IIE Transactions (2005) **37**, 533–542

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170590929009

Stochastic models for degradation-based reliability[1]

JEFFREY P. KHAROUFEH* and STEVEN M. COX

Department of Operational Sciences, Air Force Institute of Technology, AFIT/ENS, 2950 Hobson Way, Wright Patterson AFB, OH 45433-7765, USA
E-mail: Jeffrey.Kharoufeh@afit.edu

Received April 2003 and accepted June 2004

This paper presents a degradation-based procedure for estimating the full and residual lifetime distribution of a single-unit system subject to Markovian deterioration. The hybrid approach unites real degradation measures with analytical, stochastic failure models to numerically compute the distributions and their moments. Two distinct models are shown to perform well when compared with simulated data. Moreover, results obtained from the second model are compared with empirically generated lifetimes of 2024-T3 aluminum alloy specimens. The numerical experiments indicate that the proposed techniques are useful for remaining lifetime prognosis in both cases.

1. Introduction

In this paper, we consider the problem of using sensor data to estimate full and residual lifetime distributions for a single-unit system subject to a stochastically evolving environment. We are particularly motivated by applications in autonomous prognostic systems whose primary goals are to assess the current and future health of single- or multi-unit systems to avoid catastrophic failures and to eliminate superfluous preventive maintenance activities. It is well known that the difficulty in estimating lifetime probabilities is the scarcity or absence of actual failure time observations (failure-based reliability). This stems from the fact that it is often infeasible to run components to failure, whereas accelerated life tests may not be representative of the true operating environment of components; thus, yielding possibly unreliable results. Hence, novel techniques that exploit readily available sensor data for lifetime estimation are needed. Examples of such data may be the current state of the ambient environment of the single-unit system or measures of degradation suffered by the single-unit system up to some point in time.

Relative to failure-based reliability, degradation-based reliability has received a modest amount of attention in the open literature. The use of degradation measures to assess component lifetimes was addressed in the early work of Gertsbakh and Kordonsky (1969) who used a sample path approach to assess the reliability for a simple linear degradation path with random slope and intercept. Later, Lu and Meeker (1993) estimated the failure time distribution of a unit that fails once the cumulative damage exceeds a given threshold value. Their parametric statistical model allows for a bootstrap method to generate confidence intervals around the failure time distribution, but does not necessarily allow for a closed-form expression of the distribution. More recently, Meeker and Escobar (1998) provide a useful summary of degradation models, emphasizing the use of linear models with assumed log-normal rates of degradation. In such a case, the full lifetime distribution can be computed analytically. Other recent models encountered in the literature deal with the degradation of materials such as those due to Gillen and Celina (2001). Meeker *et al.* (2002) discussed general approaches to estimating lifetime distributions in accelerated life tests for highly variable environments. The models presented therein focus on the specification of the degradation path as it depends explicitly on the operating environment. In that paper, the authors note the need for a formal stochastic process model of the ambient environment and for expedient numerical techniques for the evaluation of these measures. This work provides a first step in addressing both of these issues.

Analytical lifetime distribution models for single-unit systems have been studied extensively and exist primarily in the form of stochastic shock and wear models. Bogdanoff and Kozin (1985) provide a good summary of probabilistic models of cumulative damage focusing on discrete-time versions of shock models. Such models assume that the unit sustains a random amount of damage each time a shock occurs at either fixed or random time intervals. Wear processes differ from shock models in that they assume that

[1]The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.
*Corresponding author

damage accumulates continuously over time. Some properties of the failure time distribution have been examined for such processes when the degradation path is assumed to be a Lévy process (Esary *et al.*, 1973; Abdel-Hameed, 1984). Singpurwalla (1995) gave an excellent summary of a variety of stochastic failure time models for systems in a random environment and particularly noted the difficulty of implementing many of these in a practical setting. Recently, Kharoufeh (2003) provided explicit transform results for the full lifetime distribution and moments for a single-unit system in a time-homogeneous Markovian environment. The main results of that work provide the basis for our procedures here.

In the spirit of other degradation-based approaches, we assume that a single-unit system resides in a random operating environment so that its rate of degradation depends explicitly on the influence of external factors. The evolution of the random environment is characterized as a stationary Continuous-Time Markov Chain (CTMC). The key innovation of our approach is the union of analytical models (specifically Markovian wear processes) with the use of real sensor data to estimate full and residual lifetime distributions. We present two models. The first one assumes that the sensor data provide information on the state of the random environment. The second one provides information on the cumulative degradation up to some fixed point in time. Our aim is to provide a novel, hybrid approach that will ultimately lead to a feasible, and implementable, prognostic system.

The remainder of the paper is organized as follows. The next section reviews the mathematical model describing unit lifetime distributions. In Section 3 we demonstrate the means by which real sensor data may be used to estimate the model parameters when the operating environment is observable. Section 4 presents a model for the case when only the cumulative degradation level is observable. Several illustrative examples are provided in Section 5, and our concluding remarks and future research directions are given in Section 6.

2. Mathematical model description

This section provides the general mathematical framework upon which we build our estimation technique using real degradation data. It briefly summarizes the main results of Kharoufeh (2003) for a single-unit system subject to a random environment modelled as a CTMC and extends the results to the residual lifetime of the unit.

The rate of degradation of the system at time $t > 0$ is governed by a random environment that is modelled as a finite state CTMC, $\{Z(t) : t \geq 0\}$. Let $Z(t)$ denote the state of the random environment process at time $t \in \mathbb{R}_+$, where $\mathbb{R}_+$ denotes the non-negative real line, and define the state space of Z by the set $S \subseteq \mathbb{N}$, the set of natural numbers. We specifically assume that Z has a finite state space $S = \{1, \ldots, K\}$ with $K \geq 2$. Let $R(t)$ be defined as the degradation rate of the system at time $t \in \mathbb{R}_+$ and define a positive function $r : S \to (0, \infty)$. The properties of the function $r(\cdot)$ are dictated by the type of system under consideration and its surrounding environment. Since the degradation rate of the system is explicitly dependent on the environment process, the degradation rate assumes values in the space $\mathcal{D} = \{r(1), \ldots, r(K)\}$.

