IIE Transactions

ISSN: 0740-817X (Print) 1545-8830 (Online) Journal homepage: https://www.tandfonline.com/loi/uiie20

Multistate degradation and supervised estimation methods for a condition-monitored device

Ramin Moghaddass & Ming J. Zuo

To cite this article: Ramin Moghaddass & Ming J. Zuo (2014) Multistate degradation and supervised estimation methods for a condition-monitored device, IIE Transactions, 46:2, 131-148, DOI: 10.1080/0740817X.2013.770188

To link to this article: https://doi.org/10.1080/0740817X.2013.770188

Published online: 06 Nov 2013.

IIE Transactions (2014) **46**, 131–148

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2013.770188

Multistate degradation and supervised estimation methods for a condition-monitored device

RAMIN MOGHADDASS and MING J. ZUO*

Reliability Research Laboratory, Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, T6 2G8, Canada
E-mail: ming.zuo@ualberta.ca

Received January 2012 and accepted January 2013

Multistate reliability has received significant attention over the past decades, particularly its application to mechanical devices that degrade over time. This degradation can be represented by a multistate continuous-time stochastic process. This article considers a device with discrete multistate degradation, which is monitored by a condition monitoring indicator through an observation process. A general stochastic process called the nonhomogeneous continuous-time hidden semi-Markov process is employed to model the degradation and observation processes associated with this type of device. Then, supervised parametric and nonparametric estimation methods are developed to estimate the maximum likelihood estimators of the main characteristics of the model. Finally, the correctness and empirical consistency of the estimators are evaluated using a simulation-based numerical experiment.

Keywords: Multistate reliability, degradation process, condition monitoring, supervised estimation

1. Introduction

Multistate reliability analysis has received considerable attention recently in the domain of reliability and maintenance, particularly for a mechanical device operating under stress, load, or fatigue conditions (Moghaddass *et al*., 2013). As the overall performance of this type of device deteriorates over time, its health condition may experience certain levels of health states before failure and, thus, discrete multistate deterioration. This multistate deterioration can be represented by a stochastic process, referred to as the degradation process, which corresponds to the stochastic relationship among multiple discrete health states, ranging from perfect functioning to complete failure. By tracking a single indicator or multiple indicators corresponding to certain conditions of a device, condition monitoring techniques have been used for health state estimation of such devices. A condition-monitored device with discrete multistate deterioration deals with two types of stochastic processes: (i) the degradation process and (ii) the observation process. Many applications of devices with the degradation and observation processes are reported in the literature, such as in Peng and Dong (2011) for a hydraulic pump with four levels of health states and in Hongzhi *et al*. (2011) for a gearbox with five degradation states, both under vibration monitoring.

To utilize condition monitoring information for health state diagnosis and prognosis of a device with multistate degradation, the characteristics of the stochastic processes associated with the degradation and observation processes of the device need to be determined. In other words, the parameters that characterize these stochastic processes need to be estimated from real-time data. Depending on the availability of these data, a parameter estimation method can be either a supervised or an unsupervised estimation method. The data required for estimation in a supervised estimation method include trajectories of both the degradation and observation processes, whereas in an unsupervised estimation method the data required for estimation include only the observation process. To reveal the actual health states of a device over time (degradation process), an inspection process can be implemented. During the inspection time, the system is shut down or suspended, and depending on the type of the device being inspected, methods such as visual inspection and measurement can be used for health state identification. Examples of such inspection policies are reported in Chiang and Yuan (2001) for periodic inspection and in Liao *et al*. (2006) for continuous inspection. Directly observing the health states may be too costly and technically complicated, and because of that, unsupervised estimation methods have been developed for devices with unobservable states. Such scenarios have been addressed in Moghaddass and Zuo (2012) and Moghaddass *et al*. (2013).

The problem of supervised estimation for the main characteristics of the degradation and observation processes

*Corresponding author

of a condition-monitored device is considered in this article. This article focuses only on those devices under condition monitoring where it is possible to somehow monitor the actual degradation levels continuously or at some discrete points (such as inspection points, major shutdowns, or overhaul times) in order to obtain required supervised training data for parameter estimation. It should be mentioned here that even when the states are not directly observable, certain approaches might be employed to obtain the training data required for a supervised estimation. Laboratory experiments (Peng and Dong, 2011), simulation (Saxena *et al*., 2008), and accelerated run-to-failure experiments (Huang *et al*., 2007) are helpful approaches that can provide run-to-failure data required in a supervised estimation method for a device with unobservable states. We point out here that the results of this article can also be used for devices that are not condition-monitored at all but are only under inspection (either continuous or periodical). In such cases, our method can be used to estimate the characteristic parameters of the degradation process of such devices (the observation process does not exist under such a scenario). A schematic view of the data acquisition procedure for such a type of device under inspection and condition monitoring is given in Fig. 1.

For the problem described in the previous paragraph, we propose two supervised estimation methods that employ real-time trajectories in either the form of run-to-failure sequences or suspended sequences (right-censored data). The first method is a parametric estimation method, which is based on the assumption that the statistical forms of the transition rate functions are known; that is, the characteristic parameters of transition rate functions are the only unknown parameters of the degradation process. The second method is a nonparametric estimation method, which is based on the assumption that the statistical forms of transition rate functions are not known or transition rate functions are distribution free. Therefore, instead of characteristic parameters, important quantities of interest, such as the kernel or transition rate functions, should be estimated.

In both parametric and nonparametric estimation methods, the estimation procedure depends on the type of the stochastic processes associated with the degradation and observation processes. In a multistate degradation process,

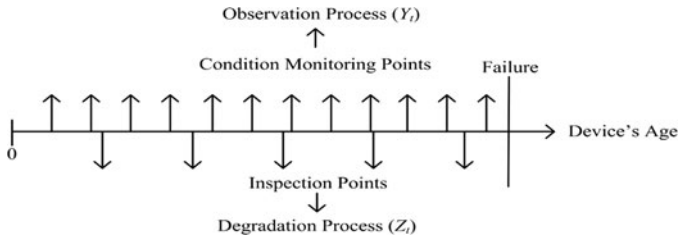

Fig. 1. A schematic view of the data collection process (a full history of a failure).

the degradation transition may depend on certain factors, such as the two states involved in the transition, the time that the device reached the current state, the time already spent at the current state, the total age of the device, or any combination of these factors. Concerning the dependency of a transition on the mentioned factors, a degradation transition between two states can follow either a Markovian or a semi-Markovian structure (Moghaddass *et al*., 2013). In a Markovian degradation structure, the transition between two states depends only on the two states involved and is independent of the history of the process. Examples of the application of Markovian degradation can be found in Lin and Makis (2004) and Kim *et al.* (2010) for a Continuous-Time Markov Process (CTMP) and in Kiessler *et al*. (2002) and Morcous *et al*. (2003) for a Discrete-Time Markov Process (DTMP). The constant transition rates and exponential sojourn time distribution at each state in a CTMP and the constant transition probabilities and geometric sojourn time distribution at each state in a DTMP limit their applications with regard to practical cases.

To overcome these limitations, researchers have considered semi-Markov structures, which take into account the history of the process and consider arbitrary sojourn time distributions at each state. With respect to the dependency of transitions on the states involved, the time that the device reached the current state, the time spent at the current state, and the total age of the device, the semi-Markov process can be in one of four main forms, namely, Discrete-Time and Continuous-Time Aging Markovian Process (DTAMP and CTAMP), Homogeneous Discrete-Time and Continuous-Time Semi-Markov Process (HDTSMP and HCTSMP), Explicit-Duration Discrete-Time and Continuous-Time Semi-Markov Process (EDDTSMP and EDCTSMP), and Nonhomogeneous Discrete-Time and Continuous-Time Semi-Markov Process (NHDTSMP and NHCTSMP). In the following, the main characteristics of each of the above structures are briefly reviewed and their applications in multistate degradation modeling are described.

In multistate degradation with an aging Markovian structure as reported in Chen and Wu (2007) and Dong and Peng (2011) for DTAMP, the device degrades from any current state *i* to a degraded state *j* according to a transition probability, which changes with the total age of the device considering an aging factor. To our knowledge, no work has been done for multistate degradation modeling using CTAMP. Finding an appropriate aging factor and the assumption of independence of transition distributions to the time that the device reached the last state and the age of the device at each state are the main limitations of using the aging Markovian process in multistate degradation modeling.

In multistate degradation with homogeneous semi-Markov structures as reported in Barbu and Limnios (2008) and Chryssaphinou *et al.* (2011) for HDTSMP and in Lisnianski and Levitin (2003), Kharoufeh *et al*. (2010), and

Moghaddass and Zou (2011) for HCTSMP, the device degrades according to a nonhomogeneous transition rate (or transition probability), which changes with the time spent at the current state. Based on this structure, the sojourn time distribution at each state follows an arbitrary distribution. Although compared to Markovian degradation, HDTSMP and HCTSMP comply with more real cases by taking the age of the device at each state into consideration, they are not applicable when the degradation transition between states is affected by the time that the device reached the current state and the total age of the device. In multistate degradation models with explicit-duration semi-Markov structures as reported in Dong *et al*. (2006) and Dong and He (2007) for EDDTSMP and in Lam and Yeh (1994) and Hongzhi *et al*. (2011) for EDCTSMP, the device stays at its current state i following an arbitrary (but explicitly defined) distribution and then transits to a degraded state j with a constant probability of $p_{i,j}$. The main shortcoming of this structure is the constant probability of a transition between two states. Additionally, finding an explicit sojourn time distribution is not a straightforward course of action (Moghaddass *et al*., 2013).

The use of any of the above-reviewed stochastic models for multistate degradation modeling is subject to two common limitations. First, all degradation transitions should follow an identical structure. Second, each transition may depend only on a maximum of two out of the following factors, namely, the two states involved in the transition, the time point that the device reached the current state, the time spent at the current state, and the total age of the device. To overcome these limitations and include more practical cases, these limitations need to be relaxed. It is interesting to note that the only strong stochastic process that can directly address the above-mentioned requirements is the nonhomogeneous semi-Markov structure. Surprisingly, a very limited amount of research has been devoted to multistate degradation modeling using either NHDTSMP or NHCTSMP. Due to the flexibility, generality, and more practical applications of this structure, we consider it as the structure of the multistate degradation model. In addition, as degradation transitions naturally evolve in the continuous-time domain and as a discrete time process can be considered as a special case of a continuous-time process (Limnios, 2011), in this article, we assume that the multistate degradation process follows a NHCTSMP.

With regard to the observation process (hidden process), based on the similar approach taken in other studies, such as Moghaddass and Zuo (2011), we have assumed that the Condition Monitoring (CM) indicator can take discrete values and therefore the stochastic relationship between each unobservable state and the CM indicator can be represented by an observation probability distribution. It should be noted here that, as the structure of the observation probability distribution can be represented by a matrix, the parameters of the observation process are the elements of this matrix. That is why nonparametric estimation and parametric estimation are equivalent for such a type of observation process.

We note here that recent research studies have focused more on nonparametric estimation methods, particularly for semi-Markov structures. The main reason for this is that nonparametric estimation methods can reduce modeling bias by considering no specific model structure. In addition, they are more useful when limited information is available or flexibility about the underlying model is required (Zhao, 2008). The nonparametric maximum-likelihood and empirical estimators of an HDTSMP are investigated in Barbu and Limnios (2006a, 2006b). The nonparametric estimation for reliability, availability, and failure rate of a CTMP was discussed in Sadek and Limnios (2005). Nonparametric estimation for transition rate functions of an HCTSMP was provided in Moore and Pyke (1968) and Ouhbi and Limnios (1999). Nonparametric estimators of the mean time to failure, the mean up-time, and the mean downtime of an HCTSMP were presented in Limnios and Ouhbi (2006). The focus of the present article is different from the above-reviewed papers because it focuses on devices with multistate degradation evolving according to an NHCTHSMP, which includes two types of processes: (i) a nonhomogeneous continuous-time semi-Markov process, which is a more general stochastic process to represent a degradation process than those used in Moore and Pyke (1968), Ouhbi and Limnios (1999), Sadek and Limnios (2005), and Limnios and Ouhbi (2006); and (ii) an observation process reflecting indirectly the actual health states of the device. In addition, unlike Moore and Pyke (1968), Ouhbi and Limnios (1999), Sadek and Limnios (2005), and Limnios and Ouhbi (2006), both parametric and nonparametric estimation methods, which use inspection and condition monitoring data to characterize the degradation and observation processes, are considered in this article. However, as CTMP and HCTSMP are special cases of NHCTSMP, some of the results in Moore and Pyke (1968), Ouhbi and Limnios (1999), Sadek and Limnios (2005), and Limnios and Ouhbi (2006) can be used to verify the results that we develop in this article.

