An Ameliorated Improvement Factor Model for Imperfect Maintenance and Its Goodness of Fit

Mimi Zhang[1] and Min Xie[2]

[1]Department of Management Science, University of Strathclyde, Glasgow, G1 1XQ, UK

[2]Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong, China

ABSTRACT

Maintenance actions can be classified, according to their efficiency, into three categories: perfect maintenance, imperfect maintenance and minimal maintenance. To date, the literature on imperfect maintenance is voluminous, and many models have been developed to treat imperfect maintenance. Yet, there are two important problems in the community of maintenance that still remain wide open: how to give practical grounds for an imperfect-maintenance model, and how to test the fit of a real data set to an imperfect-maintenance model. Motivated by these two pending problems, this work develops an imperfect-maintenance model by taking a physically meaningful approach. For the practical implementation of the developed model, we advance two methods, called QMI method and spacing-likelihood algorithm, to estimate involved unknown parameters. The two methods complete each other and are widely applicable. To offer a practical guide for testing fit to an imperfect-maintenance model, this work promotes a bootstrapping approach to approximating the distribution of a test statistic. The attractions and dilemmas of QMI method and spacing-likelihood algorithm are revealed via simulated data. The utility of the developed imperfect-maintenance model is evidenced via a real data set. This article has a supplementary material online.

KEY WORDS: Censored data; Maximum spacing estimation; Model selection; Parametric bootstrap; Quasi-Monte Carlo integration.

1 INTRODUCTION

Maintenance (or, interchangeably, repair) plays a key role in the life-cycle management of a machine. A machine herein could be any piece of mechanical equipment. An optimal scheduling of inspection, maintenance and replacement can improve the reliability of a machine and, meanwhile, significantly reduce operational cost. Hence, the modeling of maintenance efficiency and the optimization of maintenance policies are attracting more and more attention; see Ahmad and Kamaruddin (2012), Wang (2012), Shafiee and Chukova (2013) and Tanwar et al. (2014) for recent reviewing work on maintenance. Maintaining actions comprise perfect repairs, imperfect repairs and minimal repairs performed to retain or restore a machine in or to an acceptable operating condition. The problems of perfect maintenance and minimal maintenance are trivial, which can be readily solved by invoking a renewal process and a non-homogeneous Poisson process (Lindqvist, 2006). Compared with the assumptions of perfect maintenance and minimal maintenance, which only have restricted applications, the assumption of imperfect maintenance is more practical and has been applied to a variety of machines, such as electric drills (Shafiee et al., 2012), a distributed energy system (Xia et al., 2012), motors (Pulcini, 2013), a coal-transportation system (Pandey et al., 2013), fire water pumps (Ramirez and Utne, 2013), a high pressure die casting machine (Tambe et al., 2013), dump trucks (de Toledo et al., 2015), machines in power plants (Dijoux and Idee, 2013; Melchor-Hernandez et al., 2015), et al..

Motivated by the practical need of modelling imperfect repairs, numerous imperfect-maintenance models have been developed. There are many imperfect-maintenance models that are built on stochastic processes and multi-state systems; for recent works, the reader is referred to Guo et al. (2013), Mercier and Castro (2013), Zhang et al. (2015), Pandey et al. (2013), Liu et al. (2013), Zhang et al. (2013), and the references therein. Here, we only review imperfect-maintenance models that are built on lifetime distributions. An early attempt at modeling imperfect maintenance is the (p, q) rule, assuming that the probability of a repair being perfect is p and being minimal is q=1-p (Nakagawa, 1979). Block et al. (1985) extended the (p, q) rule to be age-dependent, namely, the (p(t), q(t)) rule. Obviously, the (p, q) rule (or its generalization) is confined to the realm of speculation, and will be invalid in most cases where repairs are neither perfect nor minimal. Kijima (1989) introduced the concept of virtual age, assuming that an imperfect repair reduces a maintained machine's physical age (therein called virtual age) by an amount proportional to the additional age accumulated since the last repair, or by an amount proportional to the physical age just before the repair. Henceforth, virtual age models have been extensively explored and em-

ployed (see, e.g., Dijoux and Idee, 2013; Khojandi et al., 2014; Tanwar et al., 2014). Malik (1979) introduced the concept of improvement factor, assuming that an imperfect repair changes the time of the hazard rate curve to some newer time but not all the way to zero. Lin et al. (2000) extended the improvement factor model, stating that an imperfect repair could change both the time and the slope of the hazard rate curve. For recent research on the (extended) improvement factor model, the reader may refer to, e.g., Park et al. (2012), Xia et al. (2012) and Khatab et al. (2014). Note that the extended improvement factor model by Lin et al. (2000) indeed embraces many documented imperfect-maintenance models, including the family of virtual age models, which act through the hazard rate function. Since Malik (1979) put the idea out before Kijima (1989), in the current work we group virtual age models into the family of improvement factor models. One major limitation of improvement factor models is they all assuming that the improvement factor is identical for all repairs during the maintained machine's whole operational life cycle. Intuitively, each repair should have a different degree of impact on the machine's quality. Hence, it is more practical to assume that the improvement factor is random. Motivated by the idea of "improvement factor" and the limitations of existing improvement factor models, the major contribution of the current work is an imperfect-maintenance model grounded on practical considerations. Our approach is legislated by the notion of hazard potential introduced by Singpurwalla (2006).

The second contribution of the current work is a generic approach for testing whether an imperfect-maintenance model is valid in light of a real data set. For every particular application and data set, practitioners are confronted with the problem of selecting among a set of candidate models the most appropriate model. An intuitive approach to the problem is to build a statistical test of hypotheses. Lifetime data of identical machines under perfect maintenance and/or minimal maintenance are identically distributed, and therefore classical test statistics are applicable (see, e.g., Li et al., 2014). By contrast, testing fit to an imperfect-maintenance model is a big challenge in that lifetime data of identical machines under imperfect maintenance are no longer identically distributed. Furthermore, for certain imperfect-maintenance models, the distribution function of a test statistic not only depends on the number of lifetime data but also on the number of imperfect repairs. Therefore, classical test statistics are not directly applicable. Testing fit to an imperfect-maintenance model has received surprisingly little attention. To our knowledge, at the present time there exist only two works on goodness-of-fit test for imperfect-maintenance models: Liu et al. (2012) and de Toledo et al. (2015). However, the method developed by Liu et al. (2012) is questionable, because transformed data (by using probability integral transformation) are not uniformly distributed when involved parameters are unknown and estimated (David and Johnson,

1948). The methods developed by de Toledo et al. (2015) are for model selection (namely, selecting among a given set of models the best fitted model) not for testing model validation in light of data. Motivated by the practical need of testing goodness of fit for imperfect-maintenance models, we here develop a generic approach such that practitioners can readily compute the *p*-value for any particular problem.

The following exposition is divided into four parts. Section 2 is devoted to mathematically formulating our two contributions. Section 3 is devoted to parameter-estimation. In Section 4, we further detail the procedure for testing goodness-of-fit. In Section 5, we present numerical study results.

2 MODEL FORMULATION

2.1 Ameliorated Improvement Factor Model

Let random variable T denote the lifetime of a machine operating in certain environment (denoted by $\mathscr{E}$). The reliability function of T, namely $\Pr(T > t; t \geq 0)$, is assumed to be absolutely continuous. In the foregoing and following expressions, we introduce the convention that all quantities to the right of the vertical slash are conditional, and that all quantities to the right of the semicolon are known (specified). Let non-negative function $h(t)$, $t \geq 0$, denote the hazard rate function of $\Pr(T > t; t \geq 0)$. Suppose that, after being put into operation at time 0, the machine is repaired at time s (> 0). A general framework of the improvement factor model is set forth as follows (see, e.g., Lin et al., 2000). The repair is assumed to take negligible time. Right after the repair the *hazard rate function* is re-formulated to be $bh(t - s + as)$ for $t > s$. Here, the value $t - s + as$ in the set of parentheses is called the physical age (or, interchangeably, virtual age) of the machine. $t - s$ represents the time elapsed from the repair. a $(\in [0,1])$ is an age-reduction factor, indicating that the physical age of the machine may be reduced (for $0 \leq a < 1$) or the same (for $a = 1$). $b(> 0)$ is a hazard-rate-alteration factor, indicating that the rate of accumulation of hazard may be reduced (for $0 < b < 1$), the same (for $b = 1$), or advanced (for $b > 1$). The improvement factor model is widely applicable: if $b = 1$, the improvement factor model will reduce to a virtual age model; if the repair is perfect, we can set $a = 0$ and $b = 1$; if the repair is minimal, we can set $a = 1$ and $b = 1$.