Next, we describe the stochastic evolution of the system. If $Z(t) = i \in S$, then (i) $R(t) := r(Z(t)) = r(i) \in \mathcal{D}$; and (ii) the environment transitions from state $i \in S$ to state $j \in S$ at time t according to a Markov transition function $\mathbf{P}(t) := [p_{i,j}(t)]$ where $p_{i,j}(t) := P\{Z(t) = j | Z(0) = i\}$. The environment process Z is assumed to be a temporally homogeneous, finite-state Markov process so that $p_{i,j}(t)$ does not depend on t for all $i, j \in S$. Denote by $X(t)$, the cumulative degradation of the single-unit system up to time $t \in \mathbb{R}_+$. The continuous-time, continuous-state degradation process $\{X(t) : t \geq 0\}$ assumes values on the non-negative real line. Moreover, this monotonically increasing process is a continuous, additive functional of Z, and thus, (X, Z) constitutes a special case of a Markov additive process (Çinlar, 1977). The main contribution of Kharoufeh (2003) was a closed-form expression for the cumulative distribution function (cdf) and moments of the system lifetime via an analysis of the bivariate Markov process $\{(X(t), Z(t)) : t \geq 0\}$. Next, we briefly review those results without proof. For specific details and proofs, the reader is referred to Kharoufeh (2003).

The cumulative degradation of the single-unit system up to time $t \in \mathbb{R}_+$ is defined by the cumulative stochastic process:

$$X(t) = \int_0^t r(Z(u))\mathrm{d}u, \tag{1}$$

when $X(0) \equiv 0$. The system fails as soon as the magnitude of its accumulated degradation exceeds a fixed threshold value x (i.e., a soft failure). The lifetime of the system is given by the random variable:

$$T(x) = \inf\{t : X(t) > x\}, \tag{2}$$

or the first time the X process exceeds the threshold value x. Define the probability distributions:

$$V_{i,j}(x, t) = P\{X(t) \leq x, Z(t) = j | Z(0) = i\}, \quad i, j \in S, \tag{3}$$

where $V_{i,j}(x, t)$ is the joint probability that, at time t, the degradation of the system has not exceeded the value x and the environment process is in state $j \in S$ given that the environment was initially in state $i \in S$. The distribution matrix of $X(t)$ is defined as:

$$\mathbf{V}(x, t) = [V_{i,j}(x, t)]. \tag{4}$$

Due to the dual relationship of Equation (2), the cdf of $T(x)$, the unit's full lifetime, is obtained as:

$$F(x, t) := P\{T(x) \leq t\} = 1 - \boldsymbol{\alpha}\mathbf{V}(x, t)\mathbf{e}, \tag{5}$$

where $\boldsymbol{\alpha} := [\alpha_i]$ with $\alpha_i := P\{Z(0) = i\}$ and $\mathbf{e}$ is a K-dimensional column vector of ones.

Of particular importance to the problem of lifetime prognosis is evaluation of the unit's *residual* lifetime distribution given that it has not failed by time $\xi_0 > 0$. We may write this distribution directly as:

$$\begin{aligned} S(x, t \mid \xi_0) &:= P\{T(x) \geq t + \xi_0 \mid T(x) > \xi_0\}, \\ &= (1 - F(x, t + \xi_0))/(1 - F(x, \xi_0)), \\ &= \boldsymbol{\alpha}\mathbf{V}(x, t + \xi_0)\mathbf{e}/(\boldsymbol{\alpha}\mathbf{V}(x, \xi_0)\mathbf{e}). \end{aligned} \tag{6}$$

By integrating the tail, the nth moment of the unit lifetime, for $n \geq 1$, is given by:

$$\begin{aligned} m^{(n)}(x) := E[T^n(x)] &= \int_{\mathbb{R}_+} nt^{n-1} P\{T(x) > t\}\mathrm{d}t \\ &= \int_{\mathbb{R}_+} nt^{n-1}(\boldsymbol{\alpha}\mathbf{V}(x, t)\mathbf{e})\mathrm{d}t, \end{aligned} \tag{7}$$

and the nth moment of the unit's residual lifetime, for $n \geq 1$, is given by:

$$\begin{aligned} m^{(n)}(x \mid \xi_0) &:= E[T^n(x) \mid T(x) > \xi_0], \\ &= \int_{\mathbb{R}_+} nt^{n-1} S(x, t \mid \xi_0)\mathrm{d}t, \\ &= \frac{1}{\boldsymbol{\alpha}\mathbf{V}(x, \xi_0)\mathbf{e}} \int_{\xi_0}^{\infty} nt^{n-1}(\boldsymbol{\alpha}\mathbf{V}(x, t)\mathbf{e})\mathrm{d}t. \end{aligned} \tag{8}$$

Equations (6) and (8) indicate that the residual lifetime distribution and moments may be computed via the full lifetime distribution. For this reason, we need only concern ourselves with the numerical evaluation of Equation (5). Kharoufeh (2003) showed that the matrix, $\mathbf{V}(x, t)$, satisfies the partial differential equation:

$$\mathbf{V}_t(x, t) + \mathbf{V}_x(x, t)\mathbf{R}_{\mathrm{D}} = \mathbf{V}(x, t)\mathbf{Q}, \tag{9}$$

where $\mathbf{V}_t(x, t)$ and $\mathbf{V}_x(x, t)$ denote the partial derivatives of $\mathbf{V}(x, t)$ with respect to t and x, respectively, $\mathbf{R}_D = \mathrm{diag}(r(1), \ldots, r(K))$, and $\mathbf{Q}$ is the infinitesimal generator matrix for $\{Z(t) : t \geq 0\}$. The solution to the PDE, obtained in the two-dimensional transform space, is given by:

$$\tilde{\mathbf{V}}^*(u, s) = (u\mathbf{R}_{\mathrm{D}} + s\mathbf{I} - \mathbf{Q})^{-1}, \quad \mathrm{Re}(u) > 0, \quad \mathrm{Re}(s) > 0, \tag{10}$$

where $\tilde{\mathbf{V}}^*(u, s)$ is obtained by the two matrix transforms.

$$\mathbf{V}^*(x, s) = \int_{\mathbb{R}_+} e^{-st}\mathbf{V}(x, t)\mathrm{d}t, \tag{11}$$

the Laplace transform of $\mathbf{V}(x, t)$ with respect to t and

$$\tilde{\mathbf{V}}^*(u, s) = \int_{\mathbb{R}_+} e^{-ux}\mathbf{V}^*(\mathrm{d}x, s), \tag{12}$$

the Laplace-Stieltjes Transform (LST) of $\mathbf{V}^*(x, s)$ with respect to x. Furthermore, we define the following transforms:

$$F^*(x, s) = \int_{\mathbb{R}_+} e^{-st}F(x, t)\mathrm{d}t,$$

$$\tilde{F}^*(u, s) = \int_{\mathbb{R}_+} e^{-ux}F^*(\mathrm{d}x, s).$$

The full lifetime distribution of a single-unit system in a homogeneous Markovian environment is:

$$\tilde{F}^*(u, s) = s^{-1} - \boldsymbol{\alpha}\tilde{\mathbf{V}}^*(u, s)\mathbf{e}, \quad \mathrm{Re}(u) > 0, \quad \mathrm{Re}(s) > 0. \tag{13}$$

The two-dimensional transform of Equation (13) can be inverted numerically using inversion algorithms such as those due to Moorthy (1995) and Abate *et al.* (1998).

