997	0·98	0·0457	Autumn 1999	0·88	0·0009
Winter 1997	0·81	0·0190	Winter 1999	0·80	0·0273
Spring 1997	0·71	0·0149	Spring 1999	0·69	0·0131
Summer 1998	0·94	0·0398	Summer 2000	0·93	0·0387
Autumn 1998	0·94	0·0247	Autumn 2000	0·86	0·0216
Winter 1998	0·81	0·0082	Winter 2000	0·79	0·0280
Spring 1998	0·62	0·0208	Spring 2000	0·68	0·0147

Table 1. The empirical expected number of failures and their standard deviations for 16 consecutive seasons during 1997–2000

fluctuations in soil moisture are considered as the main source of soil movement resulting in pipe breakages. In order to examine the consistency of the variations in LNF values with rainfall variations, the empirical failure rates (ENOF values) are calculated using equation (15) and plotted in Fig. 3(b). The standard deviation of ENOF values has also been calculated using the following equation

17

$$\sigma_{ENOF}(S_i) = \sqrt{\sum_{k=0}^{M} (k - \text{ENOF})^2 P_k^{EMP}(S_i)}$$

The numerical results of ENOF values and their standard deviations are presented in Table 1. The small standard deviation of ENOF values in each season shows how accurately the expected number of failures is calculated using equation (15).

In Fig. 3, it is observed that the directions of variations (increasing or decreasing) of 11 failure rates (out of a total of 16 failure rates) are in contrast to rainfall variations. Thus, in the present case study, the rainfall observations explain the failure pattern variations in 11 out of the 16 (70%) of seasons. For the remaining five seasons (outliers), other factors such as excessive loading or preventative maintenance may explain the inconsistency of failure pattern variations with rainfall variations.

In order to highlight the correlation between the rainfall data and number of failures, in Fig. 4, the rainfall data are plotted against the empirical (expected) number of failures or ENOF values. It is observed that when the rainfall is significantly higher or lower than its average value (about 110 mm), there is a corresponding increase in the number of pipe failures.

In order to clarify this point, the magnitude of deviation of rainfalls from their average is plotted against the ENOF values in Fig. 5 and a regression line is fitted to the points (excluding the five outliers). The outliers are recognised by a robust estimation technique called least median estimator (LMS).[31] This method finds the optimum linear fit to the data (excluding the detected outlier samples) and automatically results in an inlier–outlier dichotomy. The outliers in this case study are mainly associated with random effects and extreme climate variations. Although the outliers are not considered in this analysis, they also contribute to the random time variations of the failure process and its non-stationarity. Indeed, their existence also demonstrates the deficiency of parametric (probabilistic) models developed for water pipe failures in the literature.

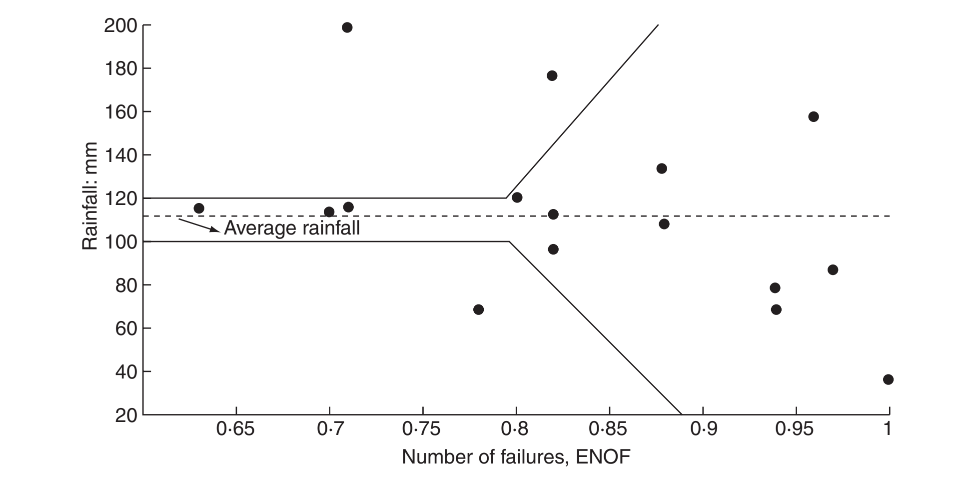

Fig. 4. Number of failures of each season in 1997–2000 for the 100 mm CICL pipes of the case study, and the corresponding rainfall records

experience a higher than average number of failures. Indeed, this fact explains the outliers in Fig. 4, where at certain periods there are high rainfalls but no appreciable increase in the number of failures.