Under the improvement factor model, an imperfect repair should lower the hazard rate of the machine and lead to $bh(as) < h(s)$. One shortcoming of the improvement factor model is it assuming that a and b are constant. This assumption could be inappropriate and even problematical in

many circumstances. For example, the improvement factor model cannot guarantee that the hazard rate will always be lowered because, if $b > 1$, $bh(as)$ can be larger than $h(s)$. Since the environment $\mathscr{E}$ may change wildly and every repair should have a different degree of impact on the machine's quality, it is more practical to assume that a and b are random. By quality, we mean a resistance to failure-causing agents. The second shortcoming of the improvement factor model is its implicit requirement that the hazard rate function should be non-decreasing. If the hazard rate function is non-increasing (e.g., the first part of a bath-tub shaped hazard rate function), then $bh(as)$ can be larger than $h(s)$. Consequently, this implicit requirement impedes the wide implementation of the improvement factor model in real-life practice. The third shortcoming of the improvement factor model is that right after the repair the reliability of the machine is still $\exp(-\int_0^s h(u)du)$. However, it is more practical in reality that an imperfect repair will improve the reliability of the maintained machine.

Let $H(t)$ denote the cumulative hazard of the machine at time t (≥ 0):

$$H(t) = \int\limits_0^t h(u)du = \int\limits_0^t \Pr(T \leq u + du | T > u).$$

Here and in the following, the symbol "d" denotes an infinitesimal change in a quantity. We note that the cumulative hazard function $H(t)$ is monotonically increasing. Motivated by the above shortcomings of the improvement factor model, we here ameliorate it by replacing the hazard rate function, $h(t)$, with the cumulative hazard function, $H(t)$. Specifically, the ameliorated improvement factor model says that right after the repair the *cumulative hazard function* is re-formulated to be $bH(t - s + as)$ for $t > s$. We might still call a ($\in [0,1]$) the age-reduction factor and $b(> 0)$ the hazard-rate-alteration factor. If the repair is perfect, we set $a = 0$ and $b = 1$; if the repair is minimal, we set $a = 1$ and $b = 1$. For the ameliorated improvement factor model, the improvement factors a and b are random, having certain probability distributions. If the variance of a (or b) is small, then we can approximate it by a constant value. In Section 1 of the supplementary material, we discuss in detail the advantages of the ameliorated improvement factor model. In what follows, unless otherwise stated, by the improvement factor model we mean the ameliorated improvement factor model.

It is well known that the reliability function and the cumulative hazard function are related via the exponentiation formula $\Pr(T > t; H(t)) = \exp(-H(t))$. The improvement factor model is not a product of conjecture. We provide in the following a novel platform for interpreting the improvement factor model by giving a new insight into the exponentiation formula. For ease of exposition, an exponential distribution with its rate parameter being 1 is called the standard exponential dis-

tribution. Let X be a random variable following the standard exponential distribution. Note that $\exp(-H(t))$ is indeed the reliability function of X evaluated at $H(t)$. Namely, we have

$$\Pr(T > t; H(t)) = \Pr(X > H(t)), \tag{1}$$

or

$$\Pr(T \leq t; H(t)) = \Pr(H(t) \geq X). \tag{2}$$

Equation (2) says that the time to failure of the machine coincides with the time at which its cumulative hazard exceeds a random threshold; the random threshold follows the standard exponential distribution. Maintenance on the machine, if any, will change its lifetime distribution, i.e., the left part of Equation (2). Note that, for any and all machines, the distribution of the random threshold is the standard exponential distribution, irrespective of the machine's quality and the environment $\mathscr{E}$. X is called by Singpurwalla (2006) the hazard potential, of which the distribution cannot be changed by, say, maintenance actions. Therefore, we claim that maintenance on the machine, if any, will re-formulate the right part of Equation (2), yet by virtue of directly re-formulating its cumulative hazard function. There can be a few ways to re-formulate its cumulative hazard function – the improvement factor model is certainly a practical and competent one.

2.2 Cumulative-Hazard-Scaling Method

Cumulative-hazard-scaling method is the name given to a generic procedure for testing fit to the improvement factor model. For testing fit to other imperfect-maintenance models, the appropriate translations are obvious. Suppose that $n(\geq 1)$ independent machines are put into operation at time 0, the lifetimes of which are denoted by random variables $\{T_1, \cdots, T_n\}$. The lifetimes of the n machines need not be identically distributed. After being put into operation, the machines will be repaired at certain time points. For $1 \leq i \leq n$, let non-negative function $H^i(t; \boldsymbol{\theta}^i)$, $t \geq 0$, denote the cumulative hazard function of the ith machine before performing any repair on it. $\boldsymbol{\theta}^i$ is a vector of parameters. Assume that the ith machine fails at time $t_i(> 0)$ (see Figure 1). Before its failure,
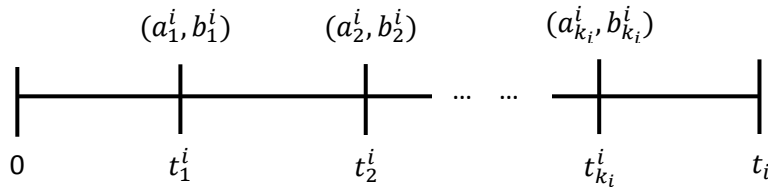
Figure 1: Repairs are performed on the ith machine at time points t^i_j's for $1 \leq j \leq k_i$.

$k_i (\geq 1)$ repairs have been performed on the machine at time points $\left\{ t_1^i, \cdots, t_{k_i}^i \right\}$ with $0 = t_0^i < t_1^i < \cdots < t_{k_i}^i < t_i$. For $1 \leq j \leq k_i$, let a_j^i (resp. b_j^i) denote the outcome of the random variable a (resp. b) at the maintenance-performing time t_j^i. Let non-negative function $\delta^i(t; \cdot)$ denote, at time t (≥ 0), the virtual age of the ith machine. It's easy to derive that, at time t_j^i right after the jth maintenance, the virtual age of the ith machine is $\delta^i\left(t_j^i; a_1^i, \cdots, a_j^i\right) = \sum_{m=1}^{j}(\Pi_{v=m}^j a_v^i)(t_m^i - t_{m-1}^i)$. Therefore, at time t_i, the virtual age of the ith machine is

$$\delta^i\left(t_i; a_1^i, \cdots, a_{k_i}^i\right) = t_i - t_{k_i}^i + \sum_{m=1}^{k_i}\left(\Pi_{v=m}^{k_i}a_v^i\right)\left(t_m^i - t_{m-1}^i\right),$$

and the cumulative hazard of the ith machine is

$$H_i = \Pi_{v=1}^{k_i}b_v^i H^i\left(\delta^i\left(t_i; a_1^i, \cdots, a_{k_i}^i\right); \boldsymbol{\theta}^i\right). \tag{3}$$

Now turn back to Equation (1). We note that for every random variable T having the cumulative hazard function $H(t)$, there exists a random variable X following the standard exponential distribution. The reliability function of T is indexed by t, whereas the reliability function of X is indexed by $H(t)$. Singpurwalla (2006, Theorem 1) claimed that the lifetime of any and all machines follows the standard exponential distribution with $H(t)$ being the scale. Consequently, we can state the following result.

Proposition 1. *If the improvement factor model is the true underlying law, and the involved parameters (e.g., the a_j^i's) are all known, then the data $\{H_1, \cdots, H_n\}$ follow the standard exponential distribution.*

Remark 1. *The conditional lifetime distribution of the ith machine after the k_ith maintenance can be formulated to be*

$$\Pr(T_i \leq t \mid T_i > t_{k_i}^i) = 1 - \exp\left(-\Pi_{v=1}^{k_i}b_v^i H^i\left(\delta^i\left(t; a_1^i, \cdots, a_{k_i}^i\right); \boldsymbol{\theta}^i\right)\right), \qquad t > t_{k_i}^i.$$

Therefore, the lifetime data $\{t_1, \cdots, t_n\}$ are independent but not identically distributed.

