# A Data-Driven Approach to Selecting Imperfect Maintenance Models

Yu Liu, Hong-Zhong Huang, *Member, IEEE*, and Xiaoling Zhang

*Abstract*—Many imperfect maintenance models have been developed to mathematically characterize the efficiency of maintenance activity from various points of view. However, the adequacy of an imperfect maintenance model must be validated before it is used in decision making. The most adequate imperfect maintenance model among the candidates to facilitate decision making is also desired.
The contributions of this paper lie in three aspects:
1  it proposes an approach to conducting a goodness-of-fit test,
2  it introduces a Bayesian approach to selecting the most adequate model among several competitive candidates, and
3  it develops a framework that incorporates the model selection results into the preventive maintenance decision making.
The effectiveness of the proposed methods is demonstrated by three designed numerical studies. The case studies show that the proposed methods are able to identify the most adequate model from the competitive candidates, and incorporating the model selection results into the maintenance decision model achieves better estimation for applications with limited data.

*Index Terms*—Bayesian model selection, bootstrap sampling, goodness-of-fit, imperfect maintenance model, u-pooling method.

## Acronyms

| | |
|---|---|
| CM | corrective maintenance |
| PM | preventive maintenance |
| MLE | maximum likelihood estimation |
| GOF | goodness-of-fit |

## Notation

| | |
|---|---|
| $t_k^p$ | time interval between the $(k-1)$th and the $k$th PM |
| $\lambda_k(t)$ | failure intensity function in the $k$th PM cycle |
| $a_k$ | parameter of the imperfect maintenance model in the $k$th PM cycle |
| $V_k$ | virtual age of the system after the $k$th PM |
| $n_{l,k}$ | number of failures of the $l$th system in the $k$th PM cycle |
| $t_{l,(k,i)}^f$ | time of the $i$th failure of the $l$th system in the $k$th PM cycle |
| $S_{l,(k,i)}(t)$ | conditional survival distribution of the $l$th system after the $(i-1)$th failure of the $k$th PM cycle |
| $T_{l,(k,i)}$ | inter-arrival time between the $(i-1)$th and the $i$th failures of the $k$th PM cycle for the $l$th repairable system |
| $\Pr\{H_k\|data\}$ | posterior probability of hypothesis $H_k$ |
| $\Pr\{\hat{\boldsymbol{a}}\|H_i\}$ | distribution of estimated parameter $\hat{\boldsymbol{a}}$ of the $i$th imperfect maintenance model |
| $T_{PM}$ | fixed time interval for periodic PM policy |
| $N_{PM}$ | number of PMs before the system is discarded |

## I. Introduction

**M**AINTENANCE is an effective way to reduce the failure frequency of repairable systems, restore system performance, and prolong the systems' remaining life. It has been adopted in a multitude of systems, such as civil aircrafts, automotive vehicles, and manufacturing systems. In general, maintenance activity can be classified into five categories based on their impacts on the physical condition of maintained systems [1]–[3]: better than new, perfect, imperfect, minimal, and worse. Better than new maintenance brings the items' condition to a state with a smaller failure intensity than a brand-new identical one. A perfect maintenance (e.g. a complete replacement) is able to recover the repaired system back to a brand new ("*as good as new*") condition; whereas a minimal maintenance (e.g. correcting a minor flaw) assumes a maintenance action can only leave the system in the same condition as it was before failure ("*as bad as old*"). Worse maintenance results in a negative impact on the systems' condition due to an ineffective maintenance. Nevertheless, it is widely recognized that most maintenance actions restore the system to a condition somewhere between "as good as new" and "as bad as old", and this sort of maintenance is called *imperfect maintenance* [2].