This article is the continuation of our previous work in Moghaddass and Zuo (2012) and Moghaddass *et al*. (2013). In Moghaddass and Zuo (2012), an unsupervised estimation method for a multistate device using NHCTHSMP was presented. In Moghaddass *et al*. (2013), which is based on Moghaddass and Zuo (2012), several stochastic models for multistate degradation modeling were reviewed and important characteristics of the NHCTHSMP and its application in multistate degradation were discussed. The current article is an extension of these two works and proposes a supervised parametric estimation method, as well as a supervised nonparametric estimation method for a device under NHCTHSMP. As for the differences in the estimation methods comparing this article to the unsupervised estimation method given in Moghaddass and Zuo (2012) and Moghaddass *et al*. (2013), it should be pointed out that

the present article considers supervised estimation, which requires both trajectories of the degradation and observation processes. In addition, supervised nonparametric estimation is also considered, wherein the statistical form of the degradation process is assumed to be unknown. This is fundamentally different from the unsupervised parametric estimation method given in Moghaddass and Zuo (2012) and Moghaddass *et al*. (2013), where the statistical distribution of the degradation process was assumed to be known. Although the general structure of the NHCTHSMP for modeling the multistate degradation process plays a significant role in the contribution made in this article, developing supervised estimation methods (parametric and nonparametric) to characterize the degradation and observation processes of a condition-monitored device with multiple health states evolving according to the NHCTHSMP is the main contribution of the article.

The remainder of this article is organized as follows: In Section 2, we review the basics of the NHCTHSMP applicable for multistate degradation modeling. Section 3 presents two parameter estimation methods (parametric and nonparametric). Simulation-based numerical experiments that were performed to evaluate the correctness and empirical consistency of the estimation results are discussed in Section 4. Finally, we conclude in Section 5 and then explore the direction of our future research.

2. Preliminaries

This section briefly reviews the fundamentals of the NHCTHSMP structure in multistate deterioration modeling for a condition-monitored device. The main notation used in this article is as follows:

N	the number of health states of the device
E	the finite state space representing the health state of the device
Z_t	the state of the device at time t (degradation process)
Y_t	the output of the observation process at time t
T_n	the time at the nth transition
X_n	the state at the nth transition
U_n	the output of the observation process at the nth transition
θ	the characteristic parameters of the NHCTHSMP model
$\mathbf{B}$	observation probability matrix (emission matrix)
M	the number of possible condition monitoring output
$V = \{v_1, \ldots, v_M\}$	the condition monitoring indicator space
$b_i(j)$	the probability of obtaining CM indicator v_j while the device is in state i
K	the number of historical sequences of observations (states)
l_k	the number of observation points for the kth sequence of observations
$t_1^{(k)}, \ldots, t_{l_k}^{(k)}$	the observation time points for the kth sequence of observations
$O^{(k)}(Q^{(k)})$	kth sequence of observations (states)
δ	inspection interval

2.1. *Transition modeling*

The use of transition rate functions as the main tool to describe a semi-Markov process was discussed in Moura and Droguett (2009) and Moghaddass *et al*. (2013). Moghaddass *et al*. (2013) presented a definition for the transition rate in a multistate deterioration model based on which a transition can depend on the two states involved in the transition, the time that the device took to reach the current state, the sojourn time at the current state, and/or the total age of the device. Based on this definition, given that the device is in state i at time u, the probability that it transitions to state j in an infinitesimal time interval $(u, u + \mathrm{d}u)$ is the transition rate between states i and j at time u given as

$$\lambda_{i,j}(s,u) = \lim_{\mathrm{d}u \to 0} \frac{\Pr\{u < T_{n+1} - T_n \le u + \mathrm{d}u,\, X_{n+1} = j | u < T_{n+1} - T_n,\, T_n = s,\, X_n = i\}}{\mathrm{d}u}, \quad (i,j) \in E, (s,u) \in [0,\infty]. \tag{1}$$

2.2. *Important measures for a multistate device under the NHCTHSMP structure*

The degradation process associated with a multistate device can be represented by a couple (X_n, T_n), where X_n is the state of the device at the nth transition and T_n is the time at the nth transition. The process (X, T) is a nonhomogeneous Markov renewal process if its kernel has the following property (Blasi *et al*., 2004):

$$Q_{i,j}(s,t) = \Pr(X_{n+1} = j,\, T_{n+1} \le t | X_n = i,\, T_n = s,\, (X_c, T_c), 0 \le c < n) = \Pr(X_{n+1} = j,\, T_{n+1} \le t | X_n = i,\, T_n = s), \quad (i,j) \in E, (s,t) \in [0,\infty]. \tag{2}$$

Now, it is possible to define the degradation process $Z_t = (Z_t, t \in \Re_0^+)$, where $Z_t = X_{N_t}$ and $N_t = \sup\{n : T_n \le t\}$, which follows an NHCTSMP. Now let $\mathbf{V} = \{v_1, \ldots, v_M\}$ be the observation space and U_n be the output of the observation process at the nth transition. Then, the following relationship holds true:

$$\Pr(U_n = v_j | (U_0, \ldots, U_{n-1}, X_0, \ldots, X_{n-1}, X_n = i)) = \Pr(U_n = v_j | X_n = i) = b_i(j), \quad i \in E, 1 \le j \le M. \tag{3}$$

Let us define Y_t ($Y_t \in V$) as the output of the observation process at time t. If Z_t is an NHCTSMP then the couple (Z, Y) is also an NHCTHSMP. Now, similar to

Moghaddass *et al*. (2013), based on the kernel function and the transition rate function as the two main describers of the degradation process, three important measures are introduced, which include important information on the stochastic characteristics of the degradation process. We express the relationship between the kernel function and the transition rate function (Moghaddass *et al*., 2013) as

$$Q_{i,j}(s,t) = \Pr(X_{n+1}=j,\, T_{n+1}\le t \mid X_n=i,\, T_n=s) = \int_0^{t-s} \lambda_{i,j}(s,u)\exp\left(-\int_0^u \sum_z \lambda_{i,z}(s,x)\mathrm{d}x\right)\mathrm{d}u, \quad (i,j)\in E,(s,t)\in[0,\infty]. \tag{4}$$

The first important measure is the embedded transition probability matrix ($\mathbf{P}=[p_{i,j}(s)]$):

$$p_{i,j}(s) = \Pr(X_{n+1}=j \mid X_n=i,\, T_n=s) = \lim_{t\to\infty} Q_{i,j}(s,t) = \int_0^{\infty} \lambda_{i,j}(s,u)\exp\left(-\int_0^u \sum_z \lambda_{i,z}(s,x)\mathrm{d}x\right)\mathrm{d}u, \quad (i,j)\in E, s\in[0,\infty]. \tag{5}$$

The second measure is the sojourn time at state i, given that state i is reached at time s:

$$H_i(s,t) = P(T_{n+1}-T_n\le t \mid X_n=i,\, T_n=s) = \sum_{j\in E} Q_{i,j}(s,t+s) = 1-\exp\left(-\int_0^t \sum_z \lambda_{i,z}(s,x)\mathrm{d}x\right), \quad i\in E,(s,t)\in[0,\infty]. \tag{6}$$

The next measure is the cumulative distribution of the conditional sojourn time, given that the state subsequently occupied is state j and state i is reached at time s:

$$G_{i,j}(s,t) = \Pr(T_{n+1}-T_n\le t \mid X_n=i,\, X_{n+1}=j,\, T_n=s) = Q_{i,j}(s,s+t)\times p_{i,j}(s)^{-1},\ p_{i,j}(s)>0, \quad (i,j)\in E,(s,t)\in[0,\infty]. \tag{7}$$

Now the expected sojourn time at state i, given that state i is reached at time s ($\bar{h}_i(s)$), and the expected conditional sojourn time at state i, given that state i is reached at time s and the next state is state j ($\bar{g}_{i,j}(s)$), can be calculated according to Equation (6) and Equation (7). The following shows how the transition rate functions can be calculated from the above-described characteristic measures:

$$\lambda_{i,j}(s,t) = \frac{Q'_{i,j}(s,t+s)}{1-H_i(s,t)} = \frac{G'_{i,j}(s,t)\times p_{i,j}(s)}{1-H_i(s,t)}, \quad p_{i,j}(s)>0,(i,j)\in E,(s,t)\in[0,\infty]. \tag{8}$$

3. Estimation methods

In this section, the assumptions for the proposed estimation methods are introduced. Then, the estimation methods are described in detail.

3.1. *Assumptions*

1. The device has N known levels of degradation states from perfect functioning to complete failure.
2. Over time, the device may degrade from its current state to a degraded state or to the failure state; that is, the device is subject to multiple competing deterioration processes. We note here that, for the proposed parametric estimation method, the set of states accessible from state i, (FS_i), need to be known.
3. A degradation transition between two states follows an arbitrary distribution, which may depend on the two states involved in the transition, the time that the last state is reached, the time spent at the current state, and/or the total age of the device. Depending on the distribution of each $\lambda_{i,j}(s,t)$, the number of characteristic parameters for each transition rate function may vary. The set of all transition rate functions in the model is represented by Γ, which is regarded as an unknown set of parameters.
4. Transitions between states are left-to-right only; that is, no repair or maintenance is considered.
5. A single condition monitoring indicator is used for health monitoring. Due to the discrete representation of health states and the condition monitoring indicator, the stochastic relationship between each health state (i) and the condition monitoring indicator is represented by a matrix called the observation probability matrix (**B**). The element at the ith row and the jth column of this matrix is denoted by $b_i(j)$, where $b_i(j)=\Pr(Y_t=v_j \mid Z_t=i)$, for $t>0$.
6. To obtain the required training data, the observation process is monitored at discrete points while the device is operating. The observation intervals do not need to be equivalent. The observation time points for the kth sequence of observations are denoted by $t_1^{(k)},\ldots,t_{l_k}^{(k)}$, where l_k is the number of observation points in the kth sequence of states. The observation interval can be larger or smaller than the inspection interval. However, the observation process deals with CM data, which may be extracted easily and relatively inexpensively through one or multiple sensors. Such CM data are usually continuously extracted while the device is performing its intended functions. However, inspections are costly and may be subject to physical complexities (e.g., shutdown, assembly, and disassembly). Therefore, the inspection time interval is practically longer than the observation time interval.
7. Except for the failure state, which may be self-announcing, all other health states are identified by inspection. In this article, it is assumed that the inspection procedure can identify the true health state level without any error. The device is inspected at limited inspection points ($n\delta, n=1,2,\ldots$) in order to identify the current state. It should be pointed out here that under continuous inspection (Liao *et al.*, 2006), δ tends to zero. Similar

to Banjevic *et al*. (2001), we focus on cases where the inspection interval for the training data is short enough, so that at most one transition may occur between two inspection points. The data observed at inspection points are naturally interval-censored. In this article, the transition times are approximated with the midpoint of the inspection intervals for mathematical simplicity; that is, $Z_t = Z_{(n-0.5)\delta}$, $(n-0.5)\delta \le t < (n+0.5)\delta$. There are also other approaches for dealing with interval-censored inspection data, as discussed in Foucher *et al*. (2007). It should be pointed out that if the inspection interval is relatively large with respect to the times of the transitions—that is, more than one transition can occur in an interval—the approximation of the transition times with the midpoints of the inspection intervals may lead to a significant error. Dealing with such scenarios is outside the scope of this article. As pointed out in Section 1, it is assumed in this article that the supervised training data are available and the inspection interval used in the supervised training data set is relatively small. This sets the boundary for the applicability of the proposed supervised estimation methods proposed in this article. Additional discussions on how such training data may be obtained are given in Section 1.