Remark 2. *We should pay attention to the condition stated in Proposition 1: parameters involved in the calculation of the H_i's should be completely known. However, in real-life practice, we have to estimate, say, the a_j^i's in order to obtain the H_i's. Consequently, according to David and Johnson (1948), the H_i's are no longer independent and may not follow the standard exponential distribution. The distribution of the H_i's can only be approximated by Monte Carlo simulation. Therefore, for testing fit to the improvement factor model, we cannot simply examine whether the H_i's are from the standard exponential distribution.*

Now we can bring up the cumulative-hazard-scaling method which involves the following steps.

Step 1 Estimate unknown parameters according to the data $\{t_1, \cdots, t_n\}$, $\{k_1, \cdots, k_n\}$ and the t_j^i's.

Step 2 For $1 \leq i \leq n$, calculate H_i by using Equation (3) in which $\boldsymbol{\theta}^i$, a_j^i and b_j^i ($1 \leq j \leq k_i$) are replaced by their estimated values.

Step 3 Employ suitable goodness-of-fit tests on the data $\{H_1, \cdots, H_n\}$ to examine whether the improvement factor model is the underlying law.

The key steps of the cumulative-hazard-scaling method are Step 1 (for parameter estimation) and Step 3 (for goodness-of-fit test). Therefore, the following two sections, Section 3 and Section 4, are devoted to parameter estimation and goodness-of-fit test, respectively.

Remark 3. *As of now, we assume that the data $\{t_1, \cdots, t_n\}$ are complete lifetime data. It should be pointed out that the cumulative-hazard-scaling method can be directly applied to incomplete life-time data. If, for example, the lifetime of the first machine is right censored at time o_1 ($t_{k_1}^1 < o_1 < t_1$), then the corresponding value, $H_1 = \Pi_{v=1}^{k_1} b_v^1 H^1(\delta^1(o_1; a_1^1, \cdots, a_{k_1}^1); \boldsymbol{\theta}^1)$, is a right censored obser-vation. Parameter-estimation methods and goodness-of-fit tests for handling incomplete lifetime data are respectively developed in Section 2 of the supplementary material and Section 4.*

3   PARAMETER ESTIMATION

For illustrative purpose, we here make the following assumptions.

A1 The age-reduction factor, a, follows a beta distribution with the probability density function given by

$$f_a(x;\ u,v) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)}x^{u-1}(1-x)^{v-1}, \qquad 0 < x < 1.$$

$u(>0)$ and $v(>0)$ are two unknown parameters. Let $F_a(x;\ u,v)$ denote the cumulative distribution function of the age-reduction factor.

A2 The hazard-rate-alteration factor, b, follows a gamma distribution with the probability density function given by

$$f_b(x;\ e,c) = \frac{c^e}{\Gamma(e)}x^{e-1}\exp(-cx), \qquad x > 0.$$

$e(>0)$ and $c(>0)$ are two unknown parameters. Let $F_b(x;\ e,c)$ denote the cumulative distribution function of the hazard-rate-alteration factor.

8

A3 The *n* independent machines are statistically identical in that, before performing any repair, the cumulative hazard function (resp. hazard rate function) of each of the *n* machines is $H(t; \boldsymbol{\theta})$ (resp. $h(t; \boldsymbol{\theta})$) for $t \geq 0$. $\boldsymbol{\theta}$ is a vector of unknown parameters. Let $F(t; \boldsymbol{\theta})$ denote the matching cumulative distribution function.

For $1 \leq i \leq n$, the conditional probability of the event $\{t_i \leq T_i \leq t_i + dt_i\}$, conditioning on the data $\{a_1^i, \cdots, a_{k_i}^i\}$ and $\{b_1^i, \cdots, b_{k_i}^i\}$, can be calculated via

$$
\Pr\left(t_i \leq T_i \leq t_i + dt_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{b_1^i, \cdots, b_{k_i}^i\}\right)
$$
$$
= \Pi_{\nu=1}^{k_i} b_\nu^i h(\delta^i(t_i; a_1^i, \cdots, a_{k_i}^i); \boldsymbol{\theta}) \exp(-\Pi_{\nu=1}^{k_i} b_\nu^i H(\delta^i(t_i; a_1^i, \cdots, a_{k_i}^i); \boldsymbol{\theta})) dt_i.
$$

For $1 \leq j \leq k_i$, the probability $\Pr(a_j^i \leq a \leq a_j^i + da_j^i | u, v)$ can be calculated via

$$
\Pr(a_j^i \leq a \leq a_j^i + da_j^i | u, v) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} (a_j^i)^{u-1} (1 - a_j^i)^{v-1} da_j^i,
$$

and the probability $\Pr(b_j^i \leq b \leq b_j^i + db_j^i | e, c)$ can be calculated via

$$
\Pr(b_j^i \leq b \leq b_j^i + db_j^i | e, c) = \frac{c^e}{\Gamma(e)} (b_j^i)^{e-1} \exp(-cb_j^i) db_j^i.
$$

Therefore, if the data $\{a_1^i, \cdots, a_{k_i}^i\}$ and $\{b_1^i, \cdots, b_{k_i}^i\}$ are available, then the log-likelihood function of $\Theta = (\boldsymbol{\theta}, u, v, e, c)$ can be formulated to be

$$
\ell(\Theta; \{t_1, \cdots, t_n\}, A, B) = \sum_{i=1}^n \log\left(\frac{1}{dt_i} \Pr\left(t_i \leq T_i \leq t_i + dt_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{b_1^i, \cdots, b_{k_i}^i\}\right)\right)
$$
$$
+ \sum_{i=1}^n \sum_{j=1}^{k_i} \log(f_a(a_j^i; u, v)) + \sum_{i=1}^n \sum_{j=1}^{k_i} \log(f_b(b_j^i; e, c)),
$$

in which A is a set collecting all the a_j^i's, and B is a set collecting all the b_j^i's.

3.1 QMI Method

Evidently, $\{a_1^i, \cdots, a_{k_i}^i\}$ and $\{b_1^i, \cdots, b_{k_i}^i\}$ are unobservable latent data. Therefore, we need to calculate the marginal probability of the event $\{t_i \leq T_i \leq t_i + dt_i\}$:

$$
\Pr\left(t_i \leq T_i \leq t_i + dt_i | \Theta\right) = \underbrace{\int_0^\infty \cdots \int_0^\infty}_{k_i} \underbrace{\int_0^1 \cdots \int_0^1}_{k_i} \Pr\left(t_i \leq T_i \leq t_i + dt_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{b_1^i, \cdots, b_{k_i}^i\}\right)
$$
$$
\prod_{j=1}^{k_i} f_a(a_j^i; u, v) \prod_{j=1}^{k_i} f_b(b_j^i; e, c) da_1^i \cdots da_{k_i}^i db_1^i \cdots db_{k_i}^i.
$$

The marginal log-likelihood function of Θ therefore can be formulated to be

$$\mathfrak{L}(\Theta; \{t_1, \cdots, t_n\}) = \sum_{i=1}^{n} \log \left(\frac{1}{dt_i} \Pr(t_i \leq T_i \leq t_i + dt_i | \Theta) \right).$$

The ML estimate of Θ, denoted by $\hat{\Theta}$, is the maximizer of the marginal log-likelihood function:

$$\hat{\Theta} = \arg\max_{\Theta} \mathfrak{L}(\Theta; \{t_1, \cdots, t_n\}).$$

Normally, the marginal probability $\Pr(t_i \leq T_i \leq t_i + dt_i | \Theta)$ shall involve high-dimensional integration and therefore is very knotty to evaluate. Two available approaches to tackling high-dimensional integration are numerical integration and Monte Carlo integration. One drawback of numerical integration is that it is inappropriate for evaluating the marginal probability when $k_i > 2$, due to the curse of dimensionality. Monte Carlo integration, on the contrary, is particularly useful for evaluating high-dimensional integrals. The strong law of large numbers guarantees that the Monte Carlo approximation converges almost surely. Moreover, it follows from the central limit theorem that the order of the expected integration error in terms of the Monte Carlo sample size is $O(N^{-1/2})$ (Gentle, 2003, Chapter 7).