To quantify the efficiency of an imperfect maintenance activity, a considerable amount of literature has been published. The most recent survey of the imperfect maintenance models for binary state systems can be found in [2]. The most relevant efforts among the existing models are: the Brown-Proschan model (See Brown and Proschan [4]), the Kijima Type I and II models (see Kijima [5], [6]), the improvement factor method (see Malik [7]), the hazard rate model (see Nakagawa [8]), the hybrid model

(see Lin [9]), the geometric process (see Lam [10]), the quasi-renewal model (see Wang and Pham [11], [12]), and the superposed renewal process (see Kallen [13]). The statistical inference of the unknown parameters in the imperfect maintenance models has also been studied [14]–[21]. Most recently, Wu and Zuo [22] presented a comprehensive study to reveal the commonality and interrelationship among some models. Also, some of the imperfect maintenance models, such as the $(p, q)$ model (see Nakagawa [8] and Brown and Proschan [4]), the $(p(t), q(t))$ model (see Block *et al.* [23]), and the $(p(n, t), q(n, t), s(n, t))$ model (see Makis and Jardine [24]), can be only used for characterizing the corrective maintenance rather than preventive maintenance [1]. In addition, several attempts have been made to investigate the imperfect maintenance models under the context of multi-state systems [25]–[27]. Nevertheless, different imperfect maintenance models are subject to different perspectives and assumptions of underlying physics. For example, from the system age view point, some imperfect maintenance models (e.g. Kijima type I and II models [5], [6], and the age reduction model [15]) assume that the effective age of the system is reduced after maintenance. Some literature, from a perspective of failure intensity, believes that the effect of maintenance leads to a direct reduction of failure intensity (e.g. the intensity reduction model [15]); or after maintenance activity, the age of the system becomes zero, but the system has a faster degradation trend in the next maintenance cycle (e.g. the hazard rate model [8], the geometric process [10], or the quasi-renewal model [11]).

Because every imperfect maintenance model characterizes the stochastic behavior of maintained systems via its own assumption of underlying physics, to better quantify the effect of maintenance actions, an imperfect maintenance model should be selected based on the knowledge of actual underlying physics. For example, for the system subject to wear, replacing the worn out component with a better one can be regarded as reducing the system effective age, and the age-based imperfect maintenance model is suitable in such a circumstance. However, it is actually very hard, even impossible, to observe and comprehensively understand the true physics and mechanisms, especially when a complicated system is maintained by replacing and fixing multiple components subject to various failure modes. When data from system operation and maintenance are the only available information, industry practitioners become concerned about whether or not the pre-assumed imperfect maintenance model is adequate enough to describe the effect of maintenance. To examine the adequacy of the pre-assumed model, two methods can be used: graphical check, and the goodness-of-fit (GOF) test [28]–[31]. The general idea of a graphical check is to verify whether the cumulative number of system failures from operation data has an identical trend with that from the pre-assumed maintenance model. Although the graphical check is straightforward and intuitive, it has two major drawbacks. First, the visual judgment is qualitative, and there is no unified criterion enabling analysts to verify whether the examined model is adequate or not. Second, to derive the expected cumulative number of failures requires Monte Carlo simulation, which is a computationally expensive task [3]. The GOF test is a mathematically rigorous method to examine the model adequacy, and it has been applied in reliability engineering extensively [28], [32]. Nonetheless, most GOF tests for repairable systems concentrate on checking the adequacy of a pre-assumed non-homogenous Poisson process (NHPP) model, i.e. whether the repairable systems can be modeled by a specified NHPP model (e.g. power law process, or log-linear process) [28], [32]. Due to the intervention of maintenance actions, the stochastic behavior of imperfect maintained systems becomes complicated. Thereby, the traditional GOF test for simple homogenous and non-homogenous Poisson processes cannot be tailored for such a situation. For the system under imperfect corrective repair, methods concerning the GOF test for the first time to failure distribution have been developed [33]. However, the imperfect repair effects characterized by the Block and Borges and Savits (BBS) model are assumed known in advance. One of the focuses of this work is to examine the adequacy of the pre-assumed imperfect maintenance model rather than the first time to failure distribution. To the best of our knowledge, the GOF test for the imperfect maintenance model has seldom been investigated, and it is one of our interests in this paper.

