Semi-Markov Models for Degradation-Based Reliability

Jeffrey P. Kharoufeh[1]
Department of Industrial Engineering
University of Pittsburgh
1048 Benedum Hall
3700 O'Hara Street
Pittsburgh, PA 15261 USA
Ph: (412) 624-9832; Fax: (412) 624-9831
Email: jkharouf@pitt.edu

Christopher J. Solo
Department of Mathematical Sciences
U.S. Air Force Academy
2354 Fairchild Drive, Rm 6D-198
United States Air Force Academy, CO 80840 USA
Ph: (719) 333-7774
E-mail: christopher.solo@usafa.edu

M. Yasin Ulukus
Department of Industrial Engineering
University of Pittsburgh
1048 Benedum Hall
3700 O'Hara Street
Pittsburgh, PA 15261 USA
Email: myu1@pitt.edu

Accepted Version: September 2009
Forthcoming in *IIE Transactions*

Abstract

We present hybrid, degradation-based reliability models for a single-unit system whose degradation is driven by a semi-Markov environment. The primary objective is to develop a mathematical framework and associated computational techniques that unite environmental data and stochastic failure models to assess the current or future health of the system. By employing phase-type distributions, we construct a surrogate environment process that is amenable to analysis by exact Markovian techniques to obtain reliability estimates. By way of two numerical experiments, we illustrate the viability of our approach and assess the quality of the approximations. The numerical results indicate that remarkably accurate lifetime distribution and moment approximations are attainable.

Keywords: Reliability, semi-Markov environment, phase-type distributions.

[1]Author to whom correspondence should be addressed.

1 Introduction

Recent advances in sensor technologies have dramatically accelerated the use of sensors for continuously monitoring critical engineering components in a host of applications (e.g., manufacturing equipment, aircraft components, structures, roadway pavement, power grids, etc.). The advantage of using sensor data to estimate the current or future health of a component or system is obvious – such data obviate the need for failure-time observations which may be scarce. For example, failure-time data may be very limited when systems are highly reliable or when it is prohibitively expensive to run systems to failure. Consequently, many researchers have been advocating *degradation-based* techniques as an alternative to the failure-based paradigm in order to exploit a plethora of degradation-related data that is now attainable through advanced sensing technologies. However, in some applications, the degradation may be difficult to measure, but the cumulative degradation can often be characterized as a function of the environment in which the component resides or the conditions under which it operates. One example is the degradation of protective chemical coatings that are exposed to time-varying environmental conditions. If the environment can be characterized appropriately (by some stochastic process), then it may be possible to characterize the cumulative degradation as a function of time. However, translating environmental data into useful lifetime distributions for reliability estimation remains a significant challenge. Therefore, there exists a critical need for novel techniques that can map environmental sensor data to degradation-based reliability indices.

In this paper, we present hybrid, degradation-based reliability models for a single-unit system whose degradation is driven by a semi-Markov environment. The primary objective is to develop a mathematical framework and associated computational techniques for uniting environmental data and stochastic failure models to assess the current or future health of the system. Specifically, we develop a general procedure that uses information about the sensed environment to estimate important reliability indices (e.g., the reliability function, the residual lifetime distribution, the mean time-to-failure, and the residual mean time-to-failure). Our procedures are most suitable when either (i) discrete (and distinct) environment states can be identified in a natural way, or (ii) discrete states can be constructed within a continuous state space in a natural way. The framework extends a host of so-called *random environment* models to the case of semi-Markovian environments that place only mild restrictions on the dynamics of the evolving environment. Moreover, unlike most existing stochastic failure models of this type, we devise a procedure for using observed data

to estimate the models' parameters and to produce reliability estimates. The need to estimate reliability indices within a *non-failure-based* paradigm has become critical as complex systems are increasingly reliable and prohibitively costly to run to failure.

The literature related to environment- and degradation-based reliability, while relatively immature as compared to that of failure-based reliability, has been growing at a rapid pace. Generally, work in this area can be partitioned into two categories broadly construed as physics- and statistics-based models and probabilistic (or stochastic) models. Physics-of-failure models are typically deterministic models that attempt to capture the basic physical principles underlying the failure process. Some examples are provided in [11, 15, 31, 36]. Statistics-based models are usually developed by collecting data over a period of time that includes a reasonable number of system failures, and then using these data to form the basis of a statistical model that predicts failure as a function of the sensor measurement. One example is the proportional hazard model (PHM) developed by Cox in 1972 to analyze medical survival data [9, 10]. The PHM has been implemented in a variety of engineering applications, such as aircraft, marine systems, and machinery [19, 20, 21, 49]. An excellent review of proportional hazard models for preventive maintenance can be found in [28]. Gebraeel, et al. [17, 18] proposed models that use actual signals to fit linear, exponential, and polynomial degradation curves from which they estimated normal residual lifetime distributions. The results were compared with real degradation signals using frequency measurements from rolling bearings. Both physics-of-failure models (by definition) and statistical models (due to the experimental methods used to develop them) are limited to the specific system under consideration. Furthermore, because it may be difficult and/or prohibitively costly to run systems to failure, their development and implementation may be very time consuming, as noted by Elsayed [14].

Techniques of the second type are typically based on stochastic shock and wear models, as well as models for systems that evolve in a random environment. Early work due to Esary, et al. [16] provided several results for both wear and shock processes. Çinlar [5, 6] generalized many of the models of [16] by showing that the joint process of the unit's wear level and the state of its ambient environment may be viewed as a Markov-additive process and provided several examples. The first considered the case when the random environment is a finite Markov process, and the wear is assumed to increase as a Lévy process. Random shocks were also assumed to occur at environment transition epochs. The second, which is similar to our approach, views cumulative wear as a continuous, additive functional of the operating environment, and the first time to failure

is a first passage time for the cumulative wear process. Other stochastic failure models also attempt to capture the impact of a randomly varying environment. An excellent survey of such models is due to Singpurwalla [44]. Li and Luo [32] considered a Markov-modulated shock process wherein the shock inter-arrival times and the random shock damage are both governed by a Markov chain. They obtain reliability bounds when the inter-arrival times have heavy- or light-tailed distributions, but their degradation model does not include a continuous wear component. Mallor and Omey [34] considered a generalized shock process and studied some asymptotic properties. Kharoufeh et al. [23, 24] presented degradation models that assume a Markovian environment. Klutke et al. [27] examined the availability of an inspected system whose inter-inspection times and wear rates are random. Kiessler et al. [26] investigated the limiting average availability of a system whose time-varying wear rates are governed by a continuous-time Markov chain. Kharoufeh et al. [25] extended the model of [26] by including damage-inducing shocks that arrive according to a time-homogeneous Poisson process and deriving the Laplace-Stieltjes transforms of transient reliability indices. Ebrahimi [12] considered a system whose degradation is comprised of a continuous wear component as well as jumps. The properties of the model were investigated and bounds were established for the reliability function. Recently, Özekici and Soyer have analytically examined reliability indices in both Markov and semi-Markov environments [39, 40, 41].

The present paper extends the models of [17, 18, 24, 26] in very important ways. First, it incorporates environmental data for the purpose of evaluating reliability indices by linking this data to degradation. Second, it generalizes the models of [23, 24, 26] by assuming that the environment *does not* evolve as a Markov process, but rather as a semi-Markov process. This generalization allows us to consider environment dynamics that are not restricted to memoryless holding times in individual environment states. Third, it provides a viable, data-driven approach for translating environment observations into failure time estimates using analytical stochastic failure models. Specifically, we exploit phase-type (PH) distributions to approximate environment state holding times, thereby inducing a Markovian structure that is amenable to exact analysis by the methods described in [24, 25]. We provide guidance on selecting an appropriate PH distribution approximation for each environment holding time. Finally, using simulated benchmarks, we illustrate the viability of our approach, and the remarkably high quality of our approximations, by way of two numerical experiments. The key innovation of our work is the use of environmental data to compute lifetime distributions within a degradation-based paradigm.

The remainder of the paper is organized as follows. Section 2 describes the semi-Markov envi-

ronment model and the stochastic failure models. In section 3, we briefly review PH distributions and provide decision rules for selecting a PH approximation. Section 4 formally describes our procedure for converting the non-Markovian environment into one that is amenable to exact analysis by Markovian techniques. Section 5 presents two randomized experiments that illustrate the viability of our approach, while section 6 provides some concluding remarks and directions for future work.

2 Model Description

This section describes our mathematical model of a single-unit system (i.e., a component) subject to continuous, environment-driven degradation. The component is placed into service at time zero in perfect working order and degrades at a rate the depends on the state of its environment which evolves as a continuous-time stochastic process with finitely many discrete states. When the component's cumulative degradation reaches a deterministic, critical threshold value x, the system is said to be failed. We denote the random time to first reach x by T_x, and we assume that T_x is proper (i.e., as $t \to \infty$, $\mathbb{P}(T_x \leq t) \to 1$ for each $x > 0$). The time-varying degradation rate is modulated by an external stochastic process commonly termed the *random environment*.