The main disadvantage of Monte Carlo integration is that it is slow. The sample size N should be fairly large – on the order of thousands or even millions – to obtain acceptable precision in the answer. On the other hand, we note that optimization algorithms for maximizing $\mathfrak{L}(\Theta; \{t_1, \cdots, t_n\})$ shall require thousands of times of integration. Hence, the computational load for ML estimation of Θ is large. To reduce computational load, we here replace Monte Carlo integration with quasi-Monte Carlo integration (Morokoff and Caflisch, 1995). It has been proved that lower error and improved convergence can be obtained by replacing random points with more uniformly distributed (known as quasi-random) points; see Niederreiter (1978). Computational experiments have shown that quasi-Monte Carlo integration is significantly more efficient than Monte Carlo integration; see, among others, Papageorgiou (2001), Schlier (2004), and Sloan and Wozniakowski (1998).

By defining $\varepsilon_j^i = \exp(-b_j^i)$ for $1 \leq i \leq n$ and $1 \leq j \leq k_i$, the marginal probability can be equivalently formulated to be

$$\Pr(t_i \leq T_i \leq t_i + dt_i | \Theta) = \int_{U_i} \Pr\left(t_i \leq T_i \leq t_i + dt_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{-\log(\varepsilon_1^i), \cdots, -\log(\varepsilon_{k_i}^i)\}\right)$$
$$\Pi_{j=1}^{k_i} f_a(a_j^i; u, v) \Pi_{j=1}^{k_i} [(\varepsilon_j^i)^{-1} f_b(-\log(\varepsilon_j^i); e, c)] da_1^i \cdots da_{k_i}^i d\varepsilon_1^i \cdots d\varepsilon_{k_i}^i,$$

in which $\boldsymbol{U}_i$ is the $2k_i$-dimensional unit cube: $\boldsymbol{U}_i = \underbrace{(0,1) \times \cdots \times (0,1)}_{2k_i}$. Let non-negative function

$g(\boldsymbol{z}^i; \Theta)$ denote the integrant:

$$g(\boldsymbol{z}^i; \Theta) = \Pr\left(t_i \le T_i \le t_i + dt_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{-\log(\varepsilon_1^i), \cdots, -\log(\varepsilon_{k_i}^i)\}\right)$$
$$\Pi_{j=1}^{k_i} f_a(a_j^i; u, v) \Pi_{j=1}^{k_i}[(\varepsilon_j^i)^{-1} f_b(-\log(\varepsilon_j^i); e, c)],$$

in which $\boldsymbol{z}^i = (a_1^i, \cdots, a_{k_i}^i, \varepsilon_1^i, \cdots, \varepsilon_{k_i}^i)$. The idea of quasi-Monte Carlo integration is to generate a sequence of N quasi-random points, denoted by $\boldsymbol{z}_1^i, \cdots, \boldsymbol{z}_N^i$, from the unit cube $\boldsymbol{U}_i$. Then the quasi-Monte Carlo approximation for the marginal probability is

$$\Pr\left(t_i \le T_i \le t_i + dt_i | \Theta\right) = \int_{\boldsymbol{U}_i} g(\boldsymbol{z}^i; \Theta) d\boldsymbol{z}^i \approx \frac{1}{N} \sum_{v=1}^{N} g(\boldsymbol{z}_v^i; \Theta).$$

For later reference, we might call the foregoing ML method by utilizing quasi-Monte Carlo integration "QMI method". R code for QMI method is available on request.

Remark 4. *It should be pointed out that $\{\varepsilon_1^i, \cdots, \varepsilon_{k_i}^i\}$ is an exchangeable sequence of random variables in that for any permutation of $\{\varepsilon_1^i, \cdots, \varepsilon_{k_i}^i\}$, denoted by $\{\ddot{\varepsilon}_1^i, \cdots, \ddot{\varepsilon}_{k_i}^i\}$, the integrant of the permutated sequence $g((a_1^i, \cdots, a_{k_i}^i, \ddot{\varepsilon}_1^i, \cdots, \ddot{\varepsilon}_{k_i}^i); \Theta)$ is identical to the integrant of the original sequence $g((a_1^i, \cdots, a_{k_i}^i, \varepsilon_1^i, \cdots, \varepsilon_{k_i}^i); \Theta)$. Therefore, the set of the N points $\{\boldsymbol{z}_1^i, \cdots, \boldsymbol{z}_N^i\}$ is in effect a set of the factorial of k_i times N points.*

3.2 Spacing-Likelihood Algorithm

QMI method has its inherent drawbacks: (1) the estimator $\hat{\Theta}$ may not be consistent, because QMI method maximizes the marginal log-likelihood function $\mathfrak{L}(\Theta; \{t_1, \cdots, t_n\})$ rather than the log-likelihood function $\ell(\Theta; \{t_1, \cdots, t_n\}, A, B)$; (2) QMI method is still time-consuming. Intuitively, a method directly maximizing the log-likelihood function $\ell(\Theta; \{t_1, \cdots, t_n\}, A, B)$ should be more efficient. Following the train of thought, now the crux of the matter is how to acquire the latent a_j^i's and b_j^i's.

Given the vector of unknown parameters, Θ, and the lifetime data $\{t_1, \cdots, t_n\}$, the log-likelihood function of A and B, denoted by $\mathbf{L}(A, B; \Theta, \{t_1, \cdots, t_n\})$, is proportional to

$$\mathbf{L}(A, B; \Theta, \{t_1, \cdots, t_n\}) \propto \sum_{i=1}^{n} \log\left(\frac{1}{dt_i} \Pr\left(t_i \le T_i \le t_i + dt_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{b_1^i, \cdots, b_{k_i}^i\}\right)\right)$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{k_i} \log(f_a(a_j^i; u, v)) + \sum_{i=1}^{n} \sum_{j=1}^{k_i} \log(f_b(b_j^i; e, c)). \quad (4)$$

Intuitively, if Θ is known, the ML estimates of the a_j^i's and b_j^i's are what jointly maximize the log-likelihood function $\mathbf{L}(A, B; \Theta, \{t_1, \cdots, t_n\})$. However, ML estimates of the b_j^i's will undoubtedly be closely spaced, because $\{b_1^i, \cdots, b_{k_i}^i\}$ is an exchangeable sequence. Even though $\{a_1^i, \cdots, a_{k_i}^i\}$ is not an exchangeable sequence, computational experiments showed that ML estimates of the a_j^i's are also closely spaced. Therefore, ML estimation (or any method grounded on maximizing the log-likelihood function $\mathbf{L}(A, B; \Theta, \{t_1, \cdots, t_n\})$) cannot be employed for estimating the a_j^i's and b_j^i's.

Note that if $\{b_1^i, \cdots, b_{k_i}^i\}$ are closely spaced, then $\{F_b(b_1^i; e, c), \cdots, F_b(b_{k_i}^i; e, c)\}$ are closely spaced. But if $\{b_1^i, \cdots, b_{k_i}^i\}$ are gamma-distributed, then $\{F_b(b_1^i; e, c), \cdots, F_b(b_{k_i}^i; e, c)\}$ should be uniformly distributed. Therefore, to guarantee that $\{b_1^i, \cdots, b_{k_i}^i\}$ are from a gamma distribution, we only need to guarantee that $\{F_b(b_1^i; e, c), \cdots, F_b(b_{k_i}^i; e, c)\}$ are from the standard uniform distribution. Following the train of thought, we come to the maximum spacing method (Cheng and Amin, 1983; Ranneby, 1984; Ekstrom, 1998). Maximum spacing estimation has been encountered in the statistics literature but is not widespread among engineers. Maximum spacing estimation was shown to give consistent, asymptotically efficient estimators even in cases where ML estimation fails (Anatolyev and Kosenok, 2005; Ekstrom, 1998, 2001).