Beyond the issue of checking model adequacy, another issue is how to determine the best model among multiple adequate models. This issue is recognized as the model selection problem, which has been investigated to choose failure models [34]. Because there is no available knowledge about the underlying physics, the only approach is to select the model which has the best fitness to the data, and it is, therefore, called the *data-driven* model selection approach in this paper.

To address the two aforementioned issues, a novel GOF test is first proposed to examine the adequacy of a pre-assumed imperfect maintenance model. By pooling all the failure data into a u-space, the adequacy of the imperfect maintenance model can be validated via statistically checking the uniformity of the u-pooling data. Along with the GOF test, a Bayesian model selection method is developed to choose the most adequate imperfect maintenance model among the competitive models. Because it is often impossible to get a sufficient amount of data to estimate the unknown model parameters accurately, the uncertainty associated with the parameter estimation is quantified via the empirical distribution obtained from the bootstrap method. The model selection results will be further incorporated into the maintenance decision framework. To demonstrate the effectiveness of the proposed methods, several sets of numerical studies are designed, and how sensitive the methods are with respect to the amount of data sets, as well as the impact from the imperfect maintenance efficiency, is also examined with the designed test problems.

The remainder of this paper is organized as follows. Section II reviews four imperfect preventive maintenance models with distinct underlying physics. The associated likelihood functions and parameter estimation are also formulated. The proposed GOF test for imperfectly maintained systems is introduced in Section III. The Bayesian approach for model selection is introduced in Section IV, along with a bootstrap method which is used to gain the distribution of the estimated model parameters. A framework of incorporating the model selection results into the preventive maintenance decision is proposed in Section V. The designed numerical studies are presented in Section VI to

illustrate and verify the effectiveness of the proposed methods, followed by the brief conclusion and discussion in Section VII.

## II. IMPERFECT MAINTENANCE MODELS AND PARAMETER ESTIMATION

In this paper, the proposed method is only demonstrated by the case where the scheduled preventive maintenance (PM) is imperfect, and corrective maintenance (CM) is the minimal repair. The PM is herein defined as the maintenance carried out at pre-determined intervals, or according to prescribed criteria, with the aim at reducing the probability of failure, or the degradation of the functioning of an item [1].

### A. Imperfect Preventive Maintenance Models

We introduce four imperfect preventive maintenance models which have different underlying physics, and they will act as the candidate models in the following designed test problems. Some assumptions and notations are given below before presenting the imperfect preventive maintenance models.

1. The PM is executed at time $t_1^p, (t_1^p + t_2^p), \ldots, (t_1^p +, \ldots, + t_{k-1}^p + t_k^p), \ldots$ where $t_k^p$ $(k = 1, 2, \ldots)$ is the time interval between the $(k-1)$th and the $k$th PMs. $t_k^p$ $(k = 1, 2, \ldots)$ can be either identical, or non-identical, representing periodic PM, and non-periodic PM respectively.

2. It is assumed that PMs are imperfect. The CMs which are treated as minimal repairs are carried out upon failures between adjacent PMs.

3. $\lambda_k(t)$ denotes the failure intensity function in the $k$th PM cycle, and $t$ is reset to zero at the beginning of each PM cycle. $\lambda_1(t)$ represents the initial failure intensity function of a brand-new system, and it is a *monotonic increasing function*.

4. Compared to system operation time, the time spent on maintenance is *negligible*.

The four imperfect maintenance models concerned in this paper are reviewed as follows.