The random environment is denoted by $Z \equiv \{Z(t) : t \geq 0\}$ with finite state space $S \equiv \{1, 2, \ldots, K\}$ and $2 \leq K < \infty$. The process visits some state $i \in S$ and spends a random amount of time there that depends on the next state it will visit, say $j \in S$, $j \neq i$. It chooses j according to a Markov chain with transition probability matrix $\boldsymbol{P}$. The time spent in state i, given that the process next visits state j, has cumulative distribution function (c.d.f.) H_{ij}. Let S_n denote the time of the nth transition of the environment process, and let $E_n \equiv Z(S_n+)$ be the state of Z just after the nth transition epoch. The environment is a temporally homogeneous semi-Markov process (SMP) with associated Markov renewal sequence $\{(E_n, S_n) : n \geq 0\}$ on the state space S. For each $t \geq 0$, $n \geq 0$, and $i, j \in S$, define the probabilities

$$G_{i,j}(t) = \mathbb{P}(E_{n+1} = j, S_{n+1} - S_n \leq t | E_n = i),$$

and the corresponding semi-Markov kernel matrix $\boldsymbol{G}(t) = [G_{i,j}(t)]$. The environment transitions from state i to state $j \neq i$ according to a Markov chain on S with transition probability matrix $\boldsymbol{P}$ where the (i, j)th element of $\boldsymbol{P}$ is given by

$$p_{i,j} \equiv \mathbb{P}(E_1 = j | E_0 = i) = \lim_{t \to \infty} G_{i,j}(t). \tag{1}$$

Although the transition probabilities ($\{p_{i,j}\}$) can be obtained by evaluating the limit of (1), the

kernel functions $G_{i,j}(t)$ may not be known in practice, or they may be difficult to characterize. We circumvent this shortcoming by estimating $p_{i,j}$ from observable data in section 4.

When $Z(t) = i \in S$, the c.d.f. of the environment's holding time in state i, given that it next visits state $j \neq i$, is

$$H_{i,j}(t) = \mathbb{P}(S_{n+1} - S_n \leq t | E_{n+1} = j, E_n = i) = \mathbb{P}(S_1 \leq t | E_1 = j, E_0 = i)$$

where the second equality implies temporal homogeneity. Therefore, the c.d.f. of the holding time in state i, independent of the next state, is

$$H_i(t) = \mathbb{P}(S_1 \leq t | E_0 = i) = \sum_{j \in S} G_{i,j}(t), \quad i \in S.$$

For $t \geq 0$ and $i, j \in S$, define the transition functions of Z by

$$\pi_{i,j}(t) \equiv \mathbb{P}(Z(t) = j | Z(0) = i),$$

and the transition matrix $\boldsymbol{P}(t) = [\pi_{i,j}(t)]$. It can be shown (cf. [29]) that $\pi_{i,j}(t)$ satisfies the Markov renewal equation

$$\pi_{i,j}(t) = \bar{H}_i(t)\, \mathbf{1}_{\{i=j\}} + \sum_{k \in S} \int_0^t \pi_{k,j}(t - v)\, G_{i,k}(\mathrm{d}v) \tag{2}$$

where $\bar{H}_i(t) = 1 - H_i(t)$, the indicator function $\mathbf{1}_{\{i=j\}}$ assumes the value 1 if $i = j$ and 0 otherwise, and the integral on the right-hand side of (2) is the convolution of $\pi_{k,j}$ with $G_{i,k}$. In general, it is difficult to obtain the transition functions, even when the kernel matrix is known.

Next, define a rate function $r : S \to (0, \infty)$ such that whenever $Z(t) = i$, the system degrades at a constant rate $r(i)$, $r(i) > 0$ for $i = 1, 2, \ldots, K$. That is, the component's degradation rate is semi-Markov modulated. We pause here to make two remarks. First, it is important to note that our technique does not require, or assume, that the overall degradation path is linear in form. Rather, the imposed assumption is that the rate of degradation is constant *within a given environment state*. This framework allows us to characterize the degradation process by a set of mean rates of degradation as a function of time (or usage). The environment-modulated rates lead to piece-wise linear sample paths of degradation that can approximate common degradation patterns (e.g., linear, exponential, polynomial, or other forms). For example, once a crack has been initiated in a metallic specimen, the crack grows exponentially in the number of load cycles (see Virkler et al. [47]). The degradation pattern can generally be represented by three distinct regions. The first region exhibits mild linear growth pattern until reaching an inflection point at which the degradation grows at a moderately higher rate. After growing moderately for some time, the

degradation path reaches a second inflection point beyond which the crack growth rate is very high until reaching a critical failure threshold (e.g., the fracture point of the material). It was shown in [24] that such a non-linear degradation path can be approximated using a piece-wise linear path with three distinct rates associated to the three phases of crack length growth. They developed a statistical procedure to (i) estimate the appropriate number of environment states, and (ii) estimate the degradation rate in each environment state. In practice, the degradation rate associated with each environment state must be estimated with great care – in consultation with subject matter experts – to ensure consistency with physical principles. Second, the assumption of strictly positive degradation rates can be justified as follows. It is presumed that, whenever the unit is in operation, degradation accrues to at least some degree, i.e., the component experiences day-to-day wear from normal usage. On the other hand, in some scenarios, even if the system is not in use, the ambient environment might induce degradation (e.g., the degradation of protective chemical coatings that are exposed to the elements). We do not consider environment states that improve the condition of the component.

The cumulative degradation up to time t, denoted by $X(t)$, is

$$X(t) = X(0) + \int_0^t r[Z(u)]\mathrm{d}u, \tag{3}$$

where we assume $X(0) \equiv 0$, and

$$\int_0^t |r[Z(u)]|\mathrm{d}u < \infty \quad \text{a.s.}$$

so that $X(t)$ is well defined for each $t \geq 0$. The process, $X \equiv \{X(t) : t \geq 0\}$, is termed the *cumulative degradation process*. The positivity of the degradation rates, $r(1), r(2), \ldots, r(K)$, ensures that the sample paths of X are almost surely monotone increasing, and consequently, that events $\{X(t) \leq x\}$ and $\{T_x \geq t\}$ are equivalent. The system's random lifetime is given by

$$T_x = \inf\{t > 0 : X(t) \geq x\}, \tag{4}$$

or the first time the degradation process X reaches x. Let

$$F(x,t) \equiv \mathbb{P}(T_x \leq t) = 1 - \mathbb{P}(X(t) \leq x)$$

denote the unconditional c.d.f. of the unit's lifetime, and let

$$F_i(x,t) \equiv \mathbb{P}(T_x \leq t | Z(0) = i) = 1 - \mathbb{P}(X(t) \leq x | Z(0) = i)$$

be the conditional c.d.f. of T_x, given the initial state of the environment. Denote by $\mathbb{E}[T_x^n]$ the nth moment of T_x, for $n \geq 1$, and let its conditional counterpart be denoted by $\mathbb{E}_i[T_x^n] \equiv \mathbb{E}[T_x^n | Z(0) = i]$.

For a semi-Markov environment process Z, computing $F(x,t)$ and $F_i(x,t)$ is nontrivial due to the difficulty in obtaining $\pi_{i,j}(t)$ of (2). However, if Z is a continuous-time Markov chain (CTMC) on the state space S, the Laplace-Stieltjes transforms (LSTs) of $F(x,t)$, $F_i(x,t)$, $\mathbb{E}[T_x^n]$ and $\mathbb{E}_i[T_x^n]$, with respect to x, can be derived explicitly (see [24, 25]). Let $\boldsymbol{\alpha}$ be the initial distribution vector of the environment, $\boldsymbol{e}$ is a column vector of ones, and $\boldsymbol{e}_i$ is a column vector whose ith entry is unity and all others are zero. The unconditional, full lifetime distribution is given by

$$F(x,t) = \mathbb{P}(T_x \leq t) = 1 - \boldsymbol{\alpha}\mathbf{V}(x,t)\,\boldsymbol{e}$$

where $\mathbf{V}(x,t) = [V_{i,j}(x,t)]$ is a $K \times K$ matrix with

$$V_{i,j}(x,t) = \mathbb{P}(X(t) \leq x, Z(t) = j | Z(0) = i),$$

the joint probability that, at time t, the degradation of the system has not exceeded x, and the environment process is in state $j \in S$, given that the environment was initially in state $i \in S$. The conditional c.d.f. of T_x is

$$F_i(x,t) \equiv \mathbb{P}_i(T_x \leq t) = 1 - \boldsymbol{e}_i'\mathbf{V}(x,t)\boldsymbol{e} \tag{5}$$

where $\boldsymbol{e}_i'$ denotes the transpose of $\boldsymbol{e}_i$. Let the Laplace-Stieltjes transforms of $F(x,t)$ and $F_i(x,t)$, with respect to x, be

$$\widetilde{F}(u,t) \equiv \int_0^\infty e^{-ux} F(\mathrm{d}x,t), \quad \mathrm{Re}(u) > 0,$$

and

$$\widetilde{F}_i(u,t) \equiv \int_0^\infty e^{-ux} F_i(\mathrm{d}x,t), \quad \mathrm{Re}(u) > 0,$$

respectively. Using Theorem 3 of [25], it can be shown that

$$\widetilde{F}(u,t) = 1 - \boldsymbol{\alpha}\exp[(\mathbf{Q} - u\mathbf{R}_d)t]\boldsymbol{e}, \tag{6}$$

and

$$\widetilde{F}_i(u,t) = 1 - \boldsymbol{e}_i'\exp[(\mathbf{Q} - u\mathbf{R}_d)t]\boldsymbol{e} \tag{7}$$

where $\mathbf{Q}$ is the generator matrix of the CTMC, $\mathbf{R}_d = \mathrm{diag}(r(1), r(2), \ldots, r(K))$, and $\exp(A)$ denotes exponentiation of the square matrix A.