For $1 \leq i \leq n$, let

$$F_i = \Pr(T_i \leq t_i | \boldsymbol{\theta}, \{a_1^i, \cdots, a_{k_i}^i\}, \{b_1^i, \cdots, b_{k_i}^i\}) = 1 - \exp(-\Pi_{v=1}^{k_i} b_v^i H(\delta^i(t_i; a_1^i, \cdots, a_{k_i}^i); \boldsymbol{\theta})).$$

We might assume that the data $\{F_1, \cdots, F_n\}$ have already been arranged in ascending order with $F_1 \leq F_2 \leq \cdots \leq F_n$. Likewise, arrange the elements in the set A (resp. B) in ascending order, denoted by $\{a_1, \cdots, a_k\}$ (resp. $\{b_1, \cdots, b_k\}$), in which $k = \sum_{i=1}^{n} k_i$. Let $F_0 = a_0 = b_0 = 0$, $F_{n+1} = a_{k+1} = 1$, and $b_{k+1} = +\infty$. Define a spacing function of A and B:

$$S(A, B; \Theta, \{t_1, \cdots, t_n\}) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log(F_i - F_{i-1}) + \frac{1}{k+1} \sum_{i=1}^{k+1} \log(F_a(a_i; u, v) - F_a(a_{i-1}; u, v))$$

$$+ \frac{1}{k+1} \sum_{i=1}^{k+1} \log(F_b(b_i; e, c) - F_b(b_{i-1}; e, c)). \tag{5}$$

We develop here an iterative algorithm, called spacing-likelihood algorithm, for efficiently and speedily estimating the unknown parameter Θ. The ML estimate $\hat{\Theta}$, obtained via QMI method, can be used for starting the algorithm. Let $\Theta^{(m)}$ denote the estimate of Θ obtained after the mth iteration of the spacing-likelihood algorithm. The $(m+1)$st iteration proceeds as follows.

S step Maximize, with respect to A and B, the spacing function $S(A, B; \Theta^{(m)}, \{t_1, \cdots, t_n\})$. Let $\hat{A}$ denote the estimate of A and $\hat{B}$ the estimate of B.

L step Maximize, with respect to Θ, the log-likelihood function $\ell(\Theta; \{t_1, \cdots, t_n\}, \hat{A}, \hat{B})$. Denote the maximizer by $\Theta^{(m+1)}$.

Stop the recursive procedure when two consecutive estimates are close enough; that is, for example, after m iterations we have $||\Theta^{(m)} - \Theta^{(m-1)}|| < \varepsilon$. ε is a pre-assigned positive tolerance limit. Then $\Theta^{(m)}$ is an optimal estimate for Θ; $\hat{A}$ and $\hat{B}$ – obtained after the mth S step – are optimal estimates for A and B respectively. We here do not prove the convergence of spacing-likelihood algorithm. In fact, the intuition behind spacing-likelihood algorithm has been around for a long time: alternate between estimating the unknown parameter Θ and the latent a_j^i's and b_j^i's. If we employ QMI method for estimating Θ, the optimal estimates for A and B can then be obtained by maximizing the spacing function $S(A, B; \hat{\Theta}, \{t_1, \cdots, t_n\})$. R code for spacing-likelihood algorithm is available on request.

Remark 5. *As will be revealed in Section 5, QMI method and spacing-likelihood algorithm are two complementary parameter-estimation methods. QMI method is more efficient when the number of repairs, k, is small, due to the $2k_i$-dimensional integration and the curse of dimensionality. Spacing-likelihood algorithm, on the other hand, is more efficient when k is moderate or large; this is because, in the spacing function (5), both n and k play the role of "sample size". Hence, the larger the values of n and k, the better the performance of spacing-likelihood algorithm.*

4 GOODNESS-OF-FIT TEST

Owning to the efforts by many statisticians, led by Pearson, Fisher, Cramér, von Mises and Kolmogorov, there have been developed a few general principles for hypothesis testing: chi-square tests, empirical-distribution-function (EDF) tests, likelihood-ratio tests, et al.. Here, by an EDF test statistic we mean a statistic (e.g., the Anderson-Darling statistic) expressed as a functional of the empirical distribution function. In the present study, we develop a generic procedure for testing fit to the improvement factor model by employing EDF tests, because all the other types of tests are not applicable. For example, likelihood-ratio tests are developed with the intention to compare two nested models, with one model being an expanded form of the other model (Swihart et al., 2014). In Section 3 of the supplementary material, we explained why chi-square goodness-of-fit tests cannot be employed for testing fit to the improvement factor model.

Among the many EDF tests, we here examine the Kolmogorov-Smirnov test, the Cramér-von Mises test and the Anderson-Darling test, due to their well-recognized power for testing fit.

Assume that the data $\{H_1, \cdots, H_n\}$ have been arranged in ascending order with $H_1 \leq H_2 \leq \cdots \leq H_n$. For $1 \leq i \leq n$, invoke the probability integral transformation and calculate $x_i = 1 - \exp(-H_i)$. Denote by $\hat{F}_n(x)$ the empirical distribution function of the data $\{x_1, \cdots, x_n\}$. The Kolmogorov-Smirnov statistic is defined by

$$\hat{K}_n^2 = \sqrt{n} \sup_{0 < x < 1} |\hat{F}_n(x) - x| = \sqrt{n} \, \max\left(\max_{1 \leq i \leq n}\left(\frac{i}{n} - x_i\right), \, \max_{1 \leq i \leq n}\left(x_i - \frac{i-1}{n}\right)\right);$$

the Cramér-von Mises statistic is defined by

$$\hat{W}_n^2 = n\int_0^1 [\hat{F}_n(x) - x]^2 dx = \sum_{i=1}^n \left(x_i - \frac{i-0.5}{n}\right)^2 + \frac{1}{12n};$$

and the Anderson-Darling statistic is defined by

$$\hat{A}_n^2 = n\int_0^1 \frac{[\hat{F}_n(x) - x]^2}{x(1-x)}dx = -n - \frac{1}{n}\sum_{i=1}^n (2i-1)\left[\log(x_i) + \log(1 - x_{n+1-i})\right].$$

Darling (1955) examined the limiting distributions of the above three statistics by analytical methods assuming that the unknown parameter is one-dimensional. Sukhatme (1972) extended Darling's (1955) result to multi-dimensional parameters. A general study of the weak convergence of the estimated empirical process, $\sqrt{n}[\hat{F}_n(x) - x]$, was carried out by Durbin (1973), stating that under some regularity conditions the estimated empirical process converges weakly to a centered Gaussian process. A survey of results connected to Durbin's (1973) work can be found in Shorack and Wellner (2009). Owning to the consistency and efficiency of maximum spacing estimation (Ranneby, 1984), the property of weak convergence of the estimated empirical process, $\sqrt{n}[\hat{F}_n(x) - x]$, still holds in our problem. Therefore, the above three statistics are applicable to testing the fit to the improvement factor model.

It is important to note that the H_i's are calculated from not only the estimate of Θ but also the estimates of the a_j^i's and b_j^i's. Hence, according to David and Johnson (1948),

- the above test statistics are no longer distribution-free; computation of critical values has to be carried through for each particular combination of the assumptions A1-A3;

- the distributions of the test statistics may depend on the unknown parameter Θ;

- the distributions of the test statistics may depend on the total number of repairs k.

In view of the complexity of our problem, the parametric bootstrap procedure (Stute et al., 1993) constitutes the only possibility for calculating critical values. Given $\hat{\Theta}$, an estimate of Θ obtained from collected data, the generic procedure for testing fit to the improvement factor model is presented as follows.

(1) Randomly generate k values from the beta distribution $F_a(x; \hat{u}, \hat{v})$ and k values from the gamma distribution $F_b(x; \hat{e}, \hat{c})$. For $1 \le i \le n$, randomly generate a value from the conditional distribution

$$\Pr(T_i \le t \mid T_i > t_{k_i}^i) = 1 - \exp(-\Pi_{v=1}^{k_i} b_v^i H(\delta^i(t; a_1^i, \cdots, a_{k_i}^i); \hat{\boldsymbol{\theta}})), \quad t > t_{k_i}^i.$$

Here, the a_j^i's and b_j^i's are simulated values.