6. CONCLUSIONS

During a period of several decades, water distribution pipes inevitably suffer from soil–pipe interactions, material ageing, design internal and external loads, and extreme loads. The induced damage may accumulate and the performance degradation owing to the above factors eventually reduces the capacity of the pipes to resist, resulting in structural failure.

The current paper introduces a probabilistic definition for the failure rate, called LNF to be applied for the analysis of the failure process of water pipes. The LNF theoretical values are derived from general lifetime models, and those values are empirically calculated using a pipe failure database. It is observed that the LNF values derived from classical lifetime models are time-invariant, while their empirical values vary from one season to the next. This demonstrates that the failure processes of water pipelines are non-stationary random processes while in the existing lifetime models utilised for the analysis of pipe failures, stationary random processes are assumed.

In order to investigate the sources of the non-stationary nature of the pipe failure processes, variations of the empirical failure rates (statistical mean of the failures occurred during each day) are studied in comparison with variations of rainfall in the area. Comparison of the concurrent plots of rainfall and the empirical failure rates shows that, most of the time, variations of failure rates can be explained with variations of rainfall.

One approach to tackle the time-varying nature of pipe failure processes is regularly to update the parameters of lifetime models. While this may require demanding computational updates and costly expert staff, the resulting predictions would be more realistic and reliable. In order to reduce the burden on water distribution authorities, research is currently underway to develop non-parametric approaches for efficient analysis of the non-stationary pipe failure processes. Such techniques should be able to handle and automatically update dynamic models of current failure processes.

The correlation coefficient of regression is 0·83, which is sufficiently large to validate the present authors' assumption of an almost linear relationship between the failure rates and rainfall (disregarding the extreme climate variations, which cause the outlier points). A similar trend has been observed in the data of other classes of pipes. This trend is expected as a significant range in soil moisture (owing to change in rainfall) would lead to swelling and shrinkage of reactive soils. This in turn would lead to movement, distortion and subsequent failure of pipes.

It is important to note that not only the amount of rainfall, but also (and more importantly) the rate of change of rainfall and soil moisture affect the failure rates. More precisely, if a very dry soil (owing to below average rainfall) receives a high amount of rainfall (well above average), the resulting soil movement would be quite large even though the total rainfall might be at about average. This high rate of change in rainfall can be observed in Fig. 3(a) by comparing the rainfalls for summer and winter of 1997. It should also be noted that if a high rainfall occurs while the soil is fully saturated from earlier events, it is unlikely that further soil movement will take place and, hence, pipes would not

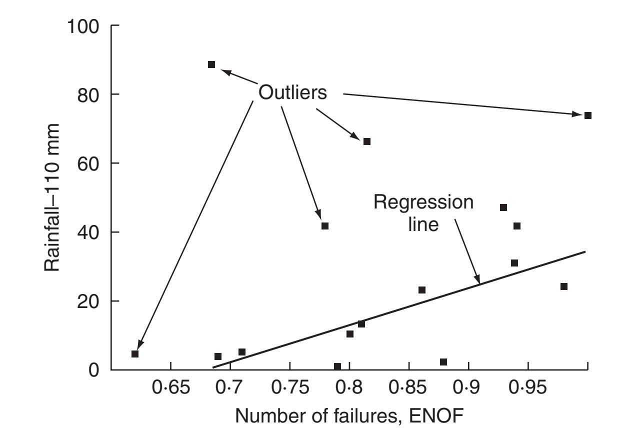

Fig. 5. Deviation of rainfalls from their average, plotted against the corresponding ENOF values. A regression line demonstrates the nearly linear correlation between the failure rates and rainfall deviations

ACKNOWLEDGEMENT

This research was supported by City West Water Limited, Victoria, Australia and it has been submitted for publication with their permission.

REFERENCES

1. SHAMIR U. and HOWARD C. Analytic approach to scheduling pipe replacement. *Journal of American Water Works Association*, 1979, **71**, No. 5, 248–258.