While these results are useful, they are applicable only when the environment evolves as a CTMC (i.e., when the environment spends an exponential time in each state before it transitions

to another state). However, in reality, these state holding times are general, non-negative random variables whose distributions are (in most cases) unknown. In fact, the assumption of memoryless holding time distributions may be completely unwarranted in many applications. Our aim is to provide a means by which to use sensor data, especially environment-related data, to characterize the holding time distributions and the environment's transition rates so that (6) or (7) can be directly applied to compute reliability indices. Our technique requires observations of the environment's current state, its transition times, and observation of its subsequent state. These will be used in an automated procedure to estimate holding time distributions by phase-type distributions. By doing so, we generalize the models presented in [17, 18, 23, 24, 26] to account for semi-Markovian environments. Sections 3 and 4 describe our approach for approximating the c.d.f. $F(x,t)$ (or $F_i(x,t)$) and its moments within the semi-Markov framework presented in this section.

3 Phase-Type Approximations

Here, we review the most commonly used techniques for approximating the c.d.f. of a non-negative random variable by a phase-type distribution. For a complete introduction to phase-type distributions, the reader is referred to Neuts [37]. A non-negative random variable, Y, is said to have a phase-type (PH) distribution if it represents the time to absorption of some Markov chain. If the chain evolves in discrete (continuous) time, then Y is a discrete (continuous) PH random variable. Owing to the fact that our environment evolves in continuous time, we review only the continuous version here.

Let $\{\phi(t) : t \geq 0\}$ be a continuous-time Markov chain (CTMC) with finite state space $M = \{1, 2, \ldots, k+1\}$ where states $1, 2, \ldots, k$ are transient, and state $k+1$ is absorbing. The infinitesimal generator matrix of $\{\phi(t) : t \geq 0\}$ is

$$\mathbf{Q}^* = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{T}^0 \\ \mathbf{0} & 0 \end{bmatrix}, \tag{8}$$

where $\boldsymbol{T}$ is a $k \times k$ matrix with $T_{i,i} < 0$ and $T_{i,j} \geq 0$, $j \neq i$, $\boldsymbol{T}^0$ is a column vector, $\mathbf{0}$ is the zero vector, and $\boldsymbol{T}\boldsymbol{e} + \boldsymbol{T}^0 = \mathbf{0}$. The vector $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}, \beta_{k+1})$ is the initial distribution vector of $\mathbf{Q}^*$, i.e., $\boldsymbol{\beta}$ is a $1 \times k$ row vector, and β_{k+1} is the probability that the Markov process begins in the absorbing state $k+1$. We assume throughout that $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}, 0)$. The vector $\boldsymbol{\beta}_0$ satisfies the usual condition $\boldsymbol{\beta}_0 \, \boldsymbol{e} = 1$. Let Y be the time to absorption of $\mathbf{Q}^*$. It is not difficult to verify (cf. Neuts [37]) that

the c.d.f. of Y is

$$F(y) \equiv \mathbb{P}(Y \leq y) = 1 - \boldsymbol{\beta} \exp(\boldsymbol{T}y)\, \boldsymbol{e}, \tag{9}$$

and the nth moment of Y is

$$\mathbb{E}[Y^n] = (-1)^n\, n!\, \boldsymbol{T}^{-n}\, \boldsymbol{e}, \quad n \geq 1.$$

The distribution function (9) is a PH distribution with representation $(\boldsymbol{\beta}, \boldsymbol{T})$; therefore, constructing a PH approximation of the c.d.f. of a non-negative random variable X involves determining the pair $(\boldsymbol{\beta}, \boldsymbol{T})$. PH distributions are attractive because they can approximate the c.d.f. of any non-negative random variable. Moreover, they can facilitate tractability when analyzing non-Markovian stochastic processes. That is, by supplementing the original non-Markov process with the phase variable, $\phi(t)$, one can obtain a Markovian structure so that standard analysis techniques for Markov processes can be employed (cf. [3, 38, 42, 48]). For these two primary reasons, we will employ PH distributions to approximate the environment state holding time distributions, $H_i(t)$, $i = 1, 2, \ldots, K$. By doing so, we will replace our semi-Markov environment by a surrogate (expanded) environment that is Markovian so that (6) and (7) can be used to compute reliability indices.

3.1 Specific PH Distributions

The general k-phase PH distribution (denoted PH_k) has an associated Markov process with k states (or phases) and a single absorbing state $k+1$. The time spent in a transient state i is exponentially distributed with parameter μ_i, $i = 1, 2, \ldots, k$. Once state $k+1$ is reached for the first time, the process is absorbed. There are no restrictions on the transitions of the process until the time of absorption. That is, a visit to state i, $1 \leq i \leq k$, can be followed by a visit to any other transient state before absorption

By contrast, the k-phase Coxian distribution (denoted C_k) is the distribution of the time to absorption of a Markov chain with k phases that are visited *sequentially*, but absorption is possible at the end of each phase. As for the PH_k distribution, the time spent in phase i is exponential with rate μ_i, $i = 1, 2, \ldots, k$, so this PH distribution has the representation

$$\boldsymbol{T} = \begin{pmatrix} -\mu_1 & a_1\mu_1 & 0 & 0 & 0 \\ 0 & -\mu_2 & a_2\mu_2 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & -\mu_{k-1} & a_{k-1}\mu_{k-1} \\ 0 & 0 & \ldots & 0 & -\mu_k \end{pmatrix} \tag{10}$$

with $\boldsymbol{\beta} = (1, 0, \ldots, 0)$ and $\boldsymbol{T}^0 = ((1-a_1)\mu_1, (1-a_2)\mu_2, \ldots, (1-a_{k-1})\mu_{k-1}, \mu_k)'$, where a_i is the transition probability from i to $i+1$, $i = 1, 2, \ldots, k-1$, and $1 - a_i$ is the probability of absorption immediately following phase i. The Coxian distribution is especially attractive because it can exactly represent any distribution that has a rational Laplace-Stieltjes transform [8].

The k-phase Erlang distribution (denoted E_k) is a special case of the C_k distribution in which $\mu_i = \mu$ for $i = 1, 2, \ldots, k$, and all the transient phases must be visited sequentially before being absorbed into phase $k+1$ (i.e., $a_i = 1$, $i = 1, 2, \ldots, k-1$) so that absorption is not possible until after the kth phase has been completed. It is clear that the k-phase Erlang is the sum of k i.i.d. exponential random variables, each having parameter μ. Its representation $(\boldsymbol{\beta}, \boldsymbol{T})$ is given by

$$\boldsymbol{T} = \begin{pmatrix} -\mu & \mu & 0 & 0 & 0 \\ 0 & -\mu & \mu & 0 & 0 \\ 0 & 0 & -\mu & \mu & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \ldots & -\mu \end{pmatrix} \tag{11}$$

with $\boldsymbol{\beta} = (1, 0, \ldots, 0)$ and $\boldsymbol{T}^0 = (0, 0, \ldots, 0, \mu)'$.

The k-phase *generalized* Erlang distribution, as defined by Marie [35], is identical to the k-phase Erlang except that absorption is possible at the end of phase 1. That is, $\{\phi(t) : t \geq 0\}$ transitions from phase 1 to phase 2 with probability a, and is absorbed after phase 1 with probability $1 - a$. However, if phase 2 is reached, all subsequent phases are visited sequentially until absorption into state $k+1$. Therefore, the representation of this PH distribution is

$$\boldsymbol{T} = \begin{pmatrix} -\mu & a\mu & 0 & 0 & 0 \\ 0 & -\mu & \mu & 0 & 0 \\ 0 & 0 & -\mu & \mu & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \ldots & -\mu \end{pmatrix} \tag{12}$$

with $\boldsymbol{\beta} = (1, 0, \ldots, 0)$ and $\boldsymbol{T}^0 = ((1-a)\mu, 0, \ldots, 0, \mu)'$.

Any one of the PH_k, C_k, E_k, and k-phase generalized Erlang distributions can be used to approximate the c.d.f. of an arbitrary, non-negative random variable. However, selecting the best distribution for a given application is not straightforward. Next, we provide some guidelines for choosing an appropriate approximation.

3.2 Phase-type Approximation Selection

PH distribution selection is a critical aspect of our proposed technique, and three important criteria must be considered when making the selection. First, the approximations must use the least possible number of phases while adequately representing the environment sojourn times. Second, the approximation should accommodate an automated procedure so that only observed sensor data is needed to construct the approximation. Finally, the computational effort should be minimal.