Using Equation (13), it was further shown by Kharoufeh (2003) that the (LST) of the nth moment of the unit lifetime, for $n \geq 1$, is given by:

$$\tilde{m}^{(n)}(u) = n!\boldsymbol{\alpha}(u\mathbf{R}_{\mathrm{D}} - \mathbf{Q})^{-n}\mathbf{e}. \tag{14}$$

The transform of Equation (14) can be inverted numerically using the one-dimensional inversion algorithm of Abate and Whitt (1995). Unfortunately, a closed-form expression for the LST of the nth residual lifetime moment ($n \geq 2$) does not appear to exist. However, the mean residual lifetime may be obtained by evaluating (numerically) the integral:

$$m^{(1)}(x \mid \xi_0) = \frac{1}{\boldsymbol{\alpha}\mathbf{V}(x, \xi_0)\mathbf{e}} \int_{\xi_0}^{\infty} (\boldsymbol{\alpha}\mathbf{V}(x, t)\mathbf{e})\mathrm{d}t. \tag{15}$$

The analytical models of this section provide a viable means for residual lifetime prognosis. In the sections that follow, we extend the basic models and apply them to two distinct scenarios in which real sensor data may be used to obtain the pertinent measures. Both models require statistical estimation of the parameters of the governing Markov chain.

3. Model I: Observable environment

In this section, it is assumed that sensor data provide information regarding the current state of the ambient environment in which the single-unit system resides while the degradation status of the unit is unobservable. Similar models, developed for engineering and medical applications, can be found in the current lifetime analysis literature (Whitmore *et al.*, 1998; Lee *et al.*, 2000). Our environment process is analogous to a "marker" process whereas the degradation process corresponds to a "latent" process that is assumed to be unobservable. In those works, the marker and latent processes were assumed to form a bivariate Brownian motion process. In this paper, we impose the assumption that the marker process is a Markov chain that directly influences the latent (degradation) process; however, we make no assumptions regarding the probability law of the latter.

Under its prevailing operating conditions, the system wears continuously (and additively) until its cumulative degradation exceeds a fixed, deterministic value x at which time the unit is said to have failed. Optimally-located sensors provide real-time data regarding the conditions of the

environment (e.g., temperature, pressure, humidity, etc.); however, the cumulative amount of degradation cannot be discerned from these readings. It is further assumed that the environment state space may be partitioned into K distinct states as $S = \{1, 2, \ldots, K\}$, $K < \infty$. Associated with each distinct environment state is a known degradation rate, namely $r(j)$, $j \in S$ so that when the environment is in state $j \in S$, the system degrades at rate $r(j)$. Knowledge of the real-valued function r is assumed to be available through physical properties of the unit under consideration. Using this information, it is possible to directly apply the results of Section 2 to characterize the full and residual lifetime distributions. However, in order to do so, we require a surrogate (estimated) stochastic process for the true, underlying environment process.

Equation (13) requires specification of three key components: (i) the initial distribution vector ($\boldsymbol{\alpha}$) for the environment process: (ii) the diagonal matrix of degradation rates ($\mathbf{R}_{\mathrm{D}}$); and (iii) the infinitesimal generator matrix ($\mathbf{Q}$) for the environment process. Since the environment process is observable and the degradation rates are assumed to be known for each environmental state, it is necessary only to approximate the matrix $\mathbf{Q}$. We do so by employing standard statistical inference techniques for Markov processes (Basawa and Rao, 1980).

We assume that the environment process $\{Z(t) : t \geq 0\}$ is observable (and homogeneous) over the time interval $[0, T]$. The system has not failed by time T, and began its lifetime in perfect working order at time 0 (i.e., $X(0) \equiv 0$). The environment is continuously observed up to time T so that, at each transition epoch, we record the current and subsequent states of the random environment. Let $q(i, j)$ denote the rate at which the environment transitions from state $i \in S$ to state $j \in S$, $j \neq i$. If the environment process $\{Z(t) : t \geq 0\}$ is a continuous-time Markov chain, then the holding time in state i, given that the subsequent state is $j \neq i$, is exponentially distributed with rate parameter $q(i, j)$. Let $N_T(i, j)$ denote the random, integer number of transitions of the process from i to j in a time interval of length T. Moreover, let $H_T(i)$ denote the total holding time in state $i \in S$ during $[0, T]$. It can be shown (Basawa and Rao, 1980) that:

$$q(i, j) = \lim_{T \to \infty} \frac{N_T(i, j)}{H_T(i)}. \tag{16}$$

Therefore, for sufficiently large T, we may approximate $q(i, j), j \neq i$ by:

$$q(i, j) \approx \hat{q}_T(i, j) = \frac{N_T(i, j)}{H_T(i)}. \tag{17}$$

The diagonal elements of the generator matrix are obtained as:

$$\hat{q}_T(i, i) = -\sum_{j \neq i} \hat{q}_T(i, j), \quad i \in S. \tag{18}$$

We pause here to note that, for a fixed observation interval, it is possible to alternatively estimate the generator matrix by observing k independent sample paths of the Z process. Define $N^{(k)}(i, j)$ as the total number of transitions from state i to j over all k trials and define $H^{(k)}(i)$ as the total holding time in state i over all k trials. The off-diagonal elements of the generator matrix are given by:

$$q(i, j) = \lim_{k \to \infty} \frac{N^{(k)}(i, j)}{H^{(k)}(i)}, \tag{19}$$

so that for k sufficiently large, the approximation of $q(i, j)$, $j \neq i$ is:

$$q(i, j) \approx \hat{q}^{(k)}(i, j) = \frac{N^{(k)}(i, j)}{H^{(k)}(i)}. \tag{20}$$

We obtain the diagonal elements by:

$$\hat{q}^{(k)}(i, i) = -\sum_{j \neq i} \hat{q}^{(k)}(i, j) \quad i \in S. \tag{21}$$

It has been shown in Basawa and Rao (1980) that:

$$q(i, j) = \lim_{T \to \infty} \frac{N_T(i, j)}{H_T(i)} = \lim_{k \to \infty} \frac{N^{(k)}(i, j)}{H^{(k)}(i)}. \tag{22}$$

In model I, we observe a single sample path over a sufficiently long period using Equations (17) and (18), whereas we utilize Equations (20) and (21) in model II. Let $\hat{\mathbf{Q}}_T := [\hat{q}_T(i, j)]$ and let $\hat{\mathbf{V}}(x, t)$ denote the matrix solution to Equation (9) when the generator matrix $\hat{\mathbf{Q}}_T$ is used as a surrogate for $\mathbf{Q}$. The estimated matrix distribution, $\hat{\mathbf{V}}(x, t)$, is obtained through the double Laplace transform inversion:

$$\hat{\mathbf{V}}(x, t) = \mathcal{L}^{-1}\{u^{-1}(u\mathbf{R}_{\mathrm{D}} + s\mathbf{I} - \hat{\mathbf{Q}}_T)^{-1}\},$$

where $\mathrm{Re}(u) > 0$, $\mathrm{Re}(s) > 0$, and $\mathcal{L}^{-1}$ denotes the (double) inverse Laplace operator.