2. O'Day D. K., Fox C. M. and Huguet G. M. Aging urban water systems, a computerized case study. *Journal of Public Works*, 1980, **111**, No. 8, 61–64.
3. Andreou S. A., Marks D. H. and Clark R. M. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Advanced Water Resources*, 1987, **10**, No. 1, 2–10.
4. Clark R. M. and Goodrich J. A. Developing a data base on infrastructure needs. *Journal of the American Water Works Association*, 1988, **81**, No. 90, 81–87.
5. Karaa F. A. and Marks D. H. Performance of water distribution networks: Integrated approach. *Journal of Performance Construction Facilities*, 1990, **4**, No. 1, 51–67.
6. Gustafson J. M. and Clancy D. V. Modelling the occurrence of breaks in cast iron water mains using methods of survival analysis. *Proceedings of the 1999 Annual Conference of the American Water Works Association*. American Water Works Association, Denver, USA, 1999.
7. Kettler A. J. and Goulter I. C. Analysis of pipe breakage in urban water distribution networks. *Canadian Journal of Civil Engineering*, 1985, **12**, No. 2, 286–293.
8. Goulter I. C. and Kazemi A. Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Canadian Journal of Civil Engineering*, 1988, **15**, No. 1, 91–97.
9. Jacobs P. and Karney B. GIS development with application to cast iron water main breakage rate. *Proceedings of 2nd International Conference on Water Pipeline Systems*. Edinburgh, Scotland, BHR Group Ltd, 1994.
10. Male J. W., Walski T. M. and Slutsky A. H. Analysis of New York City's water main replacement policy. *Proceedings of Conference on Pipeline Infrastructure*, ASCE, New York, 1988.
11. Male J. W., Walski T. M. and Slutsky A. H. Analyzing water main replacement policies. *ASCE Journal of Water Resources Planning and Management*, 1990, **116**, No. 3, 362–374.
12. Eisenbeis P., Le Gauffre P. and Saegrov S. Water infrastructure management: An overview of European models and fatabases. *Proceedings of AWWARF Infrastructure Conference*, Baltimore, Maryland, 2000.
13. Pelletier G., Mailhot A. and Villeneuve J. P. Modeling water pipe breaks—three case studies. *ASCE Journal of Water Resources Planning and Management*, 2003, **129**, No. 2, 115–123.
14. Sullivan J. P. Maintaining aging systems—Boston's approach. *Journal of American Water Works Association*, 1982, **74**, No. 11, 555–559.
15. Herz R. Ageing processes and rehabilitation needs of drinking water distribution networks. *Journal of Water Supply Research and Technology—Aqua*, 1996, **45**, No. 5, 221–231.
16. Deb A. K. Quantifying future rehabilitation and replacement needs of water mains. *AWWA Research Conference*, Denver, 1998.
17. Cox D. R. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, 1972, **34**, No. 2, 187–220, series B.
18. Kaara A. F. *A Decision Support Model for the Investment Planning of the Reconstruction and Rehabilitation of Mature Water Distribution Systems*. PhD thesis, University of MIT, Cambridge, Massachusetts, 1984.
19. Andreou S. *Predictive Models for Pipe Break Failures and their Implications on Maintenance Planning Strategies for Deteriorating Water Distribution Systems*. PhD thesis, University of MIT, Cambridge, Massachusetts, 1986.
20. Le Gat Y. Forecasting pipe failures in drinking water network using stochastic processes models—respective relevance of renewal and Poisson processes. *Proceedings of 13th Conference of EJSW, Dresden University of Technology, Montreal*, 1999.
21. Dehghan A. and McManus K. Improved estimation of water pipes reliability for urban water supply systems. *Proceedings of 2005 Conference on Structural Condition Assessment, Monitoring and Improvement, Perth, Australia*, 2005.
22. Bras R. L. and Rodriguez-Iturbe I. *Random Functions and Hydrology*. Dover Publications, New York, 1993, Ch. 1, pp. 1–11.
23. Balakrishnan A. V. *Introduction to Random Processes in Engineering*, 1st edn. John Wiley, New York, 1995, Ch. 2, pp. 63–108.
24. Rajani B. and Makar J. A methodology to estimate remaining service life of gray cast iron water mains. *Canadian Journal of Civil Engineering*, 2000, **27**, No. 6, 1259–1272.
25. Crowder M. J., Kimber A. C., Smith R. L. and Sweeting T. J. *Statistical Analysis of Reliability Data*. Chapman & Hall, London, 1994.
26. Melbourne Water. *Water Main Renewal Study*. Melbourne Water, Melbourne, Australia, 1991.
27. Water Service Association of Australia. *WSAA Facts '99*. WSAA, Melbourne, 1991.
28. Righetti T. L. Cast iron condition assessment study. In *Internal Report, Neural Connection*. City West Water, Melbourne, 2001.
29. Nelson J. D. and Debora J. M. *Expansive Soils: Problems and Practice in Foundation and Pavement Engineering*. John Wiley, New York, 1992.
30. Mann A. *The Identification of Ideal Locations of Moisture Barriers*. Masters thesis, Swinburne University of Technology, 1997.
31. Rousseeuw P. J. and Leroy A. M. *Robust Regression and Outlier Detection*. Wiley, Chichester, 2003.