Suppose that X is a non-negative random variable with unknown c.d.f. H. We want to approximate X by a PH random variable, say Y, with c.d.f. $\hat{H}$. Marie [35], Altiok [3], Johnson [22], Perros [42], and Osogami and Harchol-Balter [38] all provide excellent summaries of techniques that can be used to obtain PH-distribution approximations. With the exception of [38], most techniques require an estimate of the squared coefficient of variation of X, $c^2 = \sigma^2/m_1^2$, where m_1 and σ^2 are the mean and variance of X, respectively. (In practice, we generally will not have m_1 and σ^2 at our disposal; therefore, their sample estimators must be used to estimate c^2.) Subsequently, one may choose moment-matching, maximum likelihood, or so-called minimum-distance techniques. It is important to note that maximum likelihood and minimum-distance techniques require solving a nonlinear optimization problem, whereas moment-matching techniques require only the true, or estimated, lower moments of X. For this reason, we consider only moment-matching techniques. In what follows, let m_j, $j = 1, 2, 3$, denote the jth moment of X.

Osogami and Harchol-Balter [38] provide a nice summary of the most promising moment-matching techniques, the majority of which match the first two or three moments of X to the moments of a PH distribution. There is little dispute (see [2]) that the k-phase Erlang distribution provides very reliable approximations when $c^2 \leq 0.5$. However, when $c^2 > 0.5$, it is not obvious which technique is best. For example, Marie [35] uses two moments to match a generalized Erlang when $c^2 < 1$, and matched two moments to a C_2 distribution when $0.5 \leq c^2 \leq 1.0$. A two-moment matching algorithm using a k-phase Erlang was used for the range $0 < c^2 \leq 0.5$. Telek and Heindl [46] matched three moments of X to their two-phase canonical acyclic PH-distribution when $0.5 \leq c^2 \leq 1$. However, for this range, their technique requires bounds on m_3 of the form

$$3m_1^3(3c^2 - 1 + \sqrt{2}(1 - c^2)^{\frac{3}{2}}) \leq m_3 \leq 6m_1^3c^2.$$

By contrast, Marie [35] estimates the three parameters of C_2 using the first two moments, m_1 and m_2, with

$$\mu_1 = \frac{2}{m_1}, \quad \mu_2 = \frac{1}{m_1c^2}, \quad a = \frac{1}{2c^2},$$

resulting in a C_2 distribution representation with $\boldsymbol{\beta} = (1, 0, 0)$ and

$$\boldsymbol{T} = \begin{pmatrix} -\mu_1 & a\mu_1 \\ 0 & -\mu_2 \end{pmatrix}.$$

Whitt [48] showed that, when $c^2 > 1$, it is important to match the first three moments to reduce the maximum relative error of the approximation. He used a two-branch hyper-exponential distribution, whereas Altiok [3] used a C_2 distribution, and Telek and Heindl [46] use a two-phase canonical acyclic PH-distribution while matching three moments. Whitt's [48] hyper-exponential distribution requires the estimation of four parameters while the techniques of [3, 46] require only three parameters. However, the latter two impose the requirement

$$m_3 > \frac{3(c^2+1)^2 m_1^3}{2}.$$

For our semi-Markov environment model, it is critical to minimize the number of parameters estimated from the observed data. Therefore, a PH distribution that uses a minimal number of phases is ideal. To date, the question of determining a minimal representation with respect to the class of all PH distributions remains open. However, Osogami and Harchol-Balter [38] recently mapped general distributions to representations that are minimal with respect to the class of *acyclic* PH distributions (to which the C_k and k-phase Erlang distributions belong) when matching three moments. As previously noted, Telek and Heindl [46] examined canonical forms, a subclass of acyclic PH-distributions, that also admit minimal representations. Note that Marie's [35] two-moment approximation for $0 < c^2 \leq 0.5$ provides a minimal representation when matching only two moments.

By considering the guidance provided by [3, 35, 38, 42, 48], Table 1 summarizes our criteria for selecting the PH approximation type based on the true, or estimated, value of c^2.

Table 1: PH approximation selection criteria.

Range of c^2	PH Approximation
$0 < c^2 < 0.5$	2-moment, k-phase generalized Erlang
$0.5 \leq c^2 \leq 1$	2-moment, 2-phase Coxian
$c^2 > 1$	3-moment, 2-phase Coxian

In what follows, we summarize the specific calculations needed to match moments using the three types of PH approximations of Table 1. This discussion is largely contained in [3, 35, 38, 42].

2-moment, k-Phase Generalized Erlang Distribution

When $0 < c^2 < 0.5$, we will approximate the state holding time distribution, H, by a 2-moment, k-phase generalized Erlang approximation where the integer k ($k > 1$) is chosen such that (see [42])

$$\frac{1}{k} \leq c^2 \leq \frac{1}{(k-1)}.$$

Recall that a is the probability that $\{\phi(t) : t \geq 0\}$ transitions from phase 1 to phase 2, and μ is the common rate of the k exponential phases. These two parameters are obtained by

$$a = 1 - \frac{2kc^2 + k - 2 - (k^2 + 4 - 4kc^2)^{1/2}}{2(c^2+1)(k-1)}, \tag{13}$$

and

$$\mu = \frac{1 + (k-1)a}{m_1}, \tag{14}$$

where m_1 must be estimated from observed data. Subsequently, a and μ are used directly in the representation of (12).

2-moment, 2-Phase Coxian Distribution

When $0.5 \leq c^2 \leq 1$, a 2-moment, 2-phase Coxian distribution is used to approximate H. The representation (10) requires the exponential rates μ_i, $i = 1, 2, \ldots, k$, and the absorption probabilities, $(1 - a_i)$. Following a moment-matching algorithm similar to that for the 3-moment, 2-phase Coxian distribution (discussed next), the parameters of the 2-moment variant of the 2-phase Coxian are

$$\mu_1 = 2/m_1; \quad \mu_2 = 1/m_1 c^2; \quad a = 1/2c^2,$$

where m_1 and c^2 are estimated from the observed data. It is important to note that using either variant of the 2-phase Coxian distribution as a PH approximation helps to limit the growth of the surrogate environment state space.

3-moment, 2-Phase Coxian Distribution

When $c^2 > 1$, the c.d.f. H is approximated by a 3-moment, 2-phase Coxian distribution. As noted by [3, 48], the third moment is significant when $c^2 > 1$. By taking successive derivatives of the Laplace-Stieltjes transform of the C_2 distribution, it has been shown in [3] that

$$m_1 \equiv \mathbb{E}[Y] = \frac{1}{\mu_1} + \frac{a}{\mu_2}, \tag{15}$$

$$m_2 \equiv \mathbb{E}[Y^2] = \frac{2(1-a)}{\mu_1^2} - \frac{2a\mu_1\mu_2 - 2a(\mu_1+\mu_2)^2}{\mu_1^2\mu_2^2}, \tag{16}$$

and

$$m_3 \equiv \mathbb{E}[Y^3] = \frac{6(1-a)}{\mu_1^3} - \frac{12a\mu_1\mu_2(\mu_1+\mu_2) - 6a(\mu_1+\mu_2)^3}{\mu_1^3\mu_2^3}. \tag{17}$$

As in [3], if we set $A = \mu_1 + \mu_2$ and $B = \mu_1\mu_2$, substitute, and simplify, the parameters of the 3-moment, 2-phase Coxian distribution are given by

$$\mu_1 = A + \frac{\sqrt{A^2 - 4B}}{2}; \quad \mu_2 = A - \mu_1; \quad a = \mu_1^{-1}\mu_2(m_1\mu_1 - 1). \tag{18}$$

We statistically estimate the values m_1, m_2, and m_3 and use these estimates in equations (15)–(18) to obtain parameters μ_1, μ_2, and a for use in representation (10).

In the next section, we present our formalized procedure for approximating environment state holding times and using these within a Markovian framework to obtain lifetime distribution approximations. Subsequently, the quality of these approximations will be assessed through numerical experiments in section 5.

4 Approximation Procedure

In this section, we formalize our procedure for constructing a surrogate Markovian environment process to replace the observed semi-Markov environment. To this end, let us clarify a few assumptions. First, we assume the availability of sensor data that provides information about the current state of the environment (e.g., load, speed, temperature, pressure, humidity, etc.) in which the unit resides, or under which it operates, while the degradation status is unobservable. Second, the environment is assumed to evolve as a finite, irreducible semi-Markov process (SMP) that directly influences the degradation process $\{X(t) : t \geq 0\}$; however, no assumptions are made regarding the probability law of the latter. Third, we assume that the environment's true state space can be partitioned into K distinct states such that $S = \{1, 2, \ldots, K\}$, $2 \leq K < \infty$. However, if K is not known *a priori*, it can be estimated by using a simple clustering scheme, such as that described in [24]. Associated with each environment state is a known, positive degradation rate, $r(j)$, so that when $Z(t) = j$, the system degrades linearly at rate $r(j)$. When Z can be modeled as a CTMC, equations (6) and (7) require the environment's initial distribution vector $\boldsymbol{\alpha}$, the diagonal matrix of degradation rates $\mathbf{R}_d$, and the infinitesimal generator matrix $\mathbf{Q}$. However, because we relax the Markovian assumption here, our aim is to obtain a surrogate Markovian environment $\widehat{Z}$ with

state space $\hat{S}$ and generator matrix $\widehat{\mathbf{Q}}$ by approximating each state holding time distribution by a 2- or k-phase PH distribution, depending on the estimated value of c^2, as described in section 3. Because each PH approximation uses at least two phases, we have that $|\hat{S}| > |S|$. For instance, if $|S| = 3$, and each holding time distribution is approximated by a 2-phase PH distribution, then the surrogate environment process $\widehat{\mathbf{Q}}$ will have $3 + 2 \times 3 = 9$ states. Note that the degradation rate matrix, $\mathbf{R}_d$, must also be replaced by $\widehat{\mathbf{R}}_d$ which has the same dimension as $\widehat{\mathbf{Q}}$.