8. The data required for estimation are K temporal independent sequences of states and observations as $Q^{(k)} = [X_1^{(k)}, T_1^{(k)}, X_2^{(k)}, T_2^{(k)}, \ldots, X_{n_k}^{(k)}, T_{n_k}^{(k)}, h_{n_k}^{(k)}]$ and $O^{(k)} = [Y_{t_1^{(k)}}^{(k)}, Y_{t_2^{(k)}}^{(k)}, \ldots, Y_{t_{l_k}^{(k)}}^{(k)}]$, respectively, where $1 \le k \le K$, n_k is the discrete-time counting process of the number of transitions in $Q^{(k)}$, $h_{n_k}^{(k)}$ (backward recurrence time) is the sojourn time in the last visited state ($X_{n_k}^{(k)}$), and $t_i^{(k)}$ is the ith observation time point of the kth sequence of observations. We note here that $Q^{(k)}$ and $O^{(k)}$ contain either a complete history of a failure ($X_{n_k}^{(k)} = N$, $h_{n_k}^{(k)} = 0$) or a suspended history of a failure ($X_{n_k}^{(k)} \ne N$, $h_{n_k}^{(k)} \ne 0$). Except for all transitions to the failure state, which may be self-announcing, other transitions are realized only by inspection. As mentioned earlier, $T_i^{(k)}$, $1 \le i < n_k$, is approximated by the midpoint in the corresponding inspection interval. This can be shown as $T_i^{(k)} = (j-0.5)\times\delta \leftrightarrow Z_{j\times\delta} = X_i^{(k)}$, $Z_{(j-1)\times\delta} = X_{i-1}^{(k)}$. It is clear that under continuous inspection ($\delta \to 0$), the exact values of $T_i^{(k)}$, $1 \le i < n_k$ are known; that is, the proposed estimation methods provide more accurate estimates for the unknown parameters.

3.2. *Parametric estimation method*

This section presents a parametric estimation method to estimate the unknown parameters of an NHCTHSMP associated with a multistate device. For notational convenience, the compact notation $\theta = (N, \Gamma, V, B)$ is used to indicate the characteristic parameters of the NHCTHSMP model. As N and M are assumed to be known, Γ and $\mathbf{B}$ are the only unknown parameters to be estimated. In a supervised estimation procedure, as both the sequences of observations and states are available, using the complete likelihood function L, the following optimization problem can be employed to find the maximum likelihood estimators of the elements of Γ and $\mathbf{B}$:

$$\begin{aligned}\operatorname{Max} L(K) &= \prod_{k=1}^{K} \Pr\left(Q^{(k)}, O^{(k)}|\theta\right) \overset{L'(K)=\log L(K)}{\Rightarrow} \operatorname{Max} L'(K)\\ &= \log\left(\prod_{k=1}^{K} \Pr\left(O^{(k)}\middle|Q^{(k)},\theta\right)\times \Pr\left(Q^{(k)}|\theta\right)\right)\\ &= \sum_{k=1}^{K} \log\left(\Pr\left(O^{(k)}\middle|Q^{(k)},\theta\right)\times \Pr\left(Q^{(k)}|\theta\right)\right)\\ &= \sum_{k=1}^{K} \log\left(\Pr\left(O^{(k)}\middle|Q^{(k)},\theta\right)\right)\\ &\quad + \sum_{k=1}^{K} \log\left(\Pr\left(Q^{(k)}|\theta\right)\right). \end{aligned} \tag{9}$$

Therefore, $\theta^* = \arg\max_\theta(L'(K))$. The log-likelihood function $L'(K)$ has two independent elements that can be optimized separately. The first element, $L_1'(K) = \sum_{k=1}^{K} \log(\Pr(O^{(k)}|Q^{(k)},\theta))$, depends only on $\mathbf{B}$, and the second element, $L_2'(K) = \sum_{k=1}^{K} \log(\Pr(Q^{(k)}|\theta))$, depends only on Γ. The remaining part of this section is devoted to showing how to optimize $L_1'(K)$ and $L_2'(K)$ to estimate $\mathbf{B}$ and Γ independently.

3.2.1. *Optimizing $L_1'(K) = \sum_{k=1}^{K} \log(\Pr(O^{(k)}|Q^{(k)},\theta))$ to estimate $\boldsymbol{B}$*

Let $\psi_{i,j}^{(k)}$ be the number of cases in the observation sequence $O^{(k)}$ where v_j is observed while the device is in state i, given as

$$\psi_{i,j}^{(k)} = \sum_{m=1}^{l_k} 1_{\{Z_{t_m^{(k)}}=i,\, Y_{t_m^{(k)}}=v_j\}},$$

and $q_i^{(k)}$ be the state of the device at time $t_i^{(k)}$ for the sequence $O^{(k)}$, given as

$$q_i^{(k)} = \sum_{j=1}^{n_k-1} X_j^{(k)} \times 1_{\{T_j^{(k)} \le t_i^{(k)} < T_{j+1}^{(k)}\}} + X_{n_k}^{(k)} \times 1_{\{T_{n_k}^{(k)} \le t_i^{(k)}\}}.$$

Now, $L_1'(K)$ can be simplified as

$$\begin{aligned} L_1'(K) &= \sum_{k=1}^{K} \log\left(\Pr\left(O^{(k)}\middle|Q^{(k)},\theta\right)\right)\\ &= \sum_{k=1}^{K} \log\left(\prod_{t=1}^{l_k} b_{q_t^{(k)}}\left(O_t^{(k)}\right)\right)\\ &= \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{M} \psi_{i,j}^{(k)} \log(b_i(j)). \end{aligned} \tag{10}$$

Using the constraint $\sum_{j=1}^{M} b_i(j) = 1, i \in E$, adding the Lagrange multiplier ϑ_1, and setting the associated derivative of $L_1'(K) + \vartheta_1(1 - \sum_{j=1}^{M} b_i(j))$ with respect to $b_i(j), i \in E, 1 \le j \le M$ and ϑ_1 equal to zero, we have

$$\begin{cases} \dfrac{\partial}{\partial b_i(j)}\left(L_1'(K) + \vartheta_1\left(1 - \sum_{j=1}^{M} b_i(j)\right)\right) = 0 \\ \quad \rightarrow \dfrac{\sum_{k=1}^{K} \psi_{i,j}^{(k)}}{b_i(j)} - \vartheta_1 = 0 \rightarrow b_i(j) = \dfrac{\sum_{k=1}^{K} \psi_{i,j}^{(k)}}{\vartheta_1} \\ \dfrac{\partial}{\partial \vartheta_1}\left(L_1'(K) + \vartheta_1\left(1 - \sum_{j=1}^{M} b_i(j)\right)\right) = 0 \\ \quad \rightarrow \sum_{j=1}^{M} \dfrac{\sum_{k=1}^{K} \psi_{i,j}^{(k)}}{\vartheta_1} = 1 \rightarrow \vartheta_1 = \sum_{k=1}^{K}\sum_{j=1}^{M} \psi_{i,j}^{(k)} \\ \qquad \rightarrow \hat{b}_i(j) = \dfrac{\sum_{k=1}^{K} \psi_{i,j}^{(k)}}{\sum_{k=1}^{K}\sum_{j=1}^{M} \psi_{i,j}^{(k)}}, \\ \qquad i \in E, 1 \le j \le M. \end{cases} \quad (11)$$

Now, Equation (11) can be directly used to find the maximum likelihood estimates of all the elements of **B**.

3.2.2. *Optimizing* $L_2'(K) = \sum_{k=1}^{K} log(Pr(Q^{(k)}|\theta))$ *to estimate* Γ

Given that $(X_0^{(k)}, T_0^{(k)}) = (1, 0), 1 \le k \le K$—that is, all sequences of states begin from state one and $t = 0$—we have

$$\begin{aligned} L_2'(K) &= \sum_{k=1}^{K} \log\left(\Pr\left(Q^{(k)}|\theta\right)\right) \\ &= \sum_{k=1}^{K} \log\left(\prod_{z=1}^{n_k} \Pr\left(X_z = X_z^{(k)}, T_z - T_{z-1}\right.\right. \\ &\quad \left.\left. = d_z^{(k)} \middle| X_{z-1} = X_{z-1}^{(k)}, T_{z-1} = T_{z-1}^{(k)}, \theta\right)\right) \\ &\quad + \sum_{k=1}^{K} \log\left(\Pr\left(T_{n_k+1} - T_{n_k} > h_{n_k}^{(k)} \middle| X_{n_k} = X_{n_k}^{(k)}, \right.\right. \\ &\quad \left.\left. T_{n_k} = T_{n_k}^{(k)}, \theta\right)\right), \end{aligned} \quad (12)$$

where $d_z^{(k)}$ is the time between the $(z-1)$th and zth transition in the kth sequence of states ($d_z^k = T_z^{(k)} - T_{z-1}^{(k)}$). Now let $\chi_{i,j,u,t}$ be the total number of cases in all K sequences of states where the sequence $(X_n^{(k)} = i, X_{n+1}^{(k)} = j, T_n^{(k)} = u, T_{n+1}^{(k)} - T_n^{(k)} = t, 1 \le k \le K, 0 \le n \le n_k - 1)$ exists, $\chi'_{i,u,t}$ be the total number of cases in all K sequences of states where the sequence $(X_{n_k}^{(k)} = i, T_{n_k}^{(k)} = u, t < T_{n_k+1}^{(k)} - T_{n_k}^{(k)}, 1 \le k \le K)$ exists, $\Theta^{(K)}$ be the set of all possible combinations of (i,j,u,t) in the form of $(X_n^{(k)} = i, X_{n+1}^{(k)} = j, T_n^{(k)} = u, T_{n+1}^{(k)} - T_n^{(k)} = t, 1 \le k \le K, 0 \le n \le n_k - 1)$, and $\Theta'^{(K)}$ be the set of all possible combinations of (i,u,t) in the form of $(X_{n_k}^{(k)} = i, T_{n_k}^{(k)} = u, t < T_{n_k+1}^{(k)} - T_{n_k}^{(k)}, 1 \le k \le K)$ as

$$\chi_{i,j,u,t} = \sum_{k=1}^{K}\sum_{n=0}^{n_k-1} 1_{\{X_n^{(k)}=i, X_{n+1}^{(k)}=j, T_n^{(k)}=u, T_{n+1}^{(k)}-T_n^{(k)}=t\}}, \quad (i,j) \in E, (u,t) \in [0,\infty], \quad (13)$$

$$\chi'_{i,u,t} = \sum_{k=1}^{K} 1_{\{X_{n_k}^{(k)}=i, T_{n_k}^{(k)}=u, h_{n_k}^{(k)}=t\}}, \quad i \in E, (u,t) \in [0,\infty]. \quad (14)$$

Now, using the relationship given in Equations (12) to (14), we have

$$\begin{aligned} L_2'(K) &= \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log(\lambda_{i,j}(u,t)) \\ &\quad + \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t}\left(-\int_0^t \sum_z \lambda_{i,z}(u,x)dx\right) \\ &\quad + \sum_{(i,u,t)\in\Theta'^{(K)}} \chi'_{i,u,t}\left(-\int_0^t \sum_z \lambda_{i,z}(u,x)\mathrm{d}x\right). \end{aligned} \quad (15)$$

See Appendix A for the detailed steps in finding Equation (15). Equation (15) can be divided into $N-1$ independent expressions so that each can be optimized separately. The following shows the expression for $L_2'^{(r)}(K)$, which contains only transitions out of state r:

$$\begin{aligned} L_2'^{(r)}(K) &= \sum_{(i,j,u,t|i=r)\in\Theta^{(K)}} \left(\chi_{i,j,u,t} \log\left(\lambda_{i,j}(u,t)\right)\right. \\ &\quad \left. + \chi_{i,j,u,t}\left(-\int_0^t \sum_z \lambda_{i,z}(u,x)\mathrm{d}x\right)\right) \\ &\quad + \sum_{(i,u,t|i=r)\in\Theta'^{(K)}} \chi'_{i,u,t}\left(-\int_0^t \sum_z \lambda_{i,z}(u,x)\mathrm{d}x\right). \end{aligned} \quad (16)$$

By replacing the parametric form of $\lambda_{i,j}(u,t)$ in Equation (16), this equation becomes a function in terms of unknown characteristic parameters of transition rate functions out of state r. It should be noted here that, depending on the type (statistical form) of $\lambda_{i,j}(u,t)$, it may be possible to achieve closed-form estimators for all characteristic parameters. For example, when all transitions follow the HCTSMP structure with Weibull transition rates $(\lambda_{i,j}(u,t) = \beta_{i,j}\alpha_{i,j}^{-\beta_{i,j}} t^{\beta_{i,j}-1})$, the closed-form expression for the maximum likelihood estimation of $\alpha_{i,j}$ and the simplified likelihood equation for $\beta_{i,j}$ considering no censored

data can be expressed as follows:

$$\begin{cases} \hat{\alpha}_{i,j} = \left(\left(\sum_b \sum_u \sum_d \chi_{i,b,u,d} \times d^{\hat{\beta}_{i,j}}\right) \Big/ \left(\sum_u \sum_d \chi_{i,j,u,d}\right)\right)^{1/\hat{\beta}_{i,j}} \\ \hat{\beta}_{i,j} = \sum_u \sum_d \chi_{i,j,u,d} \Big/ \left(\left(\left(\sum_b \sum_u \sum_d \chi_{i,b,u,d} \times d^{\hat{\beta}_{i,j}} \times \log(d)\right) \times \sum_u \sum_d \chi_{i,j,u,d} \Big/ \left(\sum_b \sum_u \sum_d \chi_{i,b,u,d} \times d^{\hat{\beta}_{i,j}}\right)\right) - \left(\sum_u \sum_d \chi_{i,j,u,d} \times \log(d)\right)\right) \end{cases}. \quad (17)$$

As the CTMP and EDCTSMP are specific cases of the HCTSMP, closed-form expressions for transition rates (in a Weibull form) in these structures can also be derived from Equation (17). In Appendix B, the steps to calculate the approximate Fisher matrix confidence interval of the maximum likelihood estimates (Wayne, 1982) of parameters $\alpha_{i,j}$ and $\beta_{i,j}$ given in Equation (17) considering the asymptotic normality of the maximum likelihood estimation are illustrated.