(2) Employ QMI method and/or spacing-likelihood algorithm to obtain optimal estimates of Θ, the a_j^i's and the b_j^i's. Calculate the H_i's and, finally, the value of a test statistic.

(3) Repeat steps (1) and (2) for a sufficiently large number of times in order to obtain accurate asymptotic approximations (in n) of the p-values.

When dealing with censored data, let $\hat{\mathbb{F}}_n(x)$ be the Kaplan-Meier estimate of the distribution function (not survival function) of the data $\{x_1, \cdots, x_n\}$. The corresponding statistics are the results of replacing the empirical distribution function $\hat{F}_n(x)$ with $\hat{\mathbb{F}}_n(x)$.

5   NUMERICAL STUDY

In real-life practice, maintenance-performing time points are scheduled according to certain criteria. For example, when maintaining a safety-critical machine, the machine usually will be repaired once its reliability drops to a critical threshold. Undoubtedly, it is the number of repairs, rather than the repair-performing epochs, that affects the performance of QMI method or spacing-likelihood algorithm. Therefore, without loss of generality, we only study the case where maintenance-performing time points are equally spaced with $\Delta(> 0)$ being the fixed time between two consecutive repairs. That is, we have $t_j^i = j\Delta$ for $1 \le i \le n$ and $1 \le j \le k_i$. For each pair of n and Δ, the lifetime data $\{t_1, \cdots, t_n\}$ are simulated in the manner given in Section 4.1 of the supplementary material.

Let $\mathbb{W}(H(t; \boldsymbol{\theta}))$ denote a Weibull distribution with $H(t; \boldsymbol{\theta})$ being its cumulative hazard function. We might assume that, before performing any repair, the lifetimes of the n independent machines follow a Weibull distribution with the cumulative hazard function being $H(t; \boldsymbol{\theta}) = (wt)^\tau$, in which $\boldsymbol{\theta} = (w, \tau)$. We know that different values of τ can have marked effects on the behavior of the Weibull distribution. To show that the improvement factor model can be applied to lifetime models with different hazard rate functions, we study the cases of τ being 0.5, 1 and 3, respectively corresponding to decreasing, constant and increasing hazard rate functions.

The value of w is fixed at 1. Herein we only present the parameter-estimation results when $\{w = 1, \tau = 3, u = 3, v = 1, c = 1, e = 3\}$ to illustrate our indicative findings. Parameter-estimation results for $\tau = 0.5$ and $\tau = 1$ can be found in the supplementary material for this paper. To study the impact of sample size on the performance of QMI method and spacing-likelihood algorithm, three levels of sample size are examined: small ($n = 20$), moderate ($n = 50$), and large ($n = 100$). As pointed out in Section 4.1 of the supplementary material, the value of k_i is not fully deterministic. To study the impact of the total number of repairs on the performance of QMI method and spacing-likelihood algorithm, we gradually increase Δ from 0.4 to 1.0 with step size 0.1. The range of Δ is determined such that the total number of repairs vary from larger than n to smaller than half of n.

For a particular parameter configuration, we run the data-generating procedure in Section 4.1 of the supplementary material for 1000 times. For each of the 1000 data sets, employ a parameter-estimation method to estimate Θ; then calculate the mean and mean squared error (MSE) of the 1000 estimates. The results obtained via QMI method are summarized in Table 1 (for $n = 20$), in

Table 1: Mean and MSE of the 1000 ML estimates via QMI method ($n = 20$).

Δ	$\bar{k}$	$\hat{w}$	$\hat{\tau}$	$\hat{u}$	$\hat{v}$	$\hat{c}$	$\hat{e}$
0.4	28.441	2.4030	3.4242	3.7917	1.8867	1.8929	3.8196
		1.5602	1.3383	1.4454	1.5201	1.5708	1.4540
0.5	22.346	1.8118	3.3729	3.6280	1.7361	1.7064	3.7389
		0.9586	1.2278	1.2840	0.9700	0.9654	1.2439
0.6	18.275	1.4905	3.3021	3.4100	1.5883	1.5489	3.6225
		0.5446	0.9780	1.0320	0.6348	0.6656	0.9408
0.7	15.178	1.3890	3.2225	3.3369	1.3504	1.3963	3.5126
		0.4215	0.8862	0.8611	0.4587	0.4651	0.8364
0.8	12.223	1.2967	3.1195	3.2086	1.3009	1.3500	3.4156
		0.2915	0.7640	0.7701	0.2571	0.2598	0.7369
0.9	9.78	1.2079	3.0712	3.1345	1.2896	1.2889	3.2897
		0.1691	0.6656	0.6554	0.1918	0.1990	0.6336
1.0	7.446	1.1685	3.0504	3.1098	1.1643	1.1806	3.2168
		0.1213	0.5498	0.5690	0.1495	0.1593	0.5899

Table 2 (for $n = 50$), and in Table 3 (for $n = 100$). $\bar{k}$ is the sample mean of the 1000 simulated total number of repairs. For each pair of Δ and $\bar{k}$, the first row gives the sample mean and the second row gives the MSE. Halton sequence (for dimensions up to 6) and Sobol sequence (for dimensions

Table 2: Mean and MSE of the 1000 ML estimates via QMI method ($n = 50$).

Δ	$\bar{k}$	$\hat{w}$	$\hat{\tau}$	$\hat{u}$	$\hat{v}$	$\hat{c}$	$\hat{e}$
0.4	70.941	1.1316	3.5258	3.5171	1.1283	1.1404	3.5823
		0.1425	1.1937	1.0544	0.2818	0.2918	1.0836
0.5	56.304	1.1209	3.4190	3.4214	1.1275	1.1365	3.5199
		0.1247	1.1486	0.9656	0.2663	0.2852	1.0096
0.6	45.607	1.1144	3.3224	3.2259	1.1172	1.1305	3.4512
		0.1171	1.1376	0.9188	0.2492	0.2523	0.9876
0.7	37.705	1.0952	3.1998	3.2011	1.1084	1.1249	3.4190
		0.0823	0.9903	0.7560	0.2337	0.2471	0.7760
0.8	31.174	1.0883	3.1224	3.1213	1.1056	1.1116	3.2787
		0.0747	0.9023	0.7240	0.2112	0.2298	0.7456
0.9	24.321	1.0840	3.0842	3.0777	1.0960	1.1036	3.1207
		0.0577	0.8622	0.6696	0.2019	0.2170	0.7136
1.0	18.46	1.0747	3.0243	3.0165	1.0881	1.0966	3.0247
		0.0441	0.7089	0.6184	0.1747	0.1911	0.6992

Table 3: Mean and MSE of the 1000 ML estimates via QMI method ($n = 100$).

Δ	$\bar{k}$	$\hat{w}$	$\hat{\tau}$	$\hat{u}$	$\hat{v}$	$\hat{c}$	$\hat{e}$
0.4	141.838	1.2963	3.7303	3.6941	1.2916	1.2662	3.7270
		0.0789	0.6410	0.7138	0.1733	0.1991	0.7578
0.5	111.871	1.2596	3.6352	3.6688	1.2688	1.2196	3.5967
		0.0616	0.6370	0.6356	0.1694	0.1814	0.6412
0.6	91.043	1.2128	3.5002	3.5334	1.2274	1.2086	3.5672
		0.0477	0.5939	0.5859	0.1610	0.1700	0.5992
0.7	75.358	1.1998	3.3878	3.4457	1.2034	1.1848	3.4029
		0.0420	0.5551	0.5230	0.1495	0.1598	0.5882
0.8	61.62	1.1365	3.3693	3.3148	1.1590	1.1646	3.2948
		0.0356	0.5383	0.5047	0.1375	0.1490	0.5721
0.9	48.873	1.0985	3.1775	3.2027	1.1227	1.1377	3.2264
		0.0323	0.5302	0.4677	0.1243	0.1345	0.5099
1.0	37.19	1.0803	3.0630	3.0936	1.0961	1.1035	3.0710
		0.0221	0.5256	0.3737	0.1190	0.1283	0.4051

more than 6) are employed to generate quasi-random points for QMI method. The results obtained via spacing-likelihood algorithm are reported in Table 4 (for $n = 20$), in Table 5 (for $n = 50$), and

Table 4: Mean and MSE of the 1000 ML estimates via spacing-likelihood algorithm ($n = 20$).