Assume the homogeneous environment process $\{Z(t) : t \geq 0\}$ is perfectly observable over some time interval $[0, \tau]$, and that the unit has not failed in this interval. The environment is continuously observed up to time τ, and at each transition epoch, we record the current and subsequent state of the random environment. Let $R(i, j)$ denote the true (unknown) rate at which the environment transitions from state $i \in S$ to state $j \in S$, $j \neq i$. Let $N_\tau(i, j)$ be the number of transitions from i to j in time τ, let $N_\tau(i)$ be the number of visits to state i, and let $W_\tau(i)$ be the total time spent by Z in state i during $[0, \tau]$. It can be shown (cf. Basawa and Rao [4]) that, for each $i \in S$ and $j \neq i$, as $\tau \to \infty$,

$$\frac{N_\tau(i, j)}{W_\tau(i)} \to R(i, j) \quad \text{a.s.}$$

Therefore, for τ sufficiently large, we approximate $R(i, j)$ by

$$R(i, j) \approx \hat{R}_\tau(i, j) = \frac{N_\tau(i, j)}{W_\tau(i)}, \quad j \neq i, \tag{19}$$

and set

$$\hat{R}_\tau(i, i) = -\sum_{j \neq i} \hat{R}_\tau(i, j), \quad i \in S, \tag{20}$$

where $\hat{R}_\tau(i, i) < 0$. Now let $\widehat{\boldsymbol{P}}$ denote the estimated transition matrix of $\{E_n : n \geq 0\}$, the Markov chain embedded at environment transition epochs. The statistical estimator of the (i, j)th element of $\widehat{\boldsymbol{P}}$ is

$$\hat{p}_{i,j} = \begin{cases} \frac{\hat{R}_\tau(i,j)}{-\hat{R}_\tau(i,i)}, & \text{if } j \neq i, \\ 0, & \text{if } j = i. \end{cases} \tag{21}$$

For $n \geq 1$ and $i \in S$, let $D_n(i)$ be the duration of the environment's nth visit to state i, and assume that $D_n(i) \to D(i)$ weakly as $n \to \infty$, where $D(i)$ is the unconditional holding time in state i with proper c.d.f. H_i. For each $i \in S$, the kth moment and variance of $D(i)$ are estimated by

$$m_k(i) \equiv \frac{1}{N_\tau(i)} \sum_{n=1}^{N_\tau(i)} D_n^k(i), \quad k \geq 1, \tag{22}$$

and

$$s_i^2 \equiv \frac{1}{N_\tau(i)-1}\sum_{n=1}^{N_\tau(i)}\left[D_n(i)-m_1(i)\right]^2, \tag{23}$$

respectively. Note that (22) is the sample kth moment of the holding time in state i during $[0,\tau]$, and (23) is the sample variance of the holding time. The PH distribution selection procedure described in section 3 requires the sample squared coefficient of variation

$$\hat{c}_i^2 = \frac{s_i^2}{m_1^2(i)}, \quad i \in S. \tag{24}$$

We now formalize the full procedure to obtain $\widehat{Z}$, $\widehat{\mathbf{Q}}$, and $\widehat{\mathbf{R}}_d$.

Step 0: *Initialization.* Select a sufficiently large observation time τ.

Step 1: *Observe the Environment Process on* $[0,\tau]$

During the observation period, for each (i,j) such that $j \neq i$, record the number of transitions from $i \to j$, $N_\tau(i,j)$. Also record the duration of each visit to state i, $D_n(i)$ for $n = 1, 2, \ldots, N_\tau(i)$. The total time spent in state i during $[0,\tau]$ is approximated by

$$W_\tau(i) \approx \sum_{n=1}^{N_\tau(i)} D_n(i), \quad i \in S$$

since $W_\tau(i)$ is (almost surely) bounded above by the sum on the right-hand side.

Step 2: *Estimate the Required Parameters*

Using $N_\tau(i)$, $D_n(i)$, $n = 1, \ldots, N_\tau(i)$, $N_\tau(i,j)$, and $W_\tau(i)$, estimate the transition matrix $\widehat{\boldsymbol{P}}$ using equations (19)–(21). For each $i \in S$, compute the first three sample moments $m_k(i)$, $k = 1, 2, 3$, the sample variance s_i^2, and the sample squared coefficient of variation $\hat{c}_i^2$ using equations (22)–(24).

Step 3: *Select the PH Distribution Approximation for* H_i

Using $m_k(i)$, $k = 1, 2, 3$, and $\hat{c}_i^2$, $i \in S$, select the appropriate PH distribution approximation in accordance with Table 1. Using the moment-matching methods described in section 3, compute the matrix $\boldsymbol{T}_i$ for each $i \in S$. Note that $\boldsymbol{T}_i^0$ follows directly from $\boldsymbol{T}_i$ since $\boldsymbol{T}_i\boldsymbol{e} + \boldsymbol{T}_i^0 = \mathbf{0}$. Therefore, for each $t \geq 0$, the PH approximation of H_i is given by

$$\hat{H}_i(t) = 1 - \boldsymbol{\beta}\exp(\boldsymbol{T}_i\, t)\boldsymbol{e}, \quad i \in S.$$

Step 4: *Construct Surrogate Environment Process*

Suppose $|S| = K$ and that H_i is approximated by a PH distribution with k_i phases. That is, an absorbing Markov chain with $k_i + 1$ states replaces the original state i. Therefore, the surrogate Markovian environment, $\{\widehat{Z}(t) : t \geq 0\}$, has state space $\hat{S}$ with cardinality

$$|\hat{S}| = K + \sum_{i=1}^{K} k_i.$$

For this reason, it is important to use the smallest possible integer k_i for each $i \in S$. Next, construct the surrogate, $\hat{S}$-valued CTMC $\{\widehat{Z}(t) : t \geq 0\}$ with generator matrix $\widehat{\mathbf{Q}}$ given by

$$\widehat{\mathbf{Q}} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \cdots & \boldsymbol{A}_{1K} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \cdots & \boldsymbol{A}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{K1} & \boldsymbol{A}_{K2} & \cdots & \boldsymbol{A}_{KK} \end{pmatrix}$$

where

$$\boldsymbol{A}_{ii} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{T}_i^0 \\ \mathbf{0} & c\,\hat{R}(i,i) \end{pmatrix},$$

for $j \neq i$,

$$\boldsymbol{A}_{ij} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ c\,\hat{R}_\tau(i,j) & \mathbf{0} \end{pmatrix},$$

and c is a very large, positive real number. The parameter c is needed in the numerical implementation to ensure that $\{\widehat{Z}(t) : t \geq 0\}$ instantaneously transitions to the next state chosen by $\widehat{\boldsymbol{P}}$ when the environment's sojourn in state $i \in S$ is complete. Note that $\boldsymbol{T}_i$ is a $k_i \times k_i$ matrix while $\boldsymbol{T}_i^0$ is a k_i-dimensional column vector. The degradation rate matrix must also be expanded accordingly. The new matrix is of the form

$$\widehat{\mathbf{R}}_d = \begin{pmatrix} \Delta\left(r(1)\right) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Delta\left(r(2)\right) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Delta\left(r(K)\right) \end{pmatrix}$$

where $\Delta(a)$ is a diagonal matrix whose diagonal entries are all a.

Step 5: *Approximate the Lifetime Distribution Function and Moments*

By replacing the SMP environment $\{Z(t) : t \geq 0\}$ with state space S and generator $\mathbf{Q}$ by the CTMC environment $\{\widehat{Z}(t) : t \geq 0\}$ with state space $\hat{S}$ and generator $\widehat{\mathbf{Q}}$, the LST of the unconditional lifetime distribution is given by

$$\widetilde{F}(u,t) = 1 - \boldsymbol{\alpha} \exp[(\widehat{\mathbf{Q}} - u\widehat{\mathbf{R}}_d)t]\boldsymbol{e}, \tag{25}$$

and its conditional counterpart is

$$\widetilde{F}_i(u,t) = 1 - \boldsymbol{e}_i' \exp[(\widehat{\mathbf{Q}} - u\widehat{\mathbf{R}}_d)t]\boldsymbol{e}. \tag{26}$$

Moreover, the LST of the nth moment of T_x, conditioned on the initial state of the environment, is obtained by (see [25])

$$\widetilde{\mathbb{E}}_i[T_u^n] = n!\, \boldsymbol{e}_i' \left(u\widehat{\mathbf{R}}_d - \widehat{\mathbf{Q}}\right)^{-n} \boldsymbol{e}, \quad n \geq 1. \tag{27}$$

While the procedure described here appears somewhat involved, it has the advantage of requiring only event times and count data for the environment process. There is no need for failure time observations, or for degradation observations (besides those that will be needed initially to estimate the rates $r(j)$, $j \in S$). The next section provides two numerical experiments to assess the quality of our approximations and to illustrate the viability of the approach.