3.3. *Nonparametric estimation method*

As discussed earlier, for the nonparametric estimation method, instead of finding the parameters of the transition rate functions between states, point estimators for $\lambda_{i,j}(u, t)$ and $Q_{i,j}(u, t)$ as the two fundamental describers of the degradation process and also for other important quantities of interest ($p_{i,j}(u)$, $H_i(u, t)$, $G_{i,j}(u, t)$), given in Equations (5) to (7), are constructed. We note here that, as u and t are continuous variables, directly estimating the above important quantities (for $(s, t) \in [0, \infty]$) from available training data is impossible and, instead, the estimates are initially calculated only for limited discrete points. After the estimated values for these discrete points are calculated from available training data, the estimated values for the remaining points located between these discrete points can be approximated using a piecewise function. This idea is adopted from the concept of piecewise constant function in Ouhbi and Limnios (1999), which was used to approximate the transition rate function. As we deal only with available inspection data, the discrete points for which nonparametric estimates can be directly calculated from available training data are the elements of $\Theta^{(K)}$ and $\Theta'^{(K)}$. Now, we first start from $L_2'(K)$ and express it in terms of possible discrete points and then use it to estimate transition rate $\lambda_{i,j}(u, t)$, where $(i, j, u, t) \in \Theta^{(K)}$. Hereafter, for notational convenience, superscript p is used to represent any discrete element from $\Theta^{(K)}$, which can be directly estimated from the training data. Now, if we rewrite Equation (15) in terms of $\lambda^{\mathrm{p}}_{i,j}(u, t, K)$, $(i, j, u, t) \in \Theta^{(K)}$, we have

$$\begin{aligned} L_2'^{\mathrm{p}}(K) = & \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log\left(\lambda^{\mathrm{p}}_{i,j}(u, t, K)\right) \\ & - \sum_{(i,j,u,t)\in\Theta^{(K)}} \left(\sum_b \chi_{i,b,u,t}\right) \\ & \times \left(\varepsilon^1_{i,j}(u, t) + \left(\sum_{x\le t} \lambda^{\mathrm{p}}_{i,j}(u, x, K) \times \delta\right)\right) \\ & - \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi'_{i,u,t} \Bigg(\varepsilon^1_{i,j}(u, t) \\ & + \left(\sum_{x\le t} \lambda^{\mathrm{p}}_{i,j}(u, x, K) \times \delta\right)\Bigg), \end{aligned} \quad (18)$$

where $\varepsilon^1_{i,j}(u, t)$ is the error of finding the integration given in Equation (18) by discretization as

$$\begin{aligned} \varepsilon^1_{i,j}(u, t) = & \left(\int_0^t \lambda^{\mathrm{p}}_{i,j}(u, x, K)\mathrm{d}x\right) \\ & - \left(\sum_{x\le t} \lambda^{\mathrm{p}}_{i,j}(u, x, K) \times \delta\right). \end{aligned} \quad (19)$$

Now, if we rewrite Equation (18) in terms of a single point $\lambda^{\mathrm{p}}_{i,j}(u, t, K)$, we have

$$\begin{aligned} & L_2'^{\mathrm{p}}\left(\lambda^{\mathrm{p}}_{i,j}(u, t, K), K\right) \\ & = \chi_{i,j,u,t} \log\left(\lambda^{\mathrm{p}}_{i,j}(u, t, K)\right) - \left(\sum_b \chi_{i,b,u,t}\right) \times \varepsilon^1_{i,j}(u, t) \\ & \quad - \left(\sum_b \sum_{d\ge t} \chi_{i,b,u,d} \times \delta\right) \lambda^{\mathrm{p}}_{i,j}(u, x, K) - \chi'_{i,u,t} \\ & \quad \times \varepsilon^1_{i,j}(u, t) - \left(\sum_{d\ge t} \chi'_{i,u,d} \times \delta\right) \lambda^{\mathrm{p}}_{i,j}(u, t, K). \end{aligned} \quad (20)$$

Here, it is assumed that as $K \to \infty$, then $\varepsilon^1_{i,j}(u, t)/L_2'^{\mathrm{p}}(\lambda^{\mathrm{p}}_{i,j}(u, t, K), K) \to 0$; therefore, $\varepsilon^1_{i,j}(u, t)$ does not have a significant effect on the approached likelihood function $L_2'^{\mathrm{p}}(\lambda^{\mathrm{p}}_{i,j}(u, t, K), K)$ and can be neglected from Equation (20). Now, using Equation (20), $\lambda^{\mathrm{p}}_{i,j}(u, t, K)$ can be estimated as

$$\begin{aligned} & \frac{\partial L_2'^{\mathrm{p}}\left(\lambda^{\mathrm{p}}_{i,j}(u, t, K), K\right)}{\partial \lambda^{\mathrm{p}}_{i,j}(u, t, K)} = 0 \to \hat{\lambda}^{\mathrm{p}}_{i,j}(u, t, K) \\ & = \frac{\chi_{i,j,u,t}}{\delta \times \left(\sum_b \sum_{d\ge t} \chi_{i,b,u,d}\right) + \delta \times \left(\sum_{d\ge t} \chi'_{i,u,d}\right)}, \\ & \qquad (i, j, u, t) \in \Theta^{(K)}. \end{aligned} \quad (21)$$

If the structure of the degradation process is known to be HCTSMP or EDCTSMP, the estimators for the transition

rate functions of these processes are respectively:

$$\hat{\lambda}^{\mathrm{p}}_{i,j}(t,K) = \frac{\sum_u \chi_{i,j,u,t}}{\delta \times \left(\sum_b \sum_u \sum_{d\geq t} \chi_{i,b,u,d}\right) + \delta \times \left(\sum_u \sum_{d\geq t} \chi_{i,u,d}\right)}, \quad (i,j,u,t) \in \Theta^{(K)}, \tag{22}$$

$$\hat{\lambda}^{\mathrm{p}}_{i}(t,K) = \frac{\sum_b \sum_u \chi_{i,b,u,t}}{\delta \times \left(\sum_b \sum_u \sum_{d\geq t} \chi_{i,b,u,d}\right) + \delta \times \left(\sum_u \sum_{d\geq t} \chi_{i,u,d}\right)}, \quad (i,j,u,t) \in \Theta^{(K)}. \tag{23}$$

Now, using a piecewise function, it is possible to approximate the transition rate between two states at any point in addition to those directly calculated from Equation (21). To find $\hat{\lambda}_{i,j}(u,t,K)$, where $t_1 < t < t_2$ and the estimates for $\hat{\lambda}^{\mathrm{p}}_{i,j}(u,t_1,K)$ and $\hat{\lambda}^{\mathrm{p}}_{i,j}(u,t_2,K)$ are directly calculated from Equation (21), a Piecewise Linear Estimator (PLE) or a Piecewise Constant Estimator (PCE) can be used as follows:

$$\begin{cases} \hat{\lambda}_{i,j}(u,t,K) = \hat{\lambda}^{\mathrm{p}}_{i,j}(u,t_1,K) + (t-t_1) \\ \quad \times \dfrac{\hat{\lambda}^{\mathrm{p}}_{i,j}(u,t_2,K) - \hat{\lambda}^{\mathrm{p}}_{i,j}(u,t_1,K)}{t_2-t_1}, & \text{PLE} \cdot \\ \hat{\lambda}_{i,j}(u,t,K) = \hat{\lambda}^{\mathrm{p}}_{i,j}(u,t_2,K), & \text{PCE} \end{cases} \tag{24}$$

A similar type of approximating the transition rate function with a piecewise discrete function was employed in past studies, such as Ouhbi and Limnios (1999) and Limnios and Ouhbi (2006). To estimate $p^{\mathrm{p}}_{i,j}(u,K)$, we express $L'^{\mathrm{p}}_2(K)$ in a proper form as

$$L'^{\mathrm{p}}_2(K) = \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log\left(p^{\mathrm{p}}_{i,j}(u,K) G'^{\mathrm{p}}_{i,j}(u,t,K)\right) + \sum_{(i,u,t)\in\Theta'^{(K)}} \chi'_{i,u,t} \left(-\int_0^t \sum_z \lambda^{\mathrm{p}}_{i,z}(u,x,K)\mathrm{d}x\right). \tag{25}$$

Now, using the equalities $\sum_{b=1}^{N} p^{\mathrm{p}}_{i,b}(u,K) = 1, i \in E$ and adding the Lagrange multiplier ϑ_2, the following partial log-likelihood function can be obtained:

$$L'^{\mathrm{p}}_{2,1}\left(p^{\mathrm{p}}_{i,j}(u,K),K\right) = \sum_d \chi_{i,j,u,d} \log\left(p^{\mathrm{p}}_{i,j}(u,K)\right) + \vartheta_2 \left(1 - \sum_{b=1}^{N} p^{\mathrm{p}}_{i,b}(u,K)\right). \tag{26}$$

To find the estimator for $p^{\mathrm{p}}_{i,j}(u,K)$, Equation (26) can be maximized with respect to $p^{\mathrm{p}}_{i,j}(u,K)$ as

$$\begin{cases} \dfrac{\partial L'^{\mathrm{p}}_{2,1}\left(p^{\mathrm{p}}_{i,j}(u,K),K\right)}{\partial p^{\mathrm{p}}_{i,j}(u,K)} = 0 \rightarrow \dfrac{\sum_d \chi_{i,j,u,d}}{p^{p}_{i,j}(u,K)} - \vartheta_2 = 0 \\ \quad \rightarrow p^{\mathrm{p}}_{i,j}(u,K) = \dfrac{\sum_d \chi_{i,j,u,d}}{\vartheta_2} \\ \dfrac{\partial L'^{\mathrm{p}}_{2,1}\left(p^{\mathrm{p}}_{i,j}(u,K),K\right)}{\partial \vartheta_2} = 0 \rightarrow \left(1 - \sum_{j=1}^{N} p^{\mathrm{p}}_{i,j}(u,K)\right) = 0 \\ \quad \rightarrow \vartheta_2 = \sum_b \sum_d \chi_{i,b,u,d} \\ \qquad\qquad \rightarrow \hat{p}^{\mathrm{p}}_{i,j}(u,K) = \sum_d \chi_{i,j,u,d} \\ \qquad \times \left(\sum_b \sum_d \chi_{i,b,u,d}\right)^{-1}, \quad (i,j,u,d) \in \Theta^{(K)}. \end{cases} \tag{27}$$

In an analogous way as we did for transition rate functions, we can find $\hat{p}_{i,j}(s,K)$, where $(i,j) \in E, s \in [0,\infty]$. Now, using the equality $\int_0^\infty G'_{i,j}(u,x,K)\mathrm{d}x = 1, (i,j) \in E, u \in [0,\infty]$, and adding the Lagrange multiplier ϑ_3, the following partial log-likelihood function can be obtained from Equation (25):

$$L'_{2,2}(K) = \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log(p_{i,j}(u,K)) + \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log(G'_{i,j}(u,t,K)) + \vartheta_3 \left(1 - \int_0^\infty G'_{i,j}(u,x,K)\mathrm{d}x\right). \tag{28}$$