The resulting estimates for the full and residual lifetime distributions are, respectively.

$$\hat{F}(x, t) = 1 - \boldsymbol{\alpha}\hat{\mathbf{V}}(x, t)\mathbf{e}, \tag{23}$$

and

$$\hat{S}(x, t \mid \xi_0) = \frac{\boldsymbol{\alpha}\hat{\mathbf{V}}(x, t + \xi_0)\mathbf{e}}{\boldsymbol{\alpha}\hat{\mathbf{V}}(x, \xi_0)\mathbf{e}}. \tag{24}$$

We illustrate the estimation procedure via numerical examples in Section 5. In model II, we assume that only degradation measures are observable for the single-unit system.

4. Model II: Observable degradation

The ultimate purpose of a prognostic system is to assess the current "health" of the unit and to make inferences regarding the remaining useful lifetime of the unit. Model II provides a viable first step towards an analytical approach to address this problem when the component's degradation is governed by a stationary, Markov environment. Assume, at time $t > 0$, that the cumulative degradation of the single-unit system, $X(t)$, is observable. In contrast to model I, the

environment, its number of distinct states, and its degradation rates are unknown.

The basis of our procedure can be described as follows. We approximate the true degradation path using simple, piecewise linear functions that pass through a finite number of observations. The slope of each line segment approximates the degradation rate over a given time interval. Once the slopes have been collected, we perform K-means cluster analysis wherein the centroid of each cluster corresponds to the mean degradation rate for a distinct state of the environment. However, selection of the integer K is non-trivial. We describe a means by which to estimate this value in the next subsection.

4.1. *Estimating the number of states*

In the absence of environmental state observations, it is difficult to discern the total number of distinct states that might be encountered by the unit. In some applications, the number of states might be obvious. For example, in a manufacturing context, if a cutting tool operates at only three distinct cutting speeds (e.g., slow, moderate, fast), and the wear rate of the cutting tool is known for each speed, then three distinct "environment" states exist. In other applications, however, it may be necessary to rely on the knowledge and experience of subject matter experts to determine an appropriate number of environment states. In this subsection, we address scenarios in which the appropriate integer number of states is not obvious, and the engineer has at his/her disposal only the observed degradation of the unit. The problem is that of selecting an appropriate partition of the set of all observed degradation rates over some finite observation period.

Fraley and Raftery (1998) note that the method of K-means clustering is the most widely used non-hierarchical method for partitioning a set of real observations into K distinct clusters if the value K is known *a priori*. We now describe a method to estimate the appropriate value of K when a total of N degradation rates have been observed. The technique performs a comparison of K means via a one-way analysis of variance by computing an appropriate F-ratio that depends explicitly on K.

Denote by $\bar{y}_i$ the mean of cluster i, $i = 1, 2, \ldots, K$, and let n_i denote the number of observations in the ith cluster such that:

$$N = \sum_{i=1}^{K} n_i.$$

The overall sample average of degradation rates is:

$$\bar{y} = \frac{1}{K} \sum_{i=1}^{K} \bar{y}_i.$$

Define by SSB, the sum of squares between clusters which is:

$$SSB = \sum_{i=1}^{K} n_i (\bar{y}_i - \bar{y})^2. \tag{25}$$

Similarly, the sum of squares within clusters (SSW) is:

$$SSW = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \tag{26}$$

where y_{ij} denotes the jth degradation rate observed in the ith cluster. The F-ratio for K clusters, $2 \leq K < N$, is:

$$F_K = \frac{SSB/(K-1)}{SSW/(N-K)}. \tag{27}$$

Our objective is to find the "best" value of K via Equation (27). Calinski and Harabasz (1974) suggest choosing K such that, over some set $\mathcal{K}$, F_K is the absolute maximum, the first local maximum, or a point at which the function exhibits a comparatively rapid increase. It is worth noting that, if F_K is monotonically increasing in K, then the number of clusters (states) should be equal to the total number of observations (i.e., each observation constitutes a cluster). Let $\mathcal{K} \equiv \{2, 3, \ldots, J\}$ where J is a positive integer. For the sake of computational expedience, it is ideal to choose the smallest value of K that leads to an appropriate representation of the underlying process. As a first course of action, we apply the second criterion suggested in Calinski and Harabasz (1974), and choose a value of K that corresponds to the first local maximum over $\mathcal{K}$ given by:

$$\hat{K} = \min\{K \in \mathcal{K} : F_K > F_{K+1}\}. \tag{28}$$

However, in the case where F_K is strictly increasing in K, we resort to the remaining two criteria to estimate the smallest possible value. We illustrate and apply the estimation procedure in Section 5.

4.2. *Description of the estimation procedure*

A formal description of the full estimation procedure (model II) is now provided.

Step 0. Initialization.
At time $t_0 \equiv 0$, observe $\hat{X}(t_0) \equiv x_0$.

Step 1. Observe degradation measures.
Observe the degradation at times $t_1 < \cdots < t_M$, $M \in \mathbb{N}$ and form the set of observation times:

$$\mathcal{T} := \{t_j : j = 0, 1, 2, \ldots, M\}. \tag{29}$$

with observations $\hat{X}(t_j), j = 0, 1, 2, \ldots, M$. It is assumed that, at time t_M, the single-unit system has not failed.

Step 2. Approximate the failure path.
After observing the degradation path up to time epoch t_M, approximate the true failure path by

a simple piecewise-linear approximation that connects the observed degradation measures for each element of $\mathcal{T}$.