5 Numerical Experiments

This section provides two extensive numerical experiments to assess the quality of distribution approximations obtained by the procedures described in sections 3 and 4. In lieu of real data describing the evolution of a random environment, we randomly generated 1000 test scenarios for each of the two experiments. Specifically, we (i) chose holding time distributions (H_i) for SMP environments and randomized their parameter values, (ii) randomly generated the transition probability matrix ($\boldsymbol{P}$) of the DTMC embedded at transition epochs, and (iii) randomly generated the degradation rate ($r(i)$) associated with the ith environment state. To the outside observer, the SMP process that drives the environment's evolution is unknown. Therefore, we observed only the states visited by the environment and the time spent in each state over a time interval $[0, \tau]$. Subsequently, the statistical estimators described in section 4 were calculated. From these values, PH distribution approximations and the surrogate (Markovian) environment process were constructed to approximate the c.d.f. of T_x.

Let $\widehat{F}_i(x,t)$ denote the approximate lifetime c.d.f. generated by the surrogate environment $\{\widehat{Z}(t) : t \geq 0\}$ with generator matrix $\widehat{\mathbf{Q}}$ and degradation matrix $\widehat{\mathbf{R}}_d$. Denote by $F_i(x,t)$ the c.d.f. of T_x obtained by simulating the degradation process, $\{X(t) : t \geq 0\}$, until it first reaches the critical threshold value x. For each of the 1000 scenarios, and for both experiments, $F_i(x,t)$ was constructed using 20,000 observations of T_x, given that $Z(0) = i$. Therefore, we view the simulated c.d.f. as the completely specified, hypothesized benchmark. To compare $\widehat{F}_i(x,t)$ and $F_i(x,t)$, we employed the nonparametric two-sided Kolmogorov-Smirnov (KS) goodness-of-fit test which tests the hypothesis

$$H_0 : \widehat{F}_i(x,t) = F_i(x,t), \text{ for all } t \in \mathbb{R},$$
$$H_1 : \widehat{F}_i(x,t) \neq F_i(x,t), \text{ for at least one } t \in \mathbb{R}.$$

The distribution $\widehat{F}_i(x,t)$ was sampled on a finite set $T \subseteq \mathbb{R}_+ \equiv [0,\infty)$ whose range depends on observed simulated outcomes. The test statistic for the two-sided KS test (see [7], p. 431) is given by

$$\varepsilon = \sup_{t \in T} \left| F_i(x,t) - \widehat{F}_i(x,t) \right|,$$

the maximum absolute deviation between the two distributions over the set T. The null hypothesis, H_0, is rejected at the α level of significance if $\varepsilon > \varepsilon^*$ where ε^* is the $1-\alpha$ quantile of the test statistic. Let $n = |T|$ denote the cardinality of T. At the 0.05 level, the critical value ε^* is given by

$$\varepsilon^* = 1.36\, n^{-1/2}$$

if $n > 40$, and if $n \leq 40$, the critical value is obtained from [7] (Table A13, p. 547). All hypothesis tests were performed at the $\alpha = 0.05$ level of significance.

Numerical values of $\widehat{F}_i(x,t)$ were obtained via the inverse Laplace transform,

$$\widehat{F}_i(x,t) = \mathcal{L}^{-1}\left(u^{-1}\widetilde{F}_i(u,t)\right) \tag{28}$$

where $\widetilde{F}_i(u,t)$ is given by (26), and $\mathcal{L}^{-1}$ denotes the inverse Laplace transform operator. For illustrative purposes, we also assess the quality of the lower moment approximations given by

$$\widehat{\mathbb{E}}[T_x^n | Z(0) = i] = \mathcal{L}^{-1}\left(u^{-1}\widetilde{\mathbb{E}}_i[T_u^n]\right), \quad n \geq 1 \tag{29}$$

where $\widetilde{\mathbb{E}}_i[T_u^n]$ is given by (27). The inverse Laplace transforms (28) and (29) were obtained by coding a numerical Laplace transform inversion algorithm due to Abate and Whitt [1]. This algorithm and the discrete-event simulation models were coded in the MATLAB® computing environment and executed on a personal computer equipped with an Intel® Core™ 2 Duo CPU operating at 3.00 GHz with 2.00 GB of RAM.

5.1 Experiment 1: Degradation of a Turbine Blade

Consider an aircraft engine turbine blade that operates at high temperatures and experiences centrifugal stresses. The rotational speed of the blade varies depending on uncertain prevailing conditions. For example, the engine load varies as the aircraft is exposed to different flight conditions (i.e., takeoff, maximum climb, maximum cruise, loiter, flight idle, taxi, ground idle, and cutoff), or at different flight altitudes. Moreover, if inclement weather is encountered mid-flight, the aircraft may change altitude, accelerate, or decelerate to minimize the effects of turbulence. The induced centrifugal stresses can lead to elongated particles at the microstructure level that can result in degradation of fatigue strength that results in voids and crack initiation in the blade, and continued loading exacerbates the degradation. This phenomenon has been documented, for example, in [13, 43, 45]. The blade's dynamic and stochastic rotational speed can be viewed as a proxy for the induced stresses. Suppose the degradation process $\{X(t) : t \geq 0\}$ tracks the degradation level (e.g., the length of a crack) of the blade, and let $Z(t)$ denote the rotational speed of the blade at time t. We partition the possible range of speeds into four non-overlapping intervals that describe four distinct operating states (or flight conditions). The state space of the environment is $S = \{1, 2, 3, 4\}$. The critical threshold for the degradation level is $x = 20.0$ units. The state descriptions, holding time distributions, and degradation growth rates are summarized in Table 2.

Table 2: Summary input data for the turbine blade experiment.

State	Rotation Speed (rpm)	Holding Time Distribution	Degradation Growth Rate
1	$1000 - 4999$	Beta (a, b)	$r(1) \sim U(0, 5)$
2	$5000 - 8999$	Beta (a, b)	$r(2) \sim U(0, 5)$
3	$9000 - 12999$	Weibull (c, d)	$r(3) \sim U(0, 5)$
4	$13000 - 15999$	Weibull (c, d)	$r(4) \sim U(0, 5)$

In Table 2, Beta (a, b) denotes a beta probability distribution with shape parameters a and b. (Note that the beta distribution does not have a closed-form c.d.f.). The parameter values a and b were randomly generated for each of the 1000 scenarios so that

$$a \sim U(1, 5) \quad \text{and} \quad b \sim U(1, 5).$$

This range of values allows for a variety of distribution shapes including right-skewed, left-skewed, and symmetric forms. Weibull (c, d) denotes a Weibull distribution with shape parameter c and

scale parameter d. These parameter values were also randomly generated for each of the 1000 scenarios as follows:

$$c \sim U(0.5, 6) \quad \text{and} \quad d \sim U(0.5, 6).$$

The transition probability matrix of the SMP environment was generated by first creating a 4×4 matrix $\mathbf{B} = [b_{i,j}]$ such that $b_{i,j} \sim U(0,1)$, $j \neq i$, and $b_{i,i} = 0$, which was subsequently re-scaled to ensure a stochastic matrix. Specifically, the (i,j)th element of the randomly-generated transition matrix $\boldsymbol{P}$ is obtained by

$$p_{i,j} = \frac{b_{i,j}}{\sum_{j=1}^{4} b_{i,j}}, \quad i, j \in S$$

for each of the 1000 scenarios. From Table 2, it is seen that the matrix of degradation rates is also randomly generated in each case.

We assume Z begins in the first state with probability 1; therefore, the initial probability vector is $\boldsymbol{e}_1' = (1, 0, 0, 0)$. We simulated Z and observed its evolution for $\tau = 10,000$ time units. From the simulated environment, we estimated the parameters $\widehat{\boldsymbol{P}}$, $m_k(i)$, s_i^2, $\hat{c}_i^2$, $i = 1, 2, 3, 4$, $k = 1, 2, 3$, using equations (19)–(24). Subsequently, we applied the criteria of Table 1 to select PH distributions for each state i and constructed $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{R}}_d$ using the procedure described in section 4. For the sake of brevity, we do not include here the PH representations $(\boldsymbol{\beta}, \boldsymbol{T}_i)$, $i = 1, \ldots, 4$.

In 988 of 1000 randomized scenarios, we fail to reject the null hypothesis that $\widehat{F}_1(x,t)$ is equivalent to the benchmark c.d.f. $F_1(x,t)$ at the $\alpha = 0.05$ level. Of the 12 cases in which we reject H_0, six were the result of instability in the numerical Laplace transform inversion algorithm. Specifically, in the rare cases when the c.d.f. is not sufficiently smooth, the inversion algorithm of [1] may not perform well without modifications. Of the remaining six cases for which H_0 was rejected, the average maximum absolute deviation in probability was approximately 0.1478. In summary, we fail to reject nearly 99% of the 1000 randomized scenarios.