In order to find $\hat{G}'_{i,j}(u,t,K)$ from Equation (28), in an analogous way as we did to estimate transition rates, we approximate $\int_0^\infty G'_{i,j}(u,x,K)\mathrm{d}x$ by discretization and consequently find the estimator for $G^{\mathrm{p}}_{i,j}(u,t,K)$ as

$$\hat{G}^{\mathrm{p}}_{i,j}(u,t,K) = \left(\sum_{d\leq t} \chi_{i,j,u,d}\right) \times \left(\sum_d \chi_{i,j,u,d}\right)^{-1}, \quad (i,j,u,t) \in \Theta^{(K)}. \tag{29}$$

Details of steps in finding Equation (29) are given in Appendix C. Now, based on the relationship given in Equation (8), the estimator for the kernel function can be expressed as

$$\hat{Q}^{\mathrm{p}}_{i,j}(u,u+t,K) = \left(\sum_{d\leq t} \chi_{i,j,u,d}\right) \times \left(\sum_b \sum_d \chi_{i,b,u,d}\right)^{-1}$$

$$\rightarrow \hat{Q}^{\mathrm{p}}_{i,j}(u,t,K) = \left(\sum_{d\le t-u} \chi_{i,j,u,d}\right) \times \left(\sum_{b}\sum_{d} \chi_{i,b,u,d}\right)^{-1}, \quad (i,j,u,t) \in \Theta^{(K)}. \tag{30}$$

Finally, the nonparametric estimator for $\hat{H}^{\mathrm{p}}_{i}(u,t,K)$ is

$$\hat{H}^{\mathrm{p}}_{i}(u,t,K) = \left(\sum_{b}\sum_{d\le t} \chi_{i,b,u,d}\right) \times \left(\sum_{b}\sum_{d} \chi_{i,b,u,d}\right)^{-1}, \quad (i,j,u,t) \in \Theta^{(K)}. \tag{31}$$

4. Experimental results

This section presents a simulation-based numerical example to evaluate the accuracy and empirical consistency of the proposed estimation methods. Monte Carlo simulation is employed to generate random sequences of states and observations according to an NHCTHSMP as the inputs for the parameter estimation methods. For both parametric and nonparametric methods, the estimated values are compared with preset values that were originally used to simulate random trajectories (run-to-failure data) of the semi-Markov process. We used MATLAB 2008 on a standalone PC with CPU 2.3 GHz and 16 GB of RAM for the experiment. The device used in the example has four levels of health states, labeled normal, slight damage, medium damage, and failure. Figure 2 shows the transition diagram of this device.

The simulation was run for $K = 20, 50, 100, 200$, and 500. In addition, to minimize the effect of randomness, the experiment was repeated 100 times. The Mean Squared Error (MSE) is used as the main measure to evaluate the estimation results. A single condition-monitoring indicator is recorded every 2.4 hours (0.1 day) without suspending the operation. It is assumed that the observed condition monitoring indicator can take eight possible discrete outcomes ($M = 8$). Therefore, the observation probability matrix (**B**) has four rows and eight columns. For the transition distributions, the Weibull distribution is used as it is the most commonly used distribution to represent degradation. To consider different types of transitions, it is assumed that the transition 2-4 depends on the total age of the device

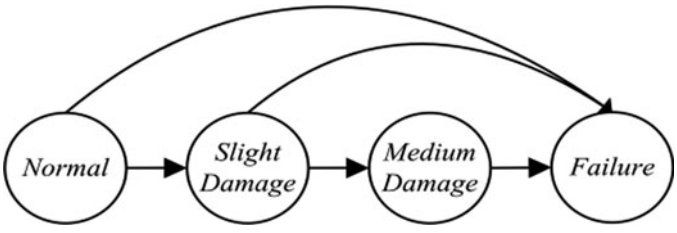

Fig. 2. Multistate device with four discrete health states.

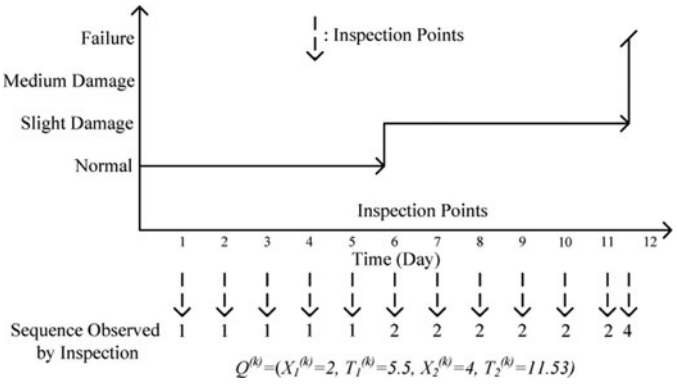

Fig. 3. A sample simulated realization of the degradation process obtained by inspection.

as $\lambda_{2,4}(s,t) = \alpha_{2,4}^{-\beta_{2,4}} \times \beta_{2,4} \times (t+s)^{\beta_{2,4}-1}$, whereas all other transitions depend only on the sojourn time at the current state as $\lambda_{i,j}(s,t) = \alpha_{i,j}^{-\beta_{i,j}} \times \beta_{i,j} \times t^{\beta_{i,j}-1}$. The preset values of the parameters used in the example are as follows:

$$\Gamma : \begin{cases} \alpha_{(1,2)} = 15, \alpha_{(1,4)} = 21, \alpha_{(2,3)} = 12, \\ \alpha_{(2,4)} = 28, \alpha_{(3,4)} = 7 \\ \beta_{(1,2)} = 8, \beta_{(1,4)} = 3, \beta_{(2,3)} = 6, \\ \beta_{(2,4)} = 12, \beta_{(3,4)} = 4. \end{cases},$$

$$\mathbf{B} = [b_i(j)] = \begin{bmatrix} 0.45 & 0.35 & 0.15 & 0.05 & 0 & 0 & 0 & 0 \\ 0.05 & 0.1 & 0.3 & 0.35 & 0.1 & .05 & 0.05 & 0 \\ 0 & 0 & 0.05 & 0.05 & 0.2 & 0.25 & 0.45 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The above values are used to simulate random trajectories of the degradation process. The total number of parameters associated with the degradation process is five each for α and β. As the failure state is assumed to be directly observable with the associated observation value of eight, the total number of unknown parameters associated with the observation process is 21 (3×7). It is assumed that the device is inspected once a day ($\delta = 1$) to identify the current health state. Figure 3 presents a sample realization of states obtained from simulation and its corresponding realization obtained from inspection ($Q^{(k)}$). Figure 4 presents a sample realization of CM indicators obtained from simulation and the corresponding realization obtained at CM points ($O^{(k)}$).

4.1. *Discussions on parametric estimation*

We first evaluate the parametric estimation method considering five different cases of K and 100 simulation runs. Table 1 presents the true values of each parameter used for simulation, the average, the STandard Deviation (STD), and the MSE for the estimated parameters for all cases of K based on the 100 simulation runs. It can be verified from Table 1 that the averages of the estimated values is reasonably close to the true values, particularly when a larger number of life histories (K) is used for estimation. In other

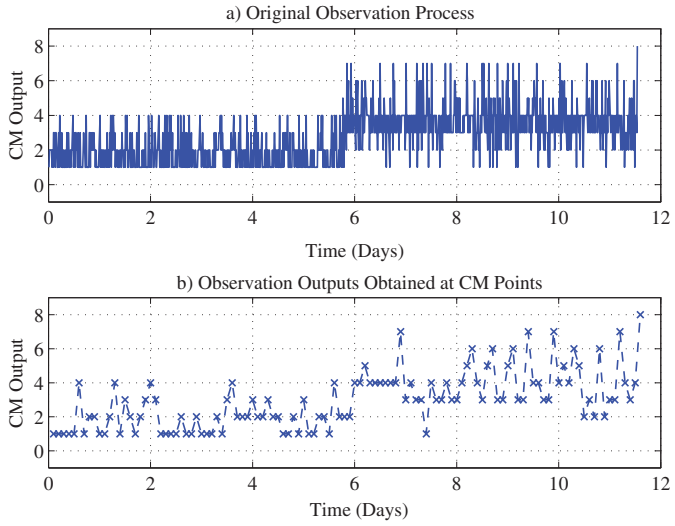

Fig. 4. A sample simulated realization of the observation process (color figure provided online).

words, as expected, the mean estimated values approach the true values as K increases. Moreover, that the MSE values for the estimated parameters converge to zero as K increases is also verified.

The average MSE values of all three sets of parameters (α, β, and **B**) are shown in Fig. 5 for all cases of K. The consistent trends in Fig. 5 verify that MSE converges to zero as the number of available life histories for estimation (K) increases. In summary, it can be concluded that the estimated values with an acceptable level of accuracy are found, in particular for a sufficient training sample.

It should be pointed out that the approximate confidence interval for each run of the simulation can be calculated by finding the approximate variance of the maximum likelihood estimator (Appendix B) considering the asymptotic normality of the maximum likelihood estimation. In Table 2, the statistical performance of the 95% confidence intervals for each parameter of interest over 100 simulation runs is shown by the average confidence interval width and the coverage rate (percentage of covering the true values of the parameter in repeated sampling; Maswadah (2003)). In can be verified from Table 2 that, overall, the average confidence interval widths are relatively narrow, particularly as K increases. Moreover, the coverage rates over 100 runs are relatively close to the nominal value of 95%, particularly for larger K.

The performance of the parametric estimation method with respect to the main describers of the degradation process is shown by comparing MSE values as shown in Figs. 6 and 7. The results in Figs. 6 and 7 verify that, as K increases, the MSE converges to zero.

4.2. *Discussions on the nonparametric estimation*

This subsection evaluates the empirical consistency of the nonparametric estimation method proposed in this article. Similar to the parametric case, five cases of K and 100 simulation runs are considered. Transition rate functions and kernel functions are focused on as the main describers of the degradation process. In this article, only the estimation results for $\lambda_{1,2}(0, t)$ and $Q_{1,2}(0, t)$ are presented. Our observations verified that the average of the estimated values converges to the true value as K increases. The MSE values of the estimated parameters are shown in Figs. 8 and 9. The results in these figures verify that as K increases, the MSE converges to zero.

Table 1. Parametric estimation results for 100 runs

Parameters	$\alpha_{1,2}$	$\alpha_{1,4}$	$\alpha_{2,3}$	$\alpha_{2,4}$	$\alpha_{3,4}$	$\beta_{1,2}$	$\beta_{1,4}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{3,4}$
True values	*15*	*21*	*12*	*28*	*7*	*8*	*3*	*6*	*12*	*4*
Mean										
$K = 20$	14.92	21.62	12.06	28.01	6.93	8.78	3.67	6.52	15.54	4.43
$K = 50$	14.97	21.40	12.04	28.07	6.98	8.45	3.21	6.28	12.88	4.11
$K = 100$	14.99	21.65	12.02	28.09	6.98	7.97	3.02	5.97	12.57	4.02
$K = 200$	14.99	21.15	12.01	27.97	6.99	7.99	3.09	5.98	12.36	3.98
$K = 500$	15.01	21.02	12.01	27.98	7.01	7.91	3.01	5.87	12.12	3.98
STD										
$K = 20$	0.49	4.64	0.64	1.99	0.63	2.09	1.73	1.92	8.82	1.25
$K = 50$	0.31	3.38	0.43	0.76	0.40	1.24	1.20	0.95	3.16	0.66
$K = 100$	0.24	2.72	0.26	0.52	0.26	0.73	0.54	0.69	2.16	0.49
$K = 200$	0.17	1.48	0.18	0.35	0.18	0.51	0.41	0.48	1.32	0.32
$K = 500$	0.10	0.91	0.13	0.24	0.10	0.32	0.22	0.27	0.87	0.20
MSE										
$K = 20$	0.24	21.72	0.41	3.92	0.39	4.96	3.42	3.91	89.64	1.73
$K = 50$	0.10	11.45	0.18	0.58	0.16	1.73	1.47	0.96	10.65	0.44
$K = 100$	0.06	7.72	0.07	0.28	0.07	0.52	0.29	0.47	4.94	0.24
$K = 200$	0.03	2.19	0.03	0.12	0.03	0.26	0.17	0.23	1.85	0.10
$K = 500$	0.01	0.82	0.02	0.06	0.01	0.11	0.05	0.09	0.77	0.04

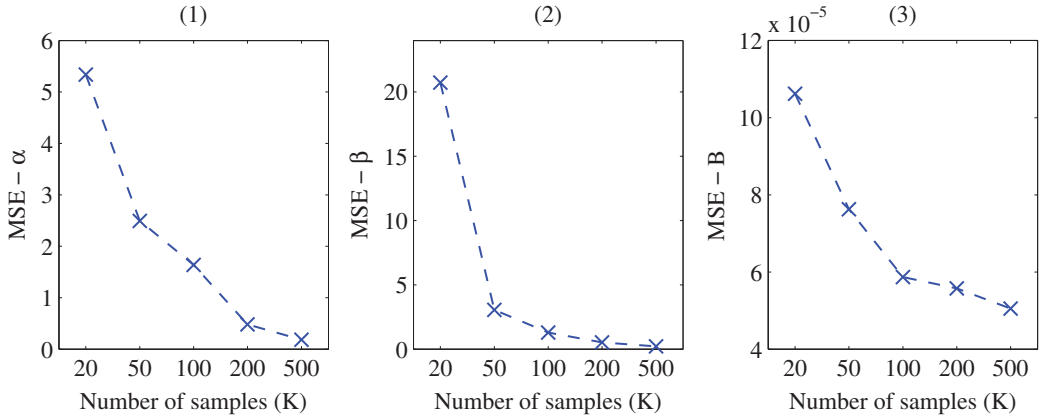

Fig. 5. Average MSE for the parametric estimation (color figure provided online).