*Model I (Kijima Type I Model [5], [6]):* The concept of virtual age has been used in the Kijima imperfect maintenance model. The general idea is that the system's health condition is related to the virtual age (called effective age), but not the actual working time. In the Kijima type I model, the virtual age of the system is formulated as

$$V_k = V_{k-1} + a_k t_k^p, (0 \le a_k \le 1), \tag{1}$$

where $V_k$ is the virtual age of the system immediately after the $k$th PM; $a_k$, with the range of [0,1], indicates the efficiency of the $k$th PM action on the virtual age of the system. If $a_k = 0$, the maintenance action is perfect, and will reduce the virtual aging incurred in the $k$th PM cycle; $a_k = 1$ represents a minimal maintenance, with no virtual age reduction after maintenance. The corresponding failure intensity function of the $k$th PM cycle is written as

$$\lambda_k(t; a_1, \ldots, a_{k-1}) = \lambda_1(t + V_{k-1}) = \lambda_1\left(t + \sum_{i=1}^{k-1} a_i t_i^p\right), \tag{2}$$

where $t$ is the working time only in the $k$th PM cycle. The PMs only reduce the age of the system since the last PM cycle.

*Model II (Kijima Type II Model [5], [6]):* Compared to the Kijima type I model, the Kijima type II model assumes that PM actions have a cumulative age-reduction effect, and will reduce the entire virtual age of the system incurred just before the maintenance. The virtual age of the system after the $k$th PM is defined as

$$V_k = a_k (V_{k-1} + t_k^p), (0 \le a_k \le 1). \tag{3}$$

If $a_k = 0$, the maintenance is perfect; $a_k = 1$ represents a minimal maintenance, with no virtual age reduction after maintenance. The failure intensity function of the $k$th PM cycle is written as

$$\lambda_k(t; a_1, \ldots, a_{k-1}) = \lambda_1(t + V_{k-1}) = \lambda_1\left(t + \sum_{i=1}^{k-1}\left(t_i^p \prod_{j=i}^{k-1} a_j\right)\right). \tag{4}$$

*Model III (Nakagawa [8]):* In [8], after the PM, the age of the system can be first brought to zero, but the failure intensity function will have a faster increasing trend than that of the previous maintenance cycle. This kind of imperfect preventive maintenance model is called a linear PM model according to the classification in Wu and Zuo [22] because the maintenance activities proportionally change the failure intensity function in a linear manner. The failure intensity function of the system in the $k$th PM cycle is given by

$$\lambda_k(t; a_1, \ldots, a_{k-1}) = \lambda_1(t) \prod_{i=1}^{k-1} a_i, \tag{5}$$

where $a_i \ge 1$, and $a_i = 1$ indicates perfect maintenance.

*Model IV (Nonlinear PM Model [22]):* Based on the classification in Wu and Zuo [22], the nonlinear PM model supposes that PMs only alter the aging process. In this paper, we consider the simplest nonlinear PM model as a candidate model here. The failure intensity function of the system in the $k$th PM cycle can be expressed as

$$\lambda_k(t; a_1, \ldots, a_{k-1}) = \lambda_1\left(t \prod_{i=1}^{k-1} a_i\right), \tag{6}$$

where $a_i \ge 1$, and a larger $a_i$ indicates the PM has a greater impact on the aging process. Again, $a_i = 1$ represents perfect maintenance.

### B. Parameter Estimation

Maximum Likelihood Estimation (MLE) is widely used to estimate the parameters of the imperfect maintenance models [14]–[19]. Suppose the data are collected from $N$ identical repairable systems. For the $l$th system, $n_{l,k}$ failures happen during the $k$th PM cycle, and the corresponding failure time in this PM cycle is denoted as $t_{l,(k,1)}^f, t_{l,(k,2)}^f, \ldots, t_{l,(k,n_{l,k})}^f$, where $0 < t_{l,(k,1)}^f, t_{l,(k,2)}^f, \ldots, t_{l,(k,n_{l,k})}^f \le t_k^p$. Assume that the system will be discarded or replaced by an identical new one at the end of
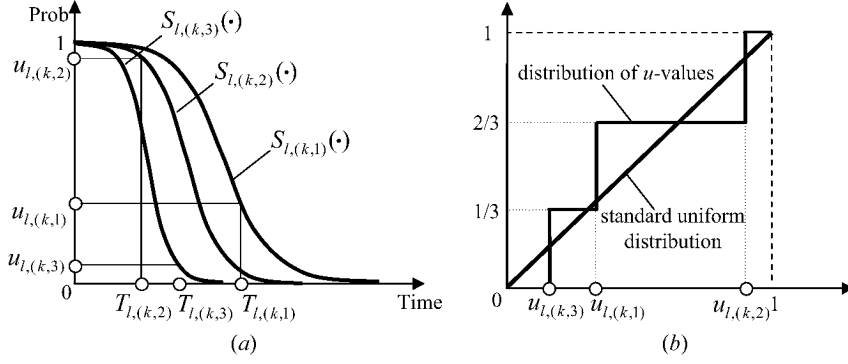
Fig. 1. Illustrations of the u-pooling method. (a) Obtain the u values according to the conditional survival distribution. (b) Pool the u values into the u-space and compare to the standard uniform distribution.