Step 3. Approximate degradation rates via finite difference methods.
For each observation time $t_j \in \mathcal{T}$, approximate the potential degradation rates of the process by the difference equation:

$$\gamma_j := \frac{\hat{X}(t_j) - \hat{X}(t_{j-1})}{t_j - t_{j-1}}, \qquad j = 1, 2, \ldots, M. \quad (30)$$

We assume that the inter-observation times ($t_j - t_{j-1}$, $j = 1, 2, \ldots, M$) are equal. For the discrete sampling interval, $[t_0, t_M]$, gather the observed rates in a set:

$$\Gamma = \{\gamma_j : j = 1, 2, \ldots, M\}.$$

Step 4. Compute the $\hat{K}$ distinct degradation rates.
Select an appropriate value $\hat{K}$ using the procedure outlined in Section 4.1 applied to observations in the set Γ. This may be accomplished by using any standard statistical software package. Define $\mathcal{C} = \{C_1, C_2, \ldots, C_{\hat{K}}\}$ as the set of $\hat{K}$ distinct clusters such that $C_i \cap C_j = \emptyset, j \neq i$, and μ_i denotes the centroid of cluster C_i, $i = 1, 2, \ldots, \hat{K}$. Each $\gamma_j \in \Gamma$ is therefore assigned to exactly one cluster in $\mathcal{C}$ such that the estimated degradation rate of environment state $j \in S$ is $\hat{r}(j) \equiv \mu_j$.

Step 5. Construct new degradation path.
Construct a new piecewise linear degradation path by replacing each $\gamma_j \in \Gamma$ by $\hat{r}(j) \equiv \mu_j$. (Note: if $\hat{r}(j) = \hat{r}(j+1)$ for some j, then the two adjacent line segments are replaced by a single line segment with slope equal to $\hat{r}(j)$.)

Step 6. Approximate generator matrix.
Using the piecewise linear estimate of the degradation path of Step 5, estimate the $\hat{K}$-dimensional generator matrix $\mathbf{Q}$ by $\hat{\mathbf{Q}}_{t_M}$ using Equations (17) and (18).

Remark 1. Three important assumptions are employed here: (i) degradation of the unit is perfectly observable; (ii) at time t_M, the single-unit system has not failed; and (iii) the degradation paths are monotonically increasing.

5. Numerical examples

This section provides illustrative examples of the failure time models (I and II) of the two preceding sections. We validate our technique by comparing the full and residual lifetime probability values with simulated values. Moreover, we test model II by comparing results with empirical observations of unit lifetimes.

5.1. *Model I examples*

For each $j \in S$, there exists a known degradation rate for the single-unit system, namely $r(j)$. The environment evolves over time and is observed over the interval $[0, T]$. We provide three test cases, namely when the environment process (Z) assumes values on a state space with $|S| = 2$, 5 and 10, respectively. For each test case, the elements of the generator matrix $\mathbf{Q}$ were drawn from a continuous uniform population on (20, 40) such that, for $j \neq i$, $q(i, j) \sim U(20, 40)$. In the usual way, the diagonal elements of the generator matrix are:

$$q(i, i) = -\sum_{j \neq i} q(i, j), \quad i \in S.$$

Moreover, the true generator matrix $\mathbf{Q}$ is estimated by the matrix $\hat{\mathbf{Q}}_T$ with $T = 100, 500, 5000$ using Equations (17) and (18). The associated degradation rates ($r(j)$, $j = 1, 2, \ldots, K$) were drawn from continuous uniform populations on the interval (20, 80). In all test cases, the failure threshold value was fixed at $x = 1.0$ units. Full and residual probability values were computed via Equations (5) and (6), respectively, using a variant of the inversion algorithm in Moorthy (1995).

We compare, at m fixed points, $\tau_1, \tau_2, \ldots, \tau_m$, distribution function values generated by the true and estimated processes. Goodness-of-fit tests were conducted (at the 0.05 level) using the Cramér-von Mises test statistic to compare full and residual lifetime distributions. The appropriate test statistic, denoted by κ_2, is given by:

$$\kappa_2 = \frac{1}{2} \sum_{i=1}^{m} \left(G_1(\tau_i) - G_2(\tau_i)\right)^2, \quad (31)$$

where G_1 denotes the distribution function obtained from the estimated environment process and G_2 is the cdf generated using the true environment process. We denote the critical value by κ^*. If $\kappa_2 < \kappa^*$, then we fail to reject the null hypothesis that $G_1 \equiv G_2$. Table 1 provides a summary of the Cramér-von Mises test statistics when comparing cumulative probability values obtained by simulating the true process with generator $\mathbf{Q}$ with those of numerical Laplace transform inversion obtained using $\hat{\mathbf{Q}}_T$. Three observation periods were considered in this experiment. For the residual life tests, we fixed $\xi_0 = E[T(x)]$.

It is noted that, in all 18 experiments, we fail to reject the null hypothesis that the estimated and true distribution functions are equivalent. Some additional remarks are warranted regarding these results.

1. If the environment process can be partitioned into K distinct and observable states, our approach provides a viable approximation procedure that does not require failure time observations; it requires only a count of environment transitions.
2. In the case where the environmental state holding times are non-exponentially distributed, the procedure

Table 1. Cramér-von Mises test statistics for model I experiments ($\kappa^* = 0.461$, $\alpha = 0.05$)

Distribution	*Run length*	$K = 2$	$K = 5$	$K = 10$
$\hat{F}(x, t)$	$T = 100$	7.013×10^{-3}	4.233×10^{-2}	9.477×10^{-3}
$(m = 48)$	$T = 500$	1.649×10^{-3}	7.895×10^{-3}	2.130×10^{-3}
	$T = 5000$	9.941×10^{-5}	9.019×10^{-5}	1.986×10^{-6}
$1 - \hat{S}(x, t\|\xi_0)$	$T = 100$	5.511×10^{-3}	3.821×10^{-2}	7.455×10^{-3}
$(m = 50)$	$T = 500$	1.223×10^{-3}	6.553×10^{-3}	2.113×10^{-3}
	$T = 5000$	5.544×10^{-5}	8.092×10^{-5}	2.576×10^{-6}

remains feasible since the holding time distributions may be approximated by phase-type distributions (e.g., generalized Erlang distributions) that retain the memoryless property. We shall elaborate further on this point in Section 6.

5.2. *Model II examples*

In this subsection, we first illustrate model II using simulated, linear degradation paths before testing the procedure on a set of non-linear degradation data. For the simulation experiment, sample paths were generated via a known environment process with K distinct states and known degradation rates $\{r(1), r(2), \ldots, r(K)\}$. Initially, we fix $K = 3$. Five hundred degradation sample paths, similar to the five shown in Fig. 1, were generated for the experiment.

Setting $t_0 \equiv 0$, we observe the level of degradation at M equally spaced points in time such that $t_M = 20$. That is, the degradation process is observed up to time 20. Figure 2 depicts a sample of observations ($M = 10$) obtained from the five degradation sample paths of Fig. 1 that were used to approximate a piecewise, linear degradation sample path by the procedure of Section 4.