4.3. *The effect of the number of states on the CPU time*

As the CPU time needed for parameter estimation could be a practical concern particularly when the number of states is large, we demonstrate the effect of the number of states on the computation time by varying the number of states from $N = 2$ to $N = 10$ for $K = 20$, 50, 100, 200, and 500. A simple degradation structure is considered for all cases where transitions are possible only to the neighboring state. Table 3 shows that increasing the number of states increases the computation time. Considering the fact that such an estimation process is performed offline in most practical cases, the results shown in Table 3 verify that even when the number of states is relatively large (e.g., $N = 10$), estimation results can be obtained in a reasonably feasible computation time (e.g., 145.34 minutes for $K = 500$).

4.4. *The effect of the inspection interval on the estimation results*

To demonstrate the effect of the inspection interval and the corresponding discretization bias on the parametric estimation results, we consider a larger inspection interval ($\delta = 4$ days) and re-estimate α and β. In this article, this effect is presented only for $K = 50$ while considering 100 simulation runs. The results shown in Table 4 verify that as the inspection interval increases, the accuracy of the estimated values decreases. To demonstrate the effect of inspection interval on the estimation results of the nonparametric estimation method, as an example, we estimated $Q_{1,2}(0, t)$ considering $\delta = 4$. Figure 10 verifies that increasing the inspection interval can influence the accuracy of the nonparametric estimated values. As stated in the Introduction of this article, the proposed estimation methods are not applicable for cases with large inspection intervals. Developing an estimation method that can use inspection data with very large inspection intervals is to be investigated in our future work.

4.5. *Comparison between the parametric and nonparametric estimation methods*

Although it is not the focus of this article, a comparison between the parametric and nonparametric estimation methods is briefly discussed here. Generally, the

Table 2. Confidence interval results for parametric estimation

Parameters	$\alpha_{1,2}$	$\alpha_{1,4}$	$\alpha_{2,3}$	$\alpha_{2,4}$	$\alpha_{3,4}$	$\beta_{1,2}$	$\beta_{1,4}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{3,4}$
True values	*15*	*21*	*12*	*28*	*7*	*8*	*3*	*6*	*12*	*4*
95% CI average width										
$K = 20$	1.86	24.80	2.54	5.37	2.26	7.17	6.48	6.73	25.88	4.77
$K = 50$	1.21	12.77	1.57	3.13	1.46	4.24	3.21	3.90	11.78	2.60
$K = 100$	0.89	9.05	1.13	2.13	1.03	2.79	2.14	2.56	7.84	1.75
$K = 200$	0.62	5.64	0.80	1.45	0.73	1.96	1.52	1.80	5.32	1.23
$K = 500$	0.40	3.52	0.51	0.93	0.46	1.23	0.94	1.12	3.26	0.78
95% CI coverage rate										
$K = 20$	96	92	94	85	90	94	93	93	85	93
$K = 50$	95	92	94	95	91	90	87	96	93	97
$K = 100$	93	94	98	97	97	98	95	95	92	92
$K = 200$	93	93	96	95	97	92	93	96	94	96
$K = 500$	95	94	95	96	98	93	97	95	93	95

Table 3. CPU time (minutes)—number of degradation states

	Number of states (N)								
K	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
20	1.37	2.64	5.09	13.07	14.00	15.55	18.87	24.94	44.06
50	5.42	10.87	13.41	19.51	22.75	29.05	42.33	47.37	65.22
100	8.57	13.72	18.87	28.14	37.67	41.72	52.05	53.97	66.70
200	14.93	17.31	26.11	29.19	41.41	48.41	65.64	77.72	94.71
500	35.77	44.86	47.48	55.49	69.46	85.11	102.06	120.81	145.34

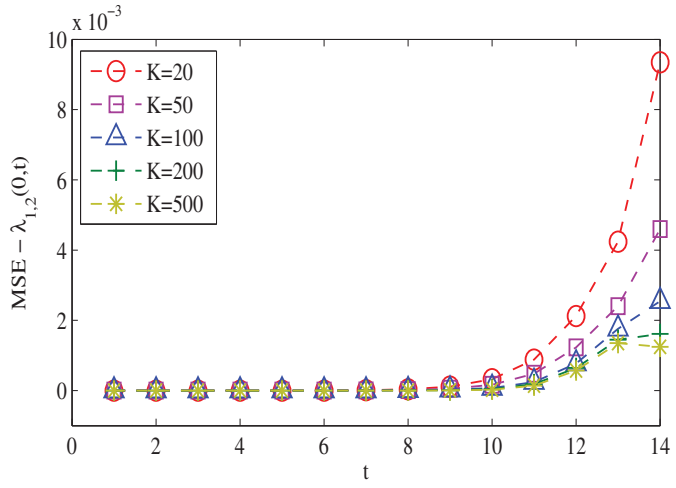

Fig. 6. MSE for parametric estimation of $\lambda_{1,2}(0, t)$ (color figure provided online).

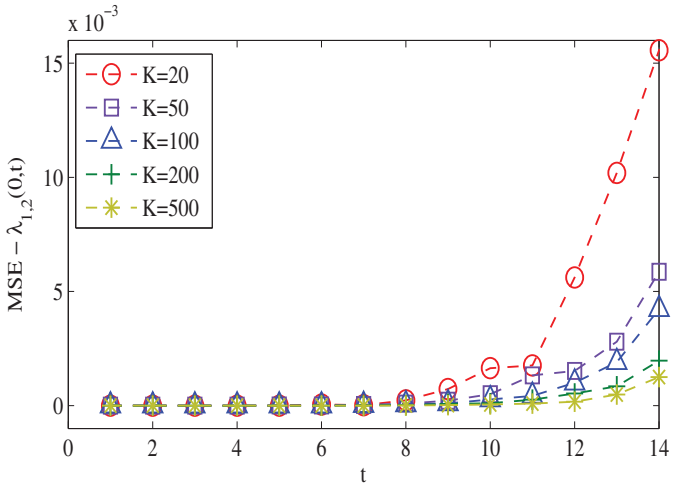

Fig. 8. MSE for nonparametric estimation of $\lambda_{1,2}(0, t)$ (color figure provided online).

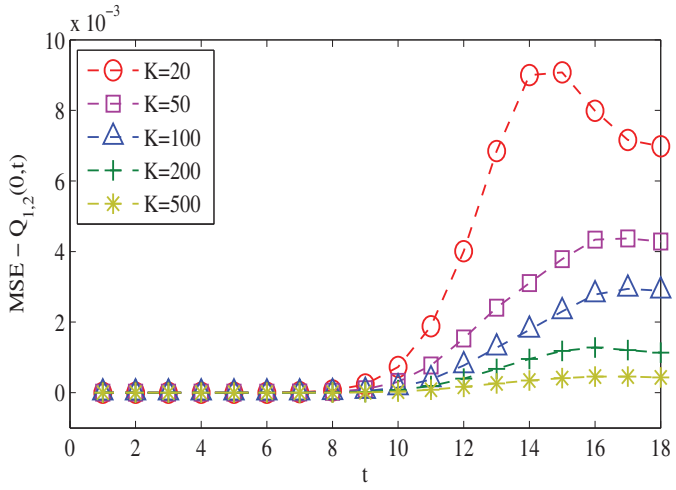

Fig. 7. MSE for parametric estimation of $Q_{1,2}(0, t)$ (color figure provided online).

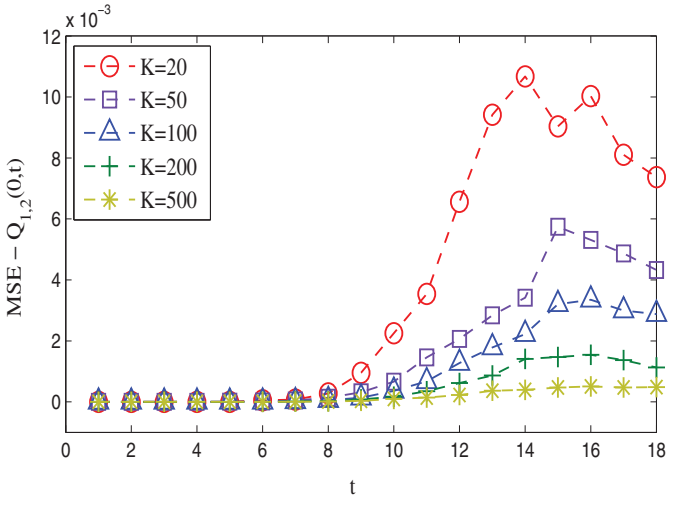

Fig. 9. MSE for nonparametric estimation of $Q_{1,2}(0, t)$ (color figure provided online).

Table 4. The effect of the inspection interval on parametric estimation ($K = 50$)

		$\delta = 1$		$\delta = 4$				$\delta = 1$		$\delta = 4$	
Parameter	*True value*	*Mean*	*MSE*	*Mean*	*MSE*	*Parameter*	*True value*	*Mean*	*MSE*	*Mean*	*MSE*
$\alpha_{1,2}$	15	14.97	0.1	15.12	0.18	$\beta_{1,2}$	8	8.45	1.73	7.07	2.00
$\alpha_{1,4}$	21	21.4	11.45	21.65	11.81	$\beta_{1,4}$	3	3.21	1.47	3.14	1.72
$\alpha_{2,3}$	12	12.04	0.18	12.37	0.41	$\beta_{2,3}$	6	6.28	0.96	5.20	1.68
$\alpha_{2,4}$	28	28.07	0.58	28.26	0.87	$\beta_{2,4}$	12	12.88	10.65	12.41	11.31
$\alpha_{3,4}$	7	6.98	0.16	7.03	0.19	$\beta_{3,4}$	4	4.11	0.44	3.42	0.65

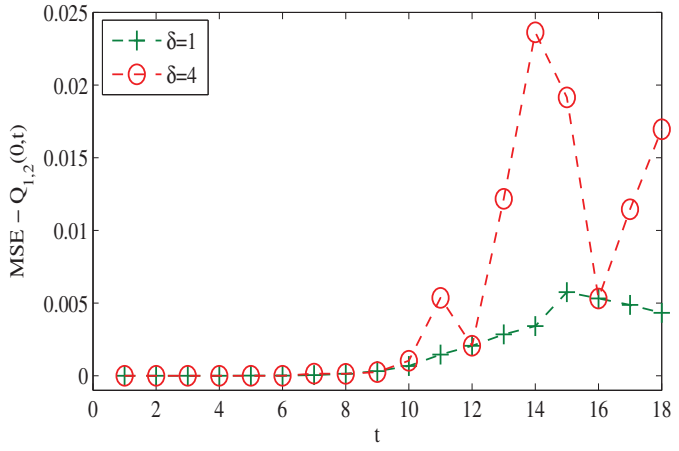

Fig. 10. The effect of the inspection interval on the nonparametric estimation ($K = 50$) (color figure provided online).

parametric estimation method deals with optimization complexities, whereas the nonparametric estimation method deals with simple closed-form solutions. Although both the parametric and nonparametric estimation methods may provide reasonable estimates for the unknown parameters of the model, it is expected that the parametric estimation method will provide relatively better estimates as it involves more information on the actual structure of the degradation process. To compare the estimation results for these two cases, as an example, the MSEs of the estimated values for $Q_{1,2}(0, t)$ and $\lambda_{1,2}(0, t)$ ($K = 50$ and 100 simulation runs) are shown in Figs. 11 and 12.