Δ	$\bar{k}$	$\hat{w}$	$\hat{\tau}$	$\hat{u}$	$\hat{v}$	$\hat{c}$	$\hat{e}$
0.4	28.608	1.2735	4.3695	3.9459	1.1826	2.0931	3.8841
		0.0889	2.3956	1.6901	0.1781	1.6477	1.4701
0.5	22.449	1.1907	4.2884	3.6869	1.1497	2.0572	3.8372
		0.0480	2.0837	1.1880	0.1345	1.6752	1.3771
0.6	18.22	1.1908	4.1832	3.7171	1.2437	2.1305	3.6604
		0.0482	1.8372	1.4069	0.2418	1.8105	1.2283
0.7	15.079	1.1558	4.2535	3.6706	1.2240	2.1352	3.5310
		0.0352	1.9679	1.2854	0.2401	1.8848	1.1747
0.8	12.403	1.1041	4.2108	3.5333	1.1742	2.1285	3.5125
		0.0190	1.9846	1.3808	0.2235	1.7825	1.0939
0.9	9.793	1.0639	4.2545	3.6277	1.1720	2.2210	3.6376
		0.0096	2.0378	1.4552	0.2311	2.1959	1.5411
1.0	7.414	1.0364	4.2625	3.6680	1.3193	2.2233	3.4400
		0.0059	2.0666	1.4954	0.4638	2.2774	1.1006

in Table 6 (for $n = 100$).

Comparing Tables 1-3, it is clear that with the sample size fixed, a larger Δ, and hence a smaller k, shall entail a more accurate ML estimate; with k fixed, a larger n shall entail a more accurate ML estimate; QMI method fails when k is not small. Therefore, for a fixed sample size n, if k is smaller than a certain value, then QMI method shall produce an accurate ML estimate of Θ; if k exceeds the value, then QMI method will break down. One reason for QMI method being only applicable to small k may be that high-dimensional integration will make the marginal log-likelihood function very flat, and therefore the returned estimate may be quite distant from the true value.

It can be concluded from Tables 4–6 that with the sample size fixed, a smaller Δ, and hence a larger k, shall produce a more accurate ML estimate; with k fixed, a larger n shall produce a more accurate ML estimate; spacing-likelihood algorithm fails when k is small. As stated in Remark 5, one reason for spacing-likelihood algorithm being only applicable to moderate or large k is that, in the spacing function (5), both n and k play the role of "sample size".

According to Sections 4.2 and 4.3 in the supplementary material and Tables 1-6, it can be concluded that (see Table 7) (1) QMI method is more efficient when the total number of repairs

Table 5: Mean and MSE of the 1000 ML estimates via spacing-likelihood algorithm ($n = 50$).

Δ	$\bar{k}$	$\hat{w}$	$\hat{\tau}$	$\hat{u}$	$\hat{v}$	$\hat{c}$	$\hat{e}$
0.4	71.224	1.0142	3.0582	3.1503	1.0681	1.0921	3.2170
		0.0033	0.4746	0.3235	0.0429	0.0559	0.3789
0.5	56.095	1.0186	3.0799	3.1753	1.0855	1.1064	3.2846
		0.0034	0.4898	0.3645	0.0698	0.0875	0.4512
0.6	45.553	1.0194	3.0978	3.2079	1.1185	1.1232	3.2909
		0.0037	0.5158	0.4891	0.0973	0.1009	0.5322
0.7	37.74	1.0224	3.1080	3.2607	1.1557	1.1651	3.3290
		0.0041	0.5727	0.5994	0.1128	0.1306	0.7039
0.8	30.765	1.0238	3.1261	3.4085	1.1812	1.1964	3.4114
		0.0042	0.6455	0.8290	0.1315	0.1572	0.8782
0.9	24.291	1.0358	3.1461	3.4986	1.2020	1.2159	3.5009
		0.0052	0.7056	0.9967	0.1854	0.2011	1.0468
1.0	18.458	1.0601	3.1522	3.5393	1.2361	1.2644	3.7225
		0.0090	0.7495	1.1977	0.2455	0.3209	1.3332

Table 6: Mean and MSE of the 1000 ML estimates via spacing-likelihood algorithm ($n = 100$).

Δ	$\bar{k}$	$\hat{w}$	$\hat{\tau}$	$\hat{u}$	$\hat{v}$	$\hat{c}$	$\hat{e}$
0.4	141.831	1.0033	3.0351	3.0794	1.0250	1.0443	3.0978
		0.0016	0.3280	0.1388	0.0157	0.0181	0.2180
0.5	111.837	1.0120	3.0485	3.0819	1.0377	1.0706	3.1055
		0.0019	0.3417	0.2127	0.0250	0.0293	0.2504
0.6	91.128	1.0249	3.0618	3.0856	1.0479	1.0715	3.1211
		0.0020	0.3653	0.2284	0.0312	0.0350	0.2801
0.7	75.717	1.0301	3.0816	3.1163	1.0650	1.1119	3.1509
		0.0021	0.4245	0.2346	0.0403	0.0542	0.3183
0.8	61.969	1.0399	3.0963	3.1243	1.0772	1.1240	3.1863
		0.0022	0.4615	0.3031	0.0523	0.0686	0.3831
0.9	48.795	1.0408	3.1183	3.1429	1.0947	1.1345	3.2424
		0.0025	0.5039	0.3704	0.0718	0.0917	0.4848
1.0	37.304	1.0456	3.1393	3.2585	1.1226	1.1404	3.3062
		0.0036	0.5681	0.5150	0.1111	0.1303	0.5397

is small, no matter how large the sample size is; (2) spacing-likelihood algorithm is more efficient when both the total numbers of repairs and the sample size are not small; (3) when the sample size is small but the total number of repairs is not small, both QMI method and spacing-likelihood algorithm fail.

Table 7: Comparative performance of QMI method and spacing-likelihood algorithm.

k / n	small	moderate	large
small	QMI	N.A.	N.A.
moderate	QMI	spacing-likelihood	spacing-likelihood
large	QMI	spacing-likelihood	spacing-likelihood

Via intensive simulation study we found that the boundary between small and moderate is around 30. Yet, when the sample size is small, for QMI method to be applicable, the total number of repairs should be smaller than the sample size; that is, the total number of repairs should be relatively small.

More numerical examples can be found in the supplementary material. A real data set from Électricité de France (a French electric utility) is analyzed in Section 4.4 of the supplementary material, in which we also compared the ameliorated improvement factor model with another two imperfect-maintenance models.

6 CONCLUSIONS

In this paper we proposed a generic maintenance model that can be used to characterize perfect, imperfect and minimal maintenance, and can be coupled to any maintenance policy as long as the maintained machine's operating duration is modeled by a lifetime distribution. The proposed maintenance model is grounded on practical considerations and is superior to most existing maintenance models. For its practical implementation, we developed two methods for parameter estimation and a framework for testing its goodness of fit. The two parameter-estimation methods complete each other, and both of them can readily handle complete data and incomplete data. For testing goodness of fit, we identified certain suitable test statistics and outlined the steps for approximating the distribution of a test statistic by using a parametric bootstrap procedure. By virtue of the parametric bootstrap procedure, the proposed test statistics can be used to measure discrepancy for any combination of sample size, number of repairs, percentage of censoring, etc.

The methods and numerical study presented here have left a number of practical issues untouched. Further studies need to be undertaken in future work.

- As shown in Table 7, both QMI method and spacing-likelihood algorithm fail when the sample size is small yet the total number of repairs is not small. We should develop other methods to solve this problem.

- For lack of space, we here didn't examine the power of the three EDF tests. Our concern is that their performance depends on several factors, such as the sample size, number of repairs, degree of censoring, underlying lifetime model, etc. Conducting a thorough study of the power of the three tests will make this paper extremely lengthy. However, it would be of interest to study the impact from sample size (or number of repairs) on the performance of the three tests – to support experimental design.