As a sample illustration of the quality of the approximations, Figure 1 depicts the benchmark (simulated) lifetime c.d.f., $F_i(x,t)$, and the approximated lifetime c.d.f. $\widehat{F}_i(x,t)$ obtained using the surrogate Markov environment $\widehat{\mathbf{Q}}$ and degradation rate matrix $\widehat{\mathbf{R}}_d$. In this particular scenario, our procedure yielded a 41×41 generator matrix ($\widehat{\mathbf{Q}}$). The maximum absolute deviation in probability of this scenario is $\varepsilon \approx 0.02562$. These results indicate that the surrogate environment process with PH-approximated holding times very closely tracks the simulated lifetime c.d.f., even when holding times are clearly non-memoryless and their parameters are randomized.

We also compared the first and second moments of the full lifetime distribution, given that the environment starts in state 1. These values are summarized in Table 3.

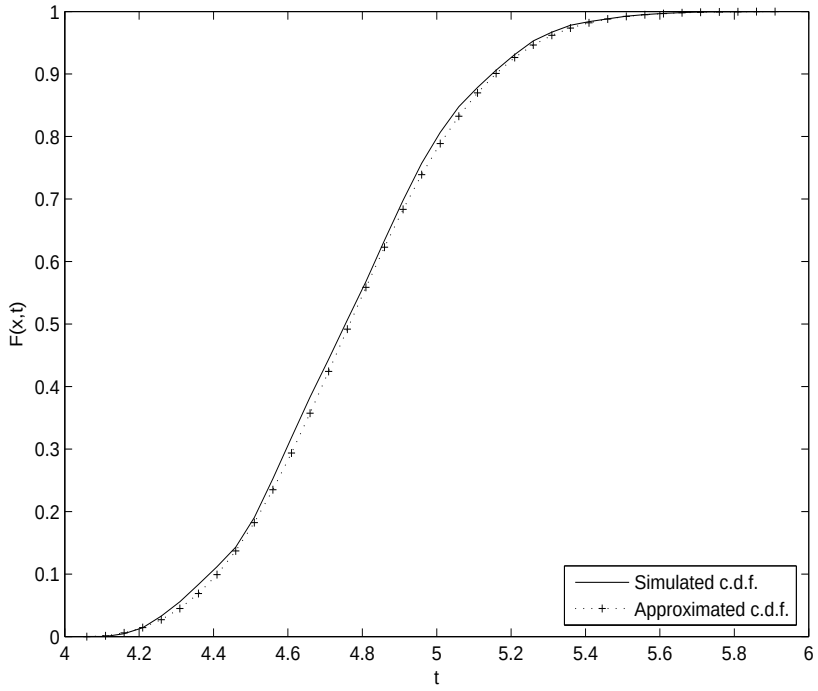

Figure 1: Simulated versus approximated lifetime c.d.f.: Turbine blade experiment.

Table 3: Lifetime moment comparisons for Experiment 1.

Parameter	Simulated	PH Approximation	% Error
$\mathbb{E}[T_x \vert Z(0) = 1]$	4.763766782	4.777832971	0.29527453
$\mathbb{E}[T_x^2 \vert Z(0) = 1]$	22.77583335	22.91001248	0.589129367

The results of Table 3 illustrate the quality of the lower moment approximations obtained using our procedure. Both values differ from their simulated counterparts by less than 1%.

5.2 Experiment 2: Chemical Coating Decomposition

Now consider a protective, automotive coating that is subject to weathering by the outdoor environment. Environmental effects, such as temperature and solar radiation, cause chemical decomposition (degradation) of the coating that can be measured in terms of gloss loss and/or color change. The failure time can be defined as the first time that the cumulative degradation of the coating reaches a critical threshold. Here, we assume the rate of degradation of the coating depends

on the ambient environment to which it is exposed, and the critical threshold is $x = 5.0$ units. The environment is characterized by five-state semi-Markov processes that are defined in Table 4 along with the degradation rates and state holding time distributions.

Table 4: Summary input data for the chemical coating experiment.

State	Sky condition	Temperature	Holding time distribution	Decomposition rate
1	Cloudy	≤ 32 deg F	Weibull (c, d)	$r(1) \sim U(0, 2)$
2	Cloudy	> 32 deg F	Beta (a, b)	$r(2) \sim U(0, 2)$
3	Sunny	≤ 32 deg F	Weibull (c, d)	$r(3) \sim U(0, 2)$
4	Sunny	> 32 deg F	Beta (a, b)	$r(4) \sim U(0, 2)$
5	Rain	> 32 deg F	Gamma (g, h)	$r(5) \sim U(0, 2)$

In Table 4, Beta (a, b) denotes a beta probability distribution with shape parameters a and b, and Weibull (c, d) denotes a Weibull distribution with shape parameter c and scale parameter d. Gamma (e, f) denotes a gamma distribution with parameters g and h. For each of the 1000 scenarios the distribution parameter values were randomly generated as follows:

$$\begin{aligned} a, b &\sim U(0.1, 3.1), \\ c, d &\sim U(1, 3), \\ g, h &\sim U(0.1, 2.1), \end{aligned}$$

and the 5×5 transition probability matrices of the SMP environments were generated randomly as in the turbine blade experiment. Here, we again assume that Z begins in state 1 (w.p. 1) so that the initial probability vector is $\boldsymbol{e}_1' = (1, 0, 0, 0)$. We simulated Z and observed its evolution for $\tau = 10,000$ time units. From the simulated environment, we estimated the parameters $\widehat{\boldsymbol{P}}$, $m_k(i)$, s_i^2, $\hat{c}_i^2$, $i = 1, 2, 3, 4, 5$, $k = 1, 2, 3$, using equations (19)–(24). We used $\hat{c}_i^2$ to determine the appropriate PH distribution approximation for each $i \in S$ using Table 1 and constructed $\widehat{\mathbf{Q}}$ and $\widehat{\mathbf{R}}_d$ using the algorithm described in section 4.

Of the 1000 randomized scenarios, we fail to reject the null hypothesis that $\widehat{F}_1(x, t)$ is equivalent to $F_1(x, t)$ in 980 cases at the $\alpha = 0.05$ level. Of the 20 cases for which we reject H_0, 10 were caused by instability in the numerical Laplace transform inversion algorithm as in the first experiment. For the remaining 10 cases, the average maximum absolute deviation in probability was approximately 0.2230. To summarize the results of this experiment, we fail to reject the null hypothesis that our

approximated c.d.f. is equivalent to the benchmark c.d.f. in 98% of the randomized scenarios.

To further illustrate the quality of the approximations, Figure 2 depicts the simulated lifetime c.d.f., $F_i(x,t)$, and the approximated lifetime c.d.f. $\widehat{F}_i(x,t)$ obtained using the surrogate Markov environment $\widehat{\mathbf{Q}}$ and degradation rate matrix $\widehat{\mathbf{R}}_d$. For this particular experiment, our procedure yielded a generator matrix $\widehat{\mathbf{Q}}$ which is 82×82, or twice as large as the illustrative case in experiment 1. Nevertheless, the resulting maximum absolute deviation in probability of this example is $\varepsilon \approx 0.02648$.

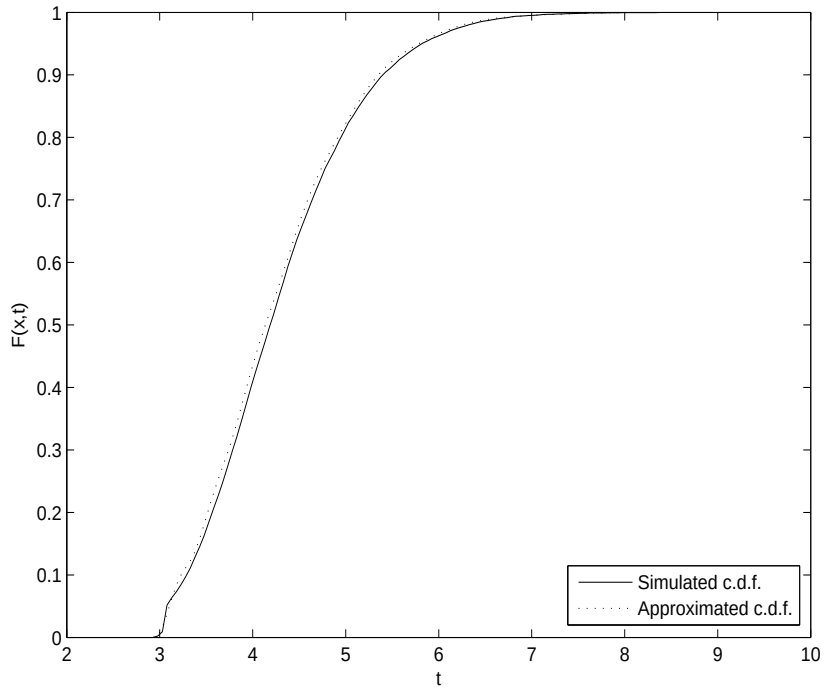

Figure 2: Simulated versus approximated lifetime c.d.f.: Chemical coating experiment.

The lower two moments of the lifetime distribution were also compared, and the results are summarized in Table 5.

Table 5: Lifetime moment comparisons for Experiment 2.

Parameter	Simulated	PH approximation	% Error
$\mathbb{E}[T_x \vert Z(0)=1]$	4.290965585	4.249862484	0.957898641
$\mathbb{E}[T_x^2 \vert Z(0)=1]$	19.0983667	18.74008389	1.875986669

As for the first example, the approximation of the lower moments is also of very high quality. The maximum percent difference is under 2%.