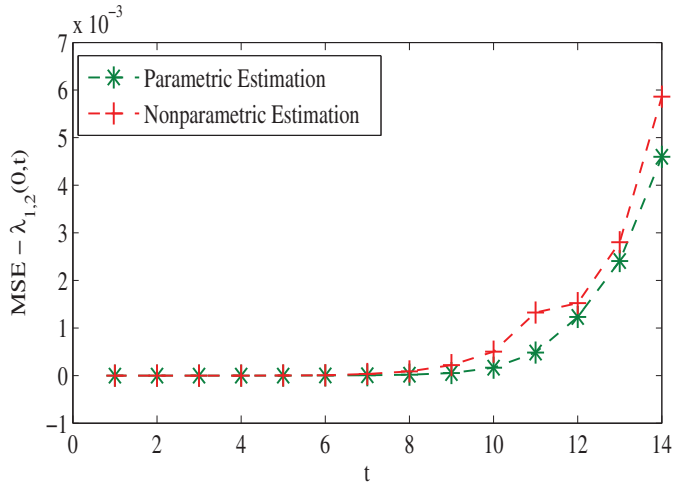

Fig. 11. Comparison between the parametric and nonparametric estimation methods ($K = 50$) for $\lambda_{1,2}(0, t)$ (color figure provided online).

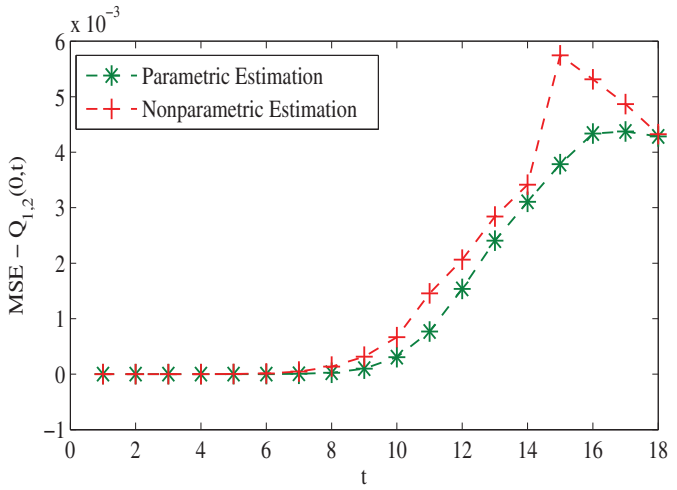

Fig. 12. Comparison between the parametric and nonparametric estimation methods ($K = 50$) for $Q_{1,2}(0, t)$ (color figure provided online).

Figures 11 and 12 verify that the parametric estimation method provides relatively more accurate estimates compared with the nonparametric estimation method. Table 5 provides the average CPU time needed for the parametric and nonparametric parameter estimation methods for each

Table 5. Average CPU time for the parametric and nonparametric estimation methods

	K				
	20	*50*	*100*	*200*	*500*
Parametric estimation (minutes)	3.58	8.70	12.6	19.24	45.26
Nonparametric estimation (seconds)	6.78	6.94	7.37	7.67	11.01

Table 6. Comparison between the supervised and unsupervised estimation methods

		Parameters									
		$\alpha_{1,2}$	$\alpha_{1,4}$	$\alpha_{2,3}$	$\alpha_{2,4}$	$\alpha_{3,4}$	$\beta_{1,2}$	$\beta_{1,4}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{3,4}$
Mean	Supervised	14.97	21.40	12.04	28.07	6.98	8.45	3.21	6.28	12.88	4.11
	Unsupervised	14.60	20.35	12.35	27.80	7.56	7.35	3.41	5.29	13.72	4.00
MSE	Supervised	0.10	11.45	0.18	0.58	0.16	1.73	1.47	0.96	10.65	0.44
	Unsupervised	0.29	12.95	0.96	0.79	0.68	1.75	1.66	1.97	15.56	0.47
CPU Time	Supervised	8.7 minutes									
	Unsupervised	127.3 minutes									

case of K. The results in Table 5 verify that, as expected, the nonparametric estimation method is much faster than the parametric estimation method because it deals with simple closed-form expressions. That the CPU time increases as the number of life histories used for estimation (K) increases is also verified for both methods.

4.6. *Comparison between the supervised and unsupervised estimation methods*

As discussed in the Introduction, the current article is an extension of our earlier work (Moghaddass and Zuo, 2012) where an unsupervised parametric estimation method was presented. It should be pointed out here that it is expected that as the supervised estimation method utilizes the degradation data, it gives better results than the unsupervised estimation method. To compare these two models, similar to our approach in this article throughout the numerical experiment section, the simulation experiment was repeated for 100 runs and the MSE for each parameter of interest based on each of the two estimation methods was calculated. The results shown in Table 6 for $K = 50$ verify that for all 10 parameters, the MSE of the supervised estimation method is lower than that from the unsupervised estimation method. Therefore, as expected, the supervised estimation method outperforms the unsupervised method for all 10 parameters. Comparing the average CPU times (100 runs) shown in Table 6 for these two cases also verifies that the supervised estimation method is faster than the unsupervised estimation method. The reason for this behavior is that while the supervised estimation method deals with maximizing a complete likelihood function employing both the degradation and observation processes, the unsupervised estimation method deals with an incomplete likelihood function, which employs only the observation process.

5. Conclusions

In this article, the problem of supervised estimation for the main characteristic parameters and important quantities of interest for the degradation and observation processes associated with a condition-monitored device is addressed. A parametric and a nonparametric estimation method are proposed to estimate the important quantities of interest of such a device. The degradation and observation processes were formulated by a flexible stochastic structure, namely, an NHCTHSMP. The application of this structure for diagnostic and prognostic health state recognition is to be investigated in our future work.

Acknowledgements

This work is financially supported by the Natural Science and Engineering Research Council of Canada (NSERC). The comments from the reviewers and the editor were very much appreciated.

References

Banjevic, D., Jardine, A.K.S., Makis, V. and Ennis, M. (2001) Control-limit policy and software for condition-based maintenance optimization. *INFOR*, **39**(1), 32–50.

Barbu, V. and Limnios, N. (2006a) Empirical estimation for discrete-time semi-Markov processes with applications in reliability. *Journal of Nonparametric Statistics*, **18**(7-8), 483–498.

Barbu, V. and Limnios, N. (2006b) Nonparametric estimation for failure rate functions of discrete time semi-Markov processes, in *Probability, Statistics and Modeling in Public Health, Special Volume in Honor of Professor Marvin Zelen*, Nikulin, M., Commenges, D. and C. Hubert, C. (eds), Springer, New York, NY, pp. 53–72.

Barbu, V.S. and Limnios, N. (2008) *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: Their Use in Reliability and DNA Analysis*, Birkhauser, New York, NY.

Blasi, A., Janssen, J. and Manca, R. (2004) Numerical treatment of homogeneous and non-homogeneous semi-Markov reliability models. *Communications in Statistics, Theory and Methods*, **33**(3), 697–714.

Chen, A. and Wu, G.S. (2007) A real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. *International Journal of Production Research*, **45**(15), 3351–3379.

Chiang, J.H. and Yuan, J. (2001) Optimal maintenance policy for a Markovian system under periodic inspection. *Reliability Engineering & System Safety*, **71**(2), 165–172.

Chryssaphinou, O., Limnios, N. and Malefaki, S. (2011) Multi-state reliability systems under discrete time semi-Markovian hypothesis. *IEEE Transactions on Reliability*, **60**(1), 80–87.

Dong, M. and He, D. (2007) Hidden semi-Markov model-based methodology for multi-sensor device health diagnosis and prognosis. *European Journal of Operational Research*, **178**(3), 858–878.

Dong, M., He, D., Banerjee, P. and Keller, J. (2006) Device health diagnosis and prognosis using hidden semi-Markov models. *International Journal of Advanced Manufacturing Technology*, **30**(7–8), 738–749.

Dong, M. and Peng, Y. (2011) Equipment PHM using non-stationary segmental hidden semi-Markov model. *Robotics and Computer-Integrated Manufacturing*, **27**(3), 581–590.

Foucher, Y., Giral, M., Soulillou, J.P. and Daures, J.P. (2007) A semi-Markov model for multistate and interval-censored data with multiple terminal events, application in renal transplantation. *Statistics in Medicine*, **26**(30), 5381–5383.

Hongzhi, T., Jianmin, Zh., Xisheng, J., Yunxian, J., Xinghui, Zh. and Liying, C. (2011) Experimental study on gearbox prognosis using total life vibration analysis, in *Proceedings of the Prognostics and System Health Management Conference*, pp. 1–6.

Huang, R., Xi, L., Li, X., Liu, R.C., Qiu, H. and Lee, J. (2007) Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, **21**(1), 193–207.

Kharoufeh, J.P., Solo, J.C. and Ulukus, M.Y. (2010) Semi-Markov models for degradation-based reliability. *IIE Transactions*, **42**(8), 599–612.

Kiessler, P.C., Klutke, G.A. and Yoonjung, Y. (2002) Availability of periodically inspected systems subject to Markovian degradation. *Journal of Applied Probability*, **39**(4), 700–711.

Kim, M.J., Makis, V. and Jiang, R. (2010) Parameter estimation in a condition-based maintenance model. *Statistics and Probability Letters*, **80**(21-22), 1633–1639.

Lam, C.T. and Yeh, R.H. (1994) Optimal replacement policies for multistate deteriorating systems. *Naval Research Logistics*, **41**(3), 303–315.

Liao, H., Elsayed, E.A. and Chan, L.-Y. (2006) Maintenance of continuously monitored degrading systems. *European Journal of Operational Research*, **175**(2), 821–835.

Limnios, N. (2012) Reliability measures of semi-Markov systems with general state space. *Methodology and Computing in Applied Probability*, **14**(4), 895–917.

Limnios, N. and Ouhbi, B. (2006) Nonparametric estimation of some important indicators in reliability for semi-Markov processes. *Statistical Methodology*, **3**(4), 341–350.

Lin, D. and Makis, V. (2004) On-line parameter estimation for a failure-prone system subject to condition monitoring. *Journal of Applied Probability*, **41**(1), 211–220.

Lisnianski, A. and Levitin, G. (2003) *Multi-State System Reliability, Assessment, Optimization and Applications*, World Scientific, Singapore.

Maswadah, M. (2003) Conditional confidence interval estimation for the inverse Weibull distribution based on censored generalized order statistics. *Journal of Statistical Computation and Simulation*, **73**(12), 888–898.

Moghaddass, R. and Zuo, M.J. (2011) A parameter estimation method for a multi-state deteriorating system with incomplete information, in *Proceedings of the Seventh International Conference on Mathematical Methods in Reliability: Theory, Methods, and Applications (MMR 2011)*, pp. 34–42.

Moghaddass, R. and Zuo, M.J. (2012) A parameter estimation method for a condition monitored device under multi-state deterioration. *Reliability Engineering and System Safety*, **106**, 94–103.

Moghaddass, R., Zuo, M.J. and Zhao, X. (2013) Modeling device with multi-state degradation with non-homogeneous continuous-time hidden semi-Markov process, in *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*, Kadry, S. (ed), IGI Global, pp. 151–181.

Moore, E.H. and Pyke, R. (1968) Estimation of the transition distributions of a Markov renewal process. *Annals of the Institute of Statistical Mathematics*, **20**(1), 411–424.

Morcous, G., Lounis, Z. and Mirza, M.S. (2003) Identification of environmental categories for Markovian deterioration models of bridge decks. *Journal of Bridge Engineering*, **8**(6), 353–361.

Moura, M.C. and Droguett, E.L. (2009) Mathematical formulation and numerical treatment based on transition frequency densities and quadrature methods for non-homogeneous semi-Markov processes. *Reliability Engineering and System Safety*, **94**(2), 342–349.

Ouhbi, B. and Limnios, N. (1999) Nonparametric estimation for semi-Markov processes based on its hazard rate functions. *Statistical Inference for Stochastic Processes*, **2**, 151–173.

Peng, Y. and Dong, M. (2011) A prognosis method using age-dependent hidden semi-Markov model for device health prediction. *Mechanical Systems and Signal Processing*, **25**(1), 237–252.