Next, we utilize the approximation to obtain the full and residual lifetime distributions. In each case, we observe the system up to time $t_M = 20$ except that we vary the inter-observation time. In particular, we consider M = 20, 100 and 200 observations on [0, 20], respectively. The assumed unit failure threshold was fixed at $x = 152.6$ units. The same 500 simulated degradation sample paths were used for each value of M to estimate a new generator matrix and new degradation rates at times 4.0, 8.0 and 12.0 respectively.

As before, we compare the estimated versus "true" full lifetime cdfs via the Cramér-von Mises goodness-of-fit test. Additionally, we compare the estimated versus true residual lifetime distribution. It should be noted that, in order to test the null hypothesis, we use the cdf of the residual lifetime distribution ($1 - \hat{S}(x, t|\xi_0)$). The summarized results of Table 2 indicate that none of the tests are significant at the 0.05 level; thus, we are able to adequately estimate the full and residual lifetime distributions for these simulated degradation paths.

To illustrate our method to select an appropriate number of states ($\hat{K}$), we simulated multiple degradation sample paths of environment processes having five and 10 states, respectively. The F-ratios for each case are displayed in Table 3. Applying the first local maximum rule, the F-ratios suggest $\hat{K} = 4$ for the five-state environment and $\hat{K} = 8$ for the 10-state environment.

Consequently, we compared the full and residual lifetime distributions using $\hat{K}$ and $\hat{K} + 1$ states in each scenario. Table 4 indicates that we fail to reject the null hypotheses

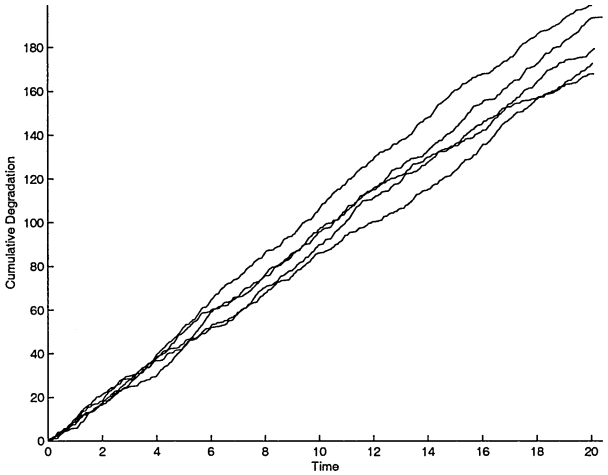

Fig. 1. A sample of five linear degradation paths.

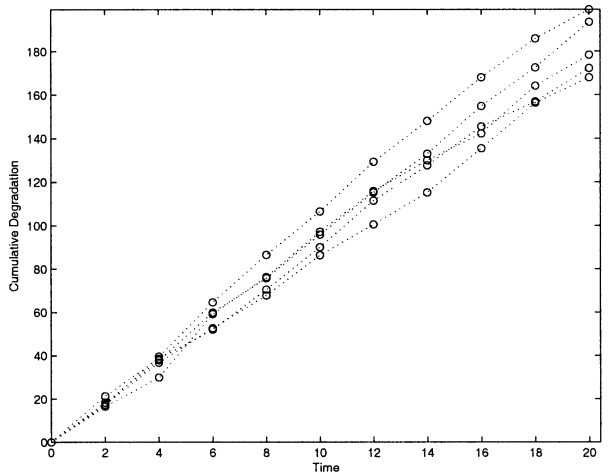

Fig. 2. Piecewise-linear approximation of the degradation paths.

Table 2. Cramér-von Mises test statistics for simulated data ($\kappa^* = 0.461$ at $\alpha = 0.05$)

M	*Interval*	$\hat{F}(x,t)$	ξ_0	$1-\hat{S}(x,t \mid \xi_0)$
20	[0.0,4.0]	0.0157	4.0	0.0189
	[0.0,8.0]	0.0210	8.0	0.1213
	[0.0,12.0]	0.0247	12.0	0.1633
100	[0.0,4.0]	0.0095	4.0	0.0197
	[0.0,8.0]	0.0088	8.0	0.0476
	[0.0,12.0]	0.0087	12.0	0.0639
200	[0.0,4.0]	0.0053	4.0	0.0149
	[0.0,8.0]	0.0051	8.0	0.0336
	[0.0,12.0]	0.0052	12.0	0.0266

that the full and residual lifetime distributions are equivalent at the 0.05 level.

Finally, we illustrate model II using real degradation data which correspond to the propagation of a fatigue crack in metallic materials. The data were originally analyzed by Virkler *et al.* (1979) who measured crack length propagation in 68 test specimens of a 2024-T3 aluminum alloy. The crack length of each sample was measured over time (in number of load cycles). Figure 3 depicts a representative sample of five degradation paths from the original data set. For a more thorough treatment of fatigue crack dynamics, the reader is referred to Ray and Tangirala (1997) and references contained therein.

To illustrate our procedure, let $X(t)$ denote the length of the crack at time t and assume the rate at which the crack grows is subject to its random environment (applied stress, ambient conditions, and other factors). It is assumed that these environmental factors can be characterized by $\{Z(t) : t \geq 0\}$, a temporally homogeneous (time-stationary) Markov process on a finite-state sample space. Although the underlying data may, in fact, exhibit non-stationary behavior, we shall observe all 68 sample paths to estimate the generator matrix (**Q**) for the Z process. In the experiment that follows, the component is said to fail

Table 3. F-ratio values for five- and 10-state environments ($\times 10^5$)

K	*Five-state*	*10-state*
2	1.654 973	0.285 050
3	3.090 848	0.400 770
4	4.067 369	0.421 102
5	3.820 857	0.660 245
6	3.249 939	0.698 204
7	6.678 639	0.955 170
8	7.278 563	1.144 632
9	10.76 289	1.047 932
10	7.383 198	0.697 347
11	5.876 550	1.306 396
12	9.920 737	1.805 007
13	8.558 721	0.857 228
14	15.161 074	0.858 865
15	12.455 254	1.386 817

Table 4. Cramér-von Mises test statistics with $\hat{K}$ states ($\kappa^* = 0.461$, $\alpha = 0.05$, $m = 100$)

K	$\hat{K}$	$\hat{F}(x,t)$	ξ_0	$1-\hat{S}(x,t\|\xi_0)$
5	4	2.3056×10^{-2}	17.5311	1.0179×10^{-1}
5	5	3.2317×10^{-2}	17.5311	1.0733×10^{-1}
10	8	3.5704×10^{-2}	13.4249	1.6783×10^{-1}
10	9	3.5397×10^{-2}	13.4249	1.4709×10^{-1}

whenever the crack length first exceeds a critical value of $x = 45$ mm. We estimate the off-diagonal generator matrix values by:

$$q(i,j) \approx \hat{q}^{(68)} = \frac{N^{(68)}(i,j)}{H^{(68)}(i)}, \tag{32}$$

as defined in Equation (20). It should be noted, however, that the empirical cumulative probabilities for the data were computed with a sample of only 68 observations. Nevertheless, we take this empirical distribution to be the "true" distribution.