SUPPLEMENTARY MATERIALS

Technical details: In the PDF file, we provided a further discussion on the ameliorated improvement factor model and a short review of chi-square goodness-of-fit tests. We extended QMI method and spacing-likelihood algorithm for dealing with censored data. The PDF file also includes parameter-estimation results for $\tau = 0.5$ and $\tau = 1$, and the performance of the S step of spacing-likelihood algorithm was examined. A real data set was analyzed (PDF file).

Source code: The zipped package contains R codes for the two parameter-estimation methods (ZIP file).

ACKNOWLEDGMENTS

REFERENCES

Ahmad, R. and Kamaruddin, S. (2012). An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering*, 63(1):135 – 149.

Anatolyev, S. and Kosenok, G. (2005). An alternative to maximum likelihood based on spacings. *Econometric Theory*, 21(2):472–476.

Block, H. W., Borges, W. S., and Savits, T. H. (1985). Age-dependent minimal repair. *Journal of Applied Probability*, 22(2):pp. 370–385.

Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3):394–403.

Darling, D. A. (1955). The cramer-smirnov test in the parametric case. *The Annals of Mathematical Statistics*, 26(1):1–20.

David, F. N. and Johnson, N. L. (1948). The probability integral transformation when parameters are estimated from the sample. *Biometrika*, 35(1/2):182–190.

de Toledo, M. L. G., Freitas, M. A., Colosimo, E. A., and Gilardoni, G. L. (2015). ARA and ARI imperfect repair models: Estimation, goodness-of-fit and reliability prediction . *Reliability Engineering & System Safety*, 140:107 – 115.

Dijoux, Y. and Idee, E. (2013). Classes of virtual age models adapted to systems with a burn-in period. *IEEE Transactions on Reliability*, 62(4):754–763.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, 1(2):279–290.

Ekstrom, M. (1998). On the consistency of the maximum spacing method. *Journal of Statistical Planning and Inference*, 70(2):209–224.

Ekstrom, M. (2001). Consistency of generalized maximum spacing estimates. *Scandinavian Journal of Statistics*, 28(2):343–354.

Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods*. Springer, 2nd edition.

Guo, C., Wang, W., Guo, B., and Si, X. (2013). A maintenance optimization model for mission-oriented systems based on wiener degradation. *Reliability Engineering & System Safety*, 111(0):183–194.

Khatab, A., Ait-Kadi, D., and Rezg, N. (2014). Availability optimisation for stochastic degrading systems under imperfect preventive maintenance. *International Journal of Production Research*, 52(14):4132–4141.

Khojandi, A., Maillart, L. M., and Prokopyev, O. A. (2014). Optimal planning of life-depleting maintenance activities. *IIE Transactions*, 46(7):636–652.

Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied probability*, 26(89-102).

Li, L., Hanson, T., Damien, P., and Popova, E. (2014). A bayesian nonparametric test for minimal repair. *Technometrics*, 56(3):393–406.

Lin, D., Zuo, M. J., and Yam, R. C. M. (2000). General sequential imperfect preventive maintenance models. *International Journal of Reliability, Quality and Safety Engineering*, 7(3):253–266.

Lindqvist, B. H. (2006). On the statistical modeling and analysis of repairable systems. *Statistical Science*, 21(4):532–551.

Liu, Y., Huang, H., Wang, Z., Li, Y., and Yang, Y. (2013). A joint redundancy and imperfect maintenance strategy optimization for multi-state systems. *IEEE Transactions on Reliability*, 62(2):368–378.

Liu, Y., Huang, H., and Zhang, X. (2012). A data-driven approach to selecting imperfect maintenance models. *IEEE Transactions on Reliability*, 61(1):101–112.

Malik, M. (1979). Reliable preventive maintenance policy. *AIIE Transactions*, 11(3):221–228.

Melchor-Hernandez, C. L., Rivas-Davalos, F., Maximov, S., Coria, V., and Guardado, J. (2015). A model for optimizing maintenance policy for power equipment. *International Journal of Electrical Power & Energy Systems*, 68:304 – 312.

Mercier, S. and Castro, I. T. (2013). On the modelling of imperfect repairs for a continuously monitored gamma wear process through age reduction. *Journal of Applied Probability*, 50(4):1057–1076.

Morokoff, W. J. and Caflisch, R. E. (1995). Quasi-monte carlo integration. *Journal of Computational Physics*, 122:218–230.

Nakagawa, T. (1979). Optimum policies when preventive maintenance is imperfect. *IEEE Transactions on Reliability*, R-28(4):331–332.

Niederreiter, H. (1978). Quasi-monte carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84(6):957–1041.

Pandey, M., Zuo, M. J., and Moghaddass, R. (2013). Selective maintenance modeling for a multistate system with multistate components under imperfect maintenance. *IIE Transactions*, 45(11):1221–1234.

Papageorgiou, A. (2001). Fast convergence of quasi-monte carlo for a class of isotropic integrals. *Mathematics of Computation*, 70:297–306.

Park, J., Chang, W., and Lie, C. (2012). Stress-reducing preventive maintenance model for a unit under stressful environment. *Reliability Engineering & System Safety*, 108(0):42–48.

Pulcini, G. (2013). A model-driven approach for the failure data analysis of multiple repairable systems without information on individual sequences. *IEEE Transactions on Reliability*, 62(3):700–713.

Ramirez, P. A. P. and Utne, I. B. (2013). Decision support for life extension of technical systems through virtual age modelling. *Reliability Engineering & System Safety*, 115(0):55–69.

Ranneby, B. (1984). The maximum spacing method-an estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, 11(2):93–112.

Schlier, C. (2004). Error trends in quasi-monte carlo integration. *Computer Physics Communications*, 159(2):93–105.

Shafiee, M. and Chukova, S. (2013). Maintenance models in warranty: A literature review. *European Journal of Operational Research*, 229(3):561–572.

Shafiee, M., Chukova, S., and Finkelstein, M. (2012). Warranty and optimal upgrade strategy for used systems: An electric drill case study. *Asia-Pacific Journal of Operational Research*, 29(04):1250023.

Shorack, G. and Wellner, J. (2009). *Empirical Processes with Applications to Statistics*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Singpurwalla, N. D. (2006). The hazard potential. *Journal of the American Statistical Association*, 101(476):1705–1717.

Sloan, I. H. and Wozniakowski, H. (1998). When are quasi-monte carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14(1):1–33.

Stute, W., Manteiga, W., and Quindimil, M. (1993). Bootstrap based goodness-of-fit-tests. *Metrika*, 40(1):243–256.

Sukhatme, S. (1972). Fredholm determinant of a positive definite kernel of a special type and its application. *The Annals of Mathematical Statistics*, 43(6):1914–1926.

Swihart, B. J., Goldsmith, J., and Crainiceanu, C. M. (2014). Restricted likelihood ratio tests for functional effects in the functional linear model. *Technometrics*, 56(4):483–493.

Tambe, P., Mohite, S., and Kulkarni, M. (2013). Optimisation of opportunistic maintenance of a multi-component system considering the effect of failures on quality and production schedule: A case study. *The International Journal of Advanced Manufacturing Technology*, 69(5-8):1743–1756.

Tanwar, M., Rai, R. N., and Bolia, N. (2014). Imperfect repair modeling using kijima type generalized renewal process. *Reliability Engineering & System Safety*, 124:24 – 31.

Wang, W. (2012). An overview of the recent advances in delay-time-based maintenance modelling. *Reliability Engineering & System Safety*, 106(0):165–178.

Xia, T., Xi, L., Zhou, X., and Du, S. (2012). Modeling and optimizing maintenance schedule for energy systems subject to degradation. *Computers & Industrial Engineering*, 63(3):607 – 614.

Zhang, M., Gaudoin, O., and Xie, M. (2015). Degradation-based maintenance decision using stochastic filtering for systems under imperfect maintenance. *European Journal of Operational Research*. In press.

Zhang, M., Xie, M., and Gaudoin, O. (2013). A bivariate maintenance policy for multi-state repairable systems with monotone process. *IEEE Transactions on Reliability*, 62(4):876–886.