Sadek, A. and Limnios, N. (2005) Nonparametric estimation of reliability and survival function for continuous-time finite Markov processes. *Journal of Statistical Planning and Inference*, **133**(1), 1–21.

Saxena, A., Goebel, K., Simon, D. and Eklund, N. (2008) Damage propagation modeling for aircraft engine run-to-failure simulation, in *The International Conference on Prognostics and Health Management*, pp. 1–9.

Wayne, N. (1982) *Applied Life Data Analysis*, John Wiley & Sons, New York, NY.

Zhao, Z. (2008) Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys*, **2**, 1–42.

Appendices

Appendix A

This appendix provides the detailed steps in finding Equation (15) from Equation (12).

$$
\begin{aligned}
L_2'(K) &= \sum_{k=1}^{K} \log\left(\Pr\left(Q^{(k)}|\theta\right)\right) \\
&= \sum_{k=1}^{K} \log\left(\prod_{z=1}^{n_k} \Pr\left(X_z = X_z^{(k)},\, T_z - T_{z-1}\right.\right. \\
&\quad \left.\left. = d_z^{(k)} \middle| X_{z-1} = X_{z-1}^{(k)},\, T_{z-1} = T_{z-1}^{(k)}, \theta\right)\right) \\
&\quad + \sum_{k=1}^{K} \log\Big(\Pr\left(T_{n_k+1} - T_{n_k} > h_{n_k}^{(k)} \middle| X_{n_k}\right. \\
&\quad \left. = X_{n_k}^{(k)},\, T_{n_k} = T_{n_k}^{(k)}, \theta\right)\Big) = \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log(\Pr(X_{n+1} \\
&\quad = j,\, T_{n+1} - T_n = t | X_n = i,\, T_n = u, \theta)) + \sum_{(i,u,t)\in\Theta'^{(K)}} \\
&\quad \chi'_{i,u,t} \log(\Pr(T_{n+1} - T_n > t | X_n = i,\, T_n = u, \theta)) \\
&= \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log(\lambda_{i,j}(u,t) \times (1 - H_i(u,t))) \\
&\quad + \sum_{(i,u,t)\in\Theta'^{(K)}} \chi'_{i,u,t} \log(1 - H_i(u,t)) \\
&= \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \log(\lambda_{i,j}(u,t)) + \sum_{(i,j,u,t)\in\Theta^{(K)}} \chi_{i,j,u,t} \\
&\quad \times \left(-\int_0^t \sum_z \lambda_{i,z}(u,x)dx\right) + \sum_{(i,u,t)\in\Theta'^{(K)}} \chi'_{i,u,t} \\
&\quad \times \left(-\int_0^t \sum_z \lambda_{i,z}(u,x)\mathrm{d}x\right). \qquad \text{(A1)}
\end{aligned}
$$

Appendix B

This appendix provides the steps to calculate the Fisher matrix confidence interval for the estimated values of the parameters of the Weibull distribution ($\lambda_{i,j}(u,t) = \beta_{i,j}\alpha_{i,j}^{-\beta_{i,j}} t^{\beta_{i,j}-1}$) associated with the degradation transition rates between states in a multistate structure (Equation (17)). This approximate confidence interval can be found based on the following two properties: (i) asymptotic normality of maximum likelihood estimators and (ii) normality of $\ln\alpha$ and $\ln\beta$. Therefore, the $(1-\omega)\%$ confidence interval for the parameter of interest θ ($\theta \in (\alpha, \beta)$) can be found as

$$\left[\hat{\theta}\exp\left[\frac{-z_{\omega/2}\times\sqrt{\mathrm{var}(\hat{\theta})}}{\hat{\theta}}\right], \hat{\theta}\exp\left[\frac{z_{\omega/2}\times\sqrt{\mathrm{var}(\hat{\theta})}}{\hat{\theta}}\right]\right].$$

Now, the important step is to find the variance of the parameters using the inverse of the Fisher information matrix, which yields the variance–covariance matrix as

$$F_{i,j} = \begin{bmatrix} -\dfrac{\partial^2 L_2'(K)}{\partial\alpha_{i,j}^2} & -\dfrac{\partial^2 L_2'(K)}{\partial\alpha_{i,j}\beta_{i,j}} \\ -\dfrac{\partial^2 L_2'(K)}{\partial\beta_{i,j}\alpha_{i,j}} & -\dfrac{\partial^2 L_2'(K)}{\partial\beta_{i,j}^2} \end{bmatrix} \to \begin{bmatrix} \mathrm{var}(\hat{\alpha}) & \mathrm{cov}(\hat{\alpha},\hat{\beta}) \\ \mathrm{cov}(\hat{\beta},\hat{\alpha}) & \mathrm{var}(\hat{\beta}) \end{bmatrix}$$
$$= \begin{bmatrix} -\dfrac{\partial^2 L_2'(K)}{\partial\alpha_{i,j}^2} & -\dfrac{\partial^2 L_2'(K)}{\partial\alpha_{i,j}\beta_{i,j}} \\ -\dfrac{\partial^2 L_2'(K)}{\partial\beta_{i,j}\alpha_{i,j}} & -\dfrac{\partial^2 L_2'(K)}{\partial\beta_{i,j}^2} \end{bmatrix}^{-1}, \tag{A2}$$

$$\begin{cases} \dfrac{\partial^2 L_2'(K)}{\partial\alpha_{i,j}^2} = \left(\sum\limits_u\sum\limits_d \chi_{i,j,u,d}\times\beta\times\alpha^{-2}\right) \\ \qquad - \sum\limits_b\sum\limits_u\sum\limits_d \chi_{i,b,u,d}\times\hat{\beta}_{i,j}\times(1+\hat{\beta}_{i,j}) \\ \qquad \times d^{\hat{\beta}_{i,j}}\times\alpha^{-\hat{\beta}_{i,j}-2} \\ \dfrac{\partial^2 L_2'(K)}{\partial\alpha_{i,j}\beta_{i,j}} = \left(-\sum\limits_u\sum\limits_d \chi_{i,j,u,d}\times\alpha^{-1}\right) \\ \qquad + \sum\limits_b\sum\limits_u\sum\limits_d \chi_{i,b,u,d}\times d^{\hat{\beta}_{i,j}}(\hat{\beta}_{i,j}\times\alpha^{-\hat{\beta}_{i,j}-1} \\ \qquad \times(\ln t - \ln\alpha_{i,j}) - \alpha^{-\hat{\beta}_{i,j}-1}) \\ \dfrac{\partial^2 L_2'(K)}{\partial\beta_{i,j}^2} = \left(-\sum\limits_u\sum\limits_d \chi_{i,j,u,d}\times\beta^{-2}\right) \\ \qquad - \sum\limits_b\sum\limits_u\sum\limits_d \chi_{i,b,u,d}\times\hat{\beta}_{i,j}\times\ln^2(t/\alpha_{i,j}) \\ \qquad \times(t/\alpha_{i,j})^{\beta_{i,j}} \\ \dfrac{\partial^2 L_2'(K)}{\partial\beta_{i,j}\alpha_{i,j}} = \left(-\sum\limits_u\sum\limits_d \chi_{i,j,u,d}\times\alpha^{-1}\right) \\ \qquad + \sum\limits_b\sum\limits_u\sum\limits_d \chi_{i,b,u,d}\times\alpha^{-1}\times((t/\alpha_{i,j})^{\beta_{i,j}} \\ \qquad + \hat{\beta}_{i,j}\times(t/\alpha_{i,j})^{\beta_{i,j}}\times\ln(t/\alpha_{i,j})) \end{cases} \tag{A3}$$

It should be pointed out that when the transition from state i to state j depends on the total age of the device (instead of the sojourn time at each state)—that is, $\lambda_{i,j}(u,t) = \beta_{i,j}\alpha_{i,j}^{-\beta_{i,j}}(u+t)^{\beta_{i,j}-1}$—the results shown in Equation (A3) need to be modified accordingly (t should be replaced by $(u+t)$ for the corresponding transitions).

Appendix C

This appendix provides the detailed steps in finding Equation (29) from Equation (25). We first use approximate formulae to calculate $\int_0^\infty G_{i,j}'(u,x,K)\mathrm{d}x$ as $\int_0^\infty G_{i,j}'(u,x,K)\mathrm{d}x = \sum_x G_{i,j}'^{\mathrm{p}}(u,x,K)\times\delta + \varepsilon_{i,j}^2(u)$. Similar to $\varepsilon_{i,j}^1(u,t)$, it is assumed that as $K\to\infty$, $\varepsilon_2/L_{2,2}'^{\mathrm{p}}(G_{i,j}'^{\mathrm{p}}(u,t,K),K)$ does not add any significant information to the likelihood function. In order to estimate $G_{i,j}'^{\mathrm{p}}(u,t,K)$, we can rewrite Equation (28) and express it in terms of $G_{i,j}'^{\mathrm{p}}(u,t,K)$ as

$$\begin{aligned} &L_{2,2}'^{\mathrm{p}}\left(G_{i,j}'^{\mathrm{p}}(u,t,K),K\right) \\ &\quad = \chi_{i,j,u,t}\log(G_{i,j}'(u,t,K)) \\ &\qquad + \vartheta_3\left(1-\sum_x G_{i,j}'^{\mathrm{p}}(u,x,K)\times\delta + \varepsilon_{i,j}^2(u)\right). \end{aligned} \tag{A4}$$

Now, Equation (A4) can be maximized with respect to $G_{i,j}'^{\mathrm{p}}(u,t,K)$ as

$$\begin{cases} \dfrac{\partial L_{2,2}'^{\mathrm{p}}\left(G_{i,j}'^{\mathrm{p}}(u,t,K),K\right)}{\partial G_{i,j}'^{\mathrm{p}}(u,t,K)} = 0 \\ \qquad \to \dfrac{\chi_{i,j,u,t}}{G_{i,j}'^{p}(u,t,K)} - \vartheta_3\times\delta = 0 \\ \dfrac{\partial L_{2,2}'^{\mathrm{p}}\left(G_{i,j}'^{\mathrm{p}}(u,t,K),K\right)}{\partial\vartheta_3} = 0 \\ \qquad \to \left(1-\sum\limits_x G_{i,j}'^{\mathrm{p}}(u,x,K)\times\delta\right) = 0 \\ \qquad\qquad \to \widehat{G}_{i,j}'^{\mathrm{p}}(u,t,K) = \chi_{i,j,u,t} \\ \qquad \times\left(\delta\times\sum\limits_d \chi_{i,j,u,d}\right)^{-1}, \quad (i,j,u,t)\in\Theta^{(K)}. \end{cases} \tag{A5}$$

Considering that $G_{i,j}^{\mathrm{p}}(u,t,K) = \int_o^t G_{i,j}'^{\mathrm{p}}(u,x,K)\mathrm{d}x$, $\hat{G}_{i,j}^{\mathrm{p}}(u,t,K)$ can be estimated as

$$\begin{aligned} &G_{i,j}^{\mathrm{p}}(u,t,K) \\ &\quad = \int_o^t G_{i,j}'^{\mathrm{p}}(u,x,K)\mathrm{d}x \cong \sum_{x\le t} G_{i,j}'^{\mathrm{p}}(u,x,K)\times\delta \\ &\to \hat{G}_{i,j}^{\mathrm{p}}(u,t,K) \\ &\quad = \left(\sum_{d\le t}\chi_{i,j,u,d}\right)\left(\sum_d \chi_{i,j,u,d}\right)^{-1}, \quad (i,j,u,t)\in\Theta^{(K)}. \end{aligned} \tag{A6}$$

Biographies

Ramin Moghaddass is currently a Ph.D. candidate at the Department of Mechanical Engineering at the University of Alberta. He received both his B.Sc. and M.Sc. from Sharif University of Technology, Tehran, Iran. His research activities are in multistate reliability and degradation modeling, condition monitoring, and condition-based maintenance.

Ming J. Zuo received an M.Sc. in 1986 and a Ph.D. in 1989, both in Industrial Engineering, from Iowa State University, Ames, Iowa. He is currently a Professor in the Department of Mechanical Engineering at the University of Alberta. His research interests include system reliability analysis, maintenance planning and optimization, condition monitoring, and fault diagnosis. He is Associate Editor of *IEEE Transactions on Reliability* and Department Editor of *IIE Transactions*. He is a Fellow of IIE, EIC, and ISEAM and Senior Member of IEEE.