The F-ratio test for selecting $\hat{K}$ was applied to the data by observing 55, 77 and 91% of the lifetime, respectively. We select two clusters at 55% of the lifetime, nine clusters at 77% of the lifetime (due to the sharp increase), and 12 clusters at 91% of the lifetime as seen in Table 5.

Estimates of the generator matrix and degradation rates were constructed for $\hat{K} = 2, 3, 4, 9$ and 12 when observing up to 91% of the lifetime of each specimen. The resulting cdfs (full and residual lifetime) for each case were compared to the empirical distribution at a fixed number of points ($m = 65$) using the Cramér-von Mises test. The results of the numerical tests are summarized in Table 6.

The results for the full lifetime distributions suggest that for $\hat{K} = 2$, 3 and 4, we fail to reject the null hypothesis that

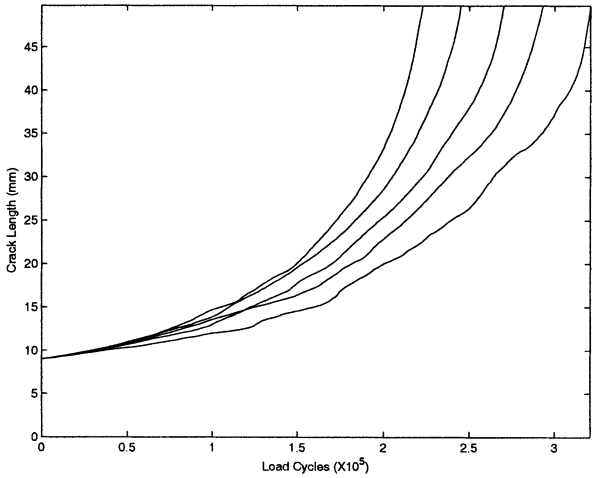

Fig. 3. The propagation of the fatigue crack length for five test specimens.

Table 5. F-ratio comparison ($\times 10^4$)

	Percentage of lifetime		
K	*55%*	*77%*	*91%*
2	0.845 912	1.716 694	1.934 667
3	0.770 946	2.045 060	2.448 228
4	0.700 849	2.267 842	2.866 000
5	0.637 918	2.412 888	3.248 069
6	1.789 474	2.506 080	3.620 849
7	1.984 733	2.627 177	3.934 574
8	2.230 180	2.658 314	4.208 318
9	2.570 793	2.669 967	4.374 201
10	2.785 035	4.566 412	4.448 147
11	3.017 830	4.974 103	4.744 162
12	2.993 362	5.272 737	4.830 633
13	3.714 449	5.683 212	4.796 518
14	4.167 025	5.850 034	4.800 008
15	4.292 031	6.608 150	4.745 232

the distributions are equal. For the residual lifetime distributions with $\hat{K} = 3$ and 4, we fail to reject the null hypothesis that the distributions are equal. The F-ratio test of Table 5 indicates an estimate of $\hat{K} = 12$. However, using an estimated 12-state generator matrix, the resulting full and residual lifetime cdfs fail to pass the goodness-of-fit tests.

Although it is apparent that there may exist greater than three or four environmental states, we surmise that the 12-state estimate is poor for a few reasons. First, to adequately estimate a 12-state generator matrix, a fairly large number of environmental transitions must occur during a finite observation period. It is obvious, by inspection of Fig. 3, that few significant transitions take place early on. Second, it is likely that transitions occur primarily between a subset of the overall environment sample space at different times. That is, some state transitions appear to dominate early in the specimen lifetimes whereas others dominate as the specimens age (i.e., as the crack grows). Consequently, we surmise that an age- or degradation-state-dependent model may be more appropriate for this data set, as it appears to violate the time-stationarity assumptions of our model. However, it is important to note that the estimation procedure indicated that only a few states could be used within our framework to provide adequate cumulative probability estimates. We surmise that there may exist a minimal representation for the underlying process that includes only the "most important" states. In the current framework, our model does not dynamically dictate which states are most important.

Table 6. Cramér-von Mises test statistic ($\kappa^* = 0.461$ at $\boldsymbol{\alpha} = 0.05$, $m = 65$)

$\hat{K}$	$\hat{F}(x, t)$	ξ_0	$1 - \hat{S}(x, t\|\xi_0)$
2	$1.398\,488 \times 10^{-1}$	2.534	$6.124\,815 \times 10^{-1}$
3	$8.671\,149 \times 10^{-2}$	2.534	$1.395\,285 \times 10^{-1}$
4	$4.525\,246 \times 10^{-1}$	2.534	$4.383\,202 \times 10^{-1}$
9	$7.657\,227 \times 10^{-1}$	2.534	2.235 864
12	$8.723\,972 \times 10^{-1}$	2.534	1.667 689

Table 7. Mean full and residual lifetimes ($\times 10^5$ load cycles), $\xi_0 =$ actual $m^{(1)}(x)$

	$m^{(1)}(x)$		$m^{(1)}(x \mid \xi_0)$	
x *(mm)*	*Actual*	*Model*	*Actual*	*Model*
10	0.318 574	0.325 974	0.349 943	0.347 018
15	1.183 405	1.150 109	1.268 688	1.293 640
20	1.638 693	1.530 864	1.738 733	1.771 963
25	1.938 014	1.800 568	2.042 441	2.113 597
30	2.158 575	2.021 006	2.286 812	2.368 369
35	2.326 253	2.116 553	2.458 462	2.508 772
41	2.464 467	2.298 747	2.611 653	2.633 815

Although the distribution results were comparatively disappointing for the empirical data, we conducted an experiment to compare the actual mean lifetime and mean residual lifetime with the results of our models for various failure threshold levels. We computed the mean residual lifetimes by conditioning on the event that a specimen survives beyond the observed, unconditional mean lifetime. For each threshold value, a new generator matrix and degradation rates were estimated using $K = 3$ environmental states. The mean lifetimes over the 68 degradation paths were compared to those computed by Equation (14). The mean residual lifetimes were computed by Equation (15). The initial measured crack length was approximately 9.0 mm (i.e., $X(0) \approx 9.00$ mm). Table 7 summarizes the results of this experiment.