While the results of these two randomized experiments are promising, it will be instructive to assess the quality of our approximations using real data. Unfortunately, such data was unavailable for use in this study. We further elaborate upon the advantages and limitations of our approach in section 6.

6 Conclusions

In this paper, we have presented a novel technique for incorporating environmental effects on component degradation for the purpose of evaluating reliability indices. The modeling framework and associated numerical techniques provide improved flexibility by allowing for semi-Markovian environments and general degradation patterns. These characteristics make the approach appealing whenever it is possible to assess physical degradation as a function of the temporal evolution of the environment. Within our framework, the environment can assume a variety of roles (e.g., time-varying operating conditions, the ambient environment, etc.).

While the procedures described in sections 3 and 4 are mathematically sound and relatively easy to implement, they do impose moderately restrictive assumptions, namely that (i) the environment evolves in a temporally homogeneous manner; (ii) the number of distinct states (K) is known; (iii) the environment transitions between states according to a Markov chain; and (iv) the future degradation of the unit is independent of the history of the degradation. Despite these limitations, the approximations provide an important extension to prior methods that either ignore the effects of the environment entirely, or assume that the environment is completely memoryless.

In the future, it will be instructive to consider models for systems whose rates of degradation depend, not only on the state of the environment, but also on the current level of degradation. It may also be useful to consider Bayesian techniques for updating the parameters of the environment as additional sensor data becomes available. If the proposed models are to be of any practical value to engineers, it is vital to provide an accurate estimate of the mapping r that describes the evolution of degradation as a function of the environment. To this end, real degradation data is required, as is the guidance and experience of subject matter experts, to ensure the degradation rates are selected in accordance with known physical principles. Finally, as noted in section 5, it will be important to evaluate the procedure using real environmental and degradation data.

Acknowledgements

The authors are grateful for the comments and suggestions of two anonymous referees, the Department Editor, and Dr. Steven M. Cox. This research was sponsored, in part, by the Air Force Office of Scientific Research (FQ8671-04-0-0359).

References

[1] Abate, J. and Whitt, W. (1995) Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, **7**(1), 36-43.

[2] Aldous, D. and Shepp, L. (1987) The least variable phase type distribution is Erlang. *Communications in Statistics-Stochastic Models*, **3**(3), 467-473.

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

[4] Basawa, I.V. and Rao, P. (1980) *Statistical Inference for Stochastic Processes.* John Wiley and Sons, New York.

[5] Çinlar, E. (1977) Shock and wear models and Markov additive processes. In *The Theory and Applications of Reliability* (I.N. Shimi and C.P. Tsokos, eds.), Academic, 193-214.

[6] Çinlar, E. (1984) Markov and semi-Markov models of deterioration. In *Reliability Theory and Models*, M. Abdel-Hameed, E. Çinlar, and J. Quinn, Eds: Academic Press, 341.

[7] Conover, W.J. (1999) *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York.

[8] Cox, D.R. (1955) A use of complex probabilities in the theory of stochastic processes. In *Proceedings of the Cambridge Philosophical Society*, **55**, 313-319.

[9] Cox, D.R. (1972) Regression models and life tables. *Journal of the Royal Statistical Society, B*, **34**, 187-220.

[10] Cox, D.R. and Oakes, S. (1984) *Analysis of Survival Data*, Chapman and Hall, London.

[11] Ebeling, C.E. (1997) *An Introduction to Reliability and Maintainability Engineering*, McGraw Hill, New York.

[12] Ebrahimi, N. (2006) System reliability based on system wear. *Stochastic Models*, **22**(1), 21-36.

[13] Ejaz, N. and Tauqir, A. (2006) Failure due to structural degradation in turbine blades *Engineering Failure Analysis*, 13, 452-463.

[14] Elsayed, E.A. (2000) Perspectives and challenges for research in quality and reliability engineering. *International Journal of Production Research*, 38, 1953-1976.

[15] Elsayed, E.A., Liao, H., and Wang, X. (2006) An extended linear hazard regression model with application to time-dependent dielectric breakdown of thermal oxides. *IIE Transactions*, 38, 329-340.

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

[17] Gebraeel, N., Lawley, M., Li, R. and Ryan, J. (2005) Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, 37, 543-557.

[18] Gebraeel, N. and Pan, J. (2008) Prognostic degradation models for computing and updating residual life distributions in a time-varying environment. *IEEE Transactions on Reliability*, 57, 539-550.

[19] Jardine, A.K.S. and Anderson, M. (1985) Use of concomitant variables for reliability estimation. *Maintenance Management International*, 5, 135-140.

[20] Jardine, A.K.S., Anderson, P.M. and Mann, D.S. (1987) Application of the Weibull proportional hazards model to aircraft and marine engine failure data. *Quality and Reliability Engineering International*, 3, 77-82.

[21] Jardine, A.K.S. Ralston, P., Reid, N. and Stafford, J. (1989) Proportional hazards analysis of diesel engine failure data. *Quality & Reliability Engineering International*, 5, 207-216.

[22] Johnson, M.A. (1993) Selecting parameters of phase distributions: Combining nonlinear programming, heuristics, and Erlang distributions. *ORSA Journal on Computing*, **5**(1), 69-83.

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

[24] Kharoufeh, J.P. and Cox, S.M. (2005) Stochastic models for degradation-based reliability. *IIE Transactions*, **37**, 533-542.

[25] Kharoufeh, J.P., Finkelstein, D., and Mixon, D. (2006) Availability of periodically inspected systems with Markovian wear and shocks. *Journal of Applied Probability*, **43**, 303-317.

[26] Kiessler, P., Klutke, G., and Yang, Y. (2002) Availability of periodically inspected systems subject to Markovian degradation. *Journal of Applied Probability*, **39**, 700-711.

[27] Klutke, G., Wortman, M., and Ayhan, H. (1996) The availability of inspected systems subject to random deterioration. *Probability in the Engineering and Informational Sciences*, **10**, 109-118.

[28] Kobbacy, K.A.H., Fawzi, B.B., Percy, D.F. and Ascher, H.E. (1997) A full history proportional hazards model for preventive maintenance scheduling. *Quality and Reliability Engineering International*, **13**, 187-198.

[29] Kulkarni, V.G. (1995), *Modeling and Analysis of Stochastic Systems*, Chapman and Hall.

[30] Lang, A. and Arthur, J.L. Arthur (1997) Parameter approximation for phase-type distributions. In *Matrix-Analytic Methods in Stochastic Models*, S.R. Chakravarthy and A.S. Alfa (Eds.), Marcel Dekker, Inc., New York, 151-206.

[31] Lawless, J.F. (1982) *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.

[32] Li, G. and Luo, J. (2005) Shock model in Markovian environment. *Naval Research Logistics*, **52**, 253-260.

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

[34] Mallor, F. and Omey, E. (2001) Shocks, runs, and random sums. *Journal of Applied Probability*, **38**, 438-448.

[35] Marie, R. (1980) Calculating equilibrium probabilities for $\lambda(n)/C_k/1/N$ queues. *Proceedings of Performance 1980*, **9**(2), 117-125.

[36] Mishra, S. and Pecht, M. (2002) In-situ sensors for product reliability monitoring. In *Proceedings of SPIE* 4755, 10-19.

[37] Neuts, M. F. (1981) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Dover Publications, Inc., New York.

[38] Osogami, T.and Harchol-Balter, M. (2006) Closed-form solutions for mapping general distributions to quasi-minimal PH distributions. *Performance Evaluation*, **63**(6), 524-552.

[39] Özekici, S. and Soyer, R. (2003) Bayesian analysis of Markov modulated Bernoulli processes. *Mathemathical Methods of Operations Research*, **57**, 125-140.

[40] Özekici, S. and Soyer, R.(2004) Reliability modeling and analysis in random environments. In *Mathematical Reliability: An Expository Perspective*, (R. Soyer, T.A. Mazzuchi, and N. D. Singpurwalla, Eds.), 249-273.

[41] Özekici, S. and Soyer, R. (2006) Semi-Markov modulated Poisson process: Probabilistic and Statistical Analysis. *Mathemathical Methods of Operations Research*, **64**, 125-144.

[42] Perros, H. (1994), *Queueing Networks with Blocking.* Oxford University Press, New York.

[43] Persson, C. and Persson, P.O. (1993) Evaluation of service-induced damage and restoration of cast turbine blades. *Journal of Materials Engineering and Performance*, **2**, 565-570.

[44] Singpurwalla, N.D. (1995) Survival in dynamic environments. *Statistical Science*, **10**, 86-103.

[45] Tamarin, Y. (2002) *Protective Coatings for Turbine Blades.* ASM International. Materials Park, OH, USA.

[46] Telek, M. and Heindl, A. (2003) Matching moments for acyclic discrete and continuous phase-type distributions of second order. *International Journal of Simulation*, **3**, 47-57.

[47] 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.

[48] Whitt, W. (1984) On approximations for queues III: Mixtures of exponential disributions. *Bell Labs Technical Journal*, **63**(1), 163-175.

[49] Zhan, Y, Makis, V. and Jardine, A.K.S. (2003) Adaptive model for vibration monitoring of rotating machinery subject to random deterioration. *Journal of Quality in Maintenance Engineering*, **9**, 351-375.