The results of this experiment indicate that the procedure adequately estimates the mean lifetimes and the mean residual lifetimes. Although not as informative as the residual life distribution, the lower moment approximations may be useful for constructing surrogate parametric distributions.

6. Conclusions

In this paper, we have presented a novel approach for the estimation of full and residual lifetime distributions (and moments) for single-unit systems subject to a Markov environment. The approach provides a necessary first step toward a formal, analytical technique for remaining lifetime prognosis via real degradation measures. Knowledge of the residual lifetime distribution is especially useful for prescribing policies that may reduce the risk of catastrophic failures while eliminating superfluous preventive maintenance activities.

Our approach is novel in that we combine analytical stochastic modelling techniques with real sensor data in two distinct cases. First, we considered the case when the sensor data provide real-time information regarding the current

state of the environment in which the single-unit system resides. Second, we considered the case in which the sensors provide the current level of degradation of the component. In the first case, our numerical results show great promise for a fully-automated technique for the estimation of lifetime distributions under our problem assumptions. In the second case, we demonstrated that our technique could be used to compute the distributions given nothing more than degradation measures (assumed to be perfectly observable). Although our procedure for estimating residual lifetime distributions performed only moderately well on the empirical data of Virkler *et al.* (1979), the comparison of mean residual lifetimes is very promising.

The techniques presented in this research have raised a few important questions that require further study. First, it may be possible that the time between environment transitions is distinctly non-exponential. In such a case, the engineer should resort to phase-type approximations for the state holding time distributions as outlined by Altiok (1985). This approach has the benefit of retaining the Markov property of the environmental process so that the main results of this paper (i.e., Equations (5) and (6)) may still be applied. The disadvantage of this approach is that it requires an expansion of the environment state space. Second, there is the issue of non-stationarity of the environment process. It will be instructive to consider the problem of estimating the parameters of an age-dependent environment process model. Due to extensive data requirements, it appears that this extension may only be feasible if the failure dynamics of the material under consideration are known. Finally, in order to facilitate a fully-automated prognostic system, it will be necessary to develop more robust techniques for evaluating the degradation rates. Such techniques may need to account explicitly for the effects of external covariates as noted in Meeker *et al.* (2002).

Acknowledgements

The authors acknowledge, with gratitude, the helpful comments of three anonymous referees and the Associate Editor. This research was supported by the Air Force Office of Scientific Research under agreement QAF185045200004.

References

Abate, J., Choudhury, G. and Whitt, W. (1998) Numerical inversion of multidimensional Laplace transforms by the Laguerre method. *Performance Evaluation*, **31**, 229–243.

Abate, J. and Whitt, W. (1995) Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, **7**, 36–43.

Abdel-Hameed, M. (1984) Life distribution properties of devices subject to Lévy wear process. *Mathematics of Operations Research*, **9**, 606–614.

Altiok, T. (1985) On the phase-type approximations of general distributions. *IIE Transactions*, **17**(2), 110–116.

Basawa, I.V. and Rao, P. (1980) *Statistical Inference for Stochastic Processes*, J. Wiley, New York, NY.

Bogdanoff, J.L. and Kozin, F. (1985) *Probabilistic Models of Cumulative Damage*, J. Wiley, New York, NY.

Calinski, R.B. and Harabasz, J. (1974) A dendrite method for cluster analysis. *Communications in Statistics*, **3**, 1–27.

Çinlar, E. (1977) Shock and wear models and Markov additive processes, in *The Theory and Applications of Reliability*, Shimi, I.N. and Tsokos, C.P. (eds.), Academic Press, New York, pp. 193–214.

Esary, J.D., Marshall, A.W. and Proschan, F. (1973) Shock models and wear processes. *Annals of Probability*, **1**, 627–649.

Fraley, C. and Raftery, A. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, **41**(8), 578–588.

Gertsbakh, I.B. and Kordonsky, K.B. (1969) *Models of Failure*, Springer-Verlag, New York, NY.

Gillen, K.T. and Celina, M. (2001) The wear-out approach for predicting the remaining lifetime of materials. *Polymer Degradation and Stability*, **71**, 15–30.

Kharoufeh, J.P. (2003) Explicit results for wear processes in a Markovian environment. *Operations Research Letters*, **31**(3), 237–244.

Lee, M.T., DeGruttola, V. and Schoenfeld, D. (2000) A model for markers and latent health status. *Journal of the Royal Statistical Society B*, **62**(4), 747–762.

Lu, C.J. and Meeker, W.Q. (1993) Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, **35**(2), 161–174.

Meeker, W.Q. and Escobar, L.A. (1998) *Statistical Methods for Reliability Data*, J. Wiley, New York, NY, pp. 316–342.

Meeker, W.Q., Escobar, L.A. and Chan, V. (2002) Using accelerated tests to predict service life in highly variable environments, in *Service Life Prediction: Methodologies and Metrologies*, Bauer, D.R. and Martin, J.W. (eds.), American Chemical Society, Washington, DC.

Moorthy, M.V. (1995) Numerical inversion of two-dimensional Laplace transforms-Fourier series representation. *Applied Numerical Mathematics*, **17**, 119–127.

Ray, A. and Tangirala, S. (1997) A nonlinear stochastic model of fatigue crack dynamics. *Probabilistic Engineering Mechanics*, **12**, 33–40.

Singpurwalla, N.D. (1995) Survival in dynamic environments. *Statistical Science*, **10**, 86–103.

Virkler, D.A., Hillberry, B. M. and Goel, P.K. (1979) The statistical nature of fatigue crack propagation. *ASME Journal of Engineering Materials and Technology*, **101**(2), 148–153.

Whitmore, G.A., Crowder, M.J. and Lawless, J.F. (1998) Failure inference from a marker process based on a bivariate Wiener process. *Lifetime Data Analysis*, **4**, 229–251.

Biographies

Jeffrey P. Kharoufeh is an Assistant Professor in the Department of Operational Sciences at the Air Force Institute of Technology in Dayton, OH. He holds a Ph.D. in Industrial Engineering and Operations Research from The Pennsylvania State University. His research interest is the development and analysis of stochastic models in reliability theory, queueing theory, and transportation systems. He is a professional member of IIE and INFORMS.

Steven M. Cox is a Ph.D. candidate in the Department of Operational Sciences at the Air Force Institute of Technology in Dayton, OH. He holds a B.S. in Mathematics from the US Air Force Academy and a M.S. in Operations Research from the Air Force Institute of Technology. He is a student member of INFORMS.

Contributed by the Reliability Engineering Department