the $H$th PM action, and failures between adjacent PMs are minimally repaired. The likelihood function under PM can be expressed as [3], [16], [28]

$$
\begin{aligned}
&L(failure\ data|\boldsymbol{\theta}, \boldsymbol{a}) \\
&= \prod_{l=1}^{N} \prod_{k=1}^{H} \left( \prod_{i=1}^{n_{l,k}} \left( \lambda_k(t^f_{l,(k,i)}) \exp\left( -\int_{t^f_{l,(k,i-1)}}^{t^f_{l,(k,i)}} \lambda_k(t)dt \right) \right) \right. \\
&\left. \times \exp\left( -\int_{t^f_{l,(k,n_{l,k})}}^{t^p_k} \lambda_k(t)dt \right) \right),
\end{aligned} \quad (7)
$$

where the vector $\boldsymbol{\theta}$ represents a set of parameters in the initial failure intensity function, and it is assumed known in advance in this paper. The estimators of the unknown parameters, denoted as $\hat{\boldsymbol{a}}$, can be computed by maximizing the logarithm of (7). With the estimates of the unknown parameters, the conditional survival distribution of the $l$th system after the $(i-1)$th failure of the $k$th PM cycle can be written as

$$
S_{l,(k,i)}(t') = \exp\left( -\int_{t^f_{l,(k,i-1)}}^{t'+t^f_{l,(k,i-1)}} \lambda_k(t; \hat{\boldsymbol{a}})dt \right), \quad (8)
$$

where $t'$ represents the time elapsed since the last CM.

### III. GOODNESS-OF-FIT TEST

Given a pre-assumed imperfect maintenance model, the unknown parameters in the model can be estimated by the MLE introduced in Section II-B. The adequacy of the maintenance model, however, must be validated before using the model in decision-making. In this paper, a u-pooling method is proposed to deal with the GOF test for the imperfect maintenance model. The basic idea of expressing the conformance of theory and data on some universal scale is rooted in the practice of model validation [35]–[37]. Let $T_{l,(k,i)}$ represent the inter-arrival time between the $(i-1)$th and the $i$th failures of the $k$th PM cycle for the $l$th repairable system. By transforming every time $T_{l,(k,i)}$

according to the corresponding conditional survival distribution $S_{l,(k,i)}(\cdot)$ derived by (8), one can obtain a value $u_{l,(k,i)}$ as

$$
u_{l,(k,i)} = S_{l,(k,i)}\left(T_{l,(k,i)}\right), \quad (9)
$$

where the inter-arrival time $T_{l,(k,i)} = t^f_{l,(k,i)} - t^f_{l,(k,i-1)}$, and $u_{l,(k,i)}$ has a range of [0, 1]. An illustration with three failure data are shown in Fig. 1(a). $u_{l,(k,1)}$, $u_{l,(k,2)}$, and $u_{l,(k,3)}$ are derived based on the inter-arrival times $T_{l,(k,1)}$, $T_{l,(k,2)}$, $T_{l,(k,3)}$, and their corresponding conditional survival distributions $S_{l,(k,1)}(\cdot)$, $S_{l,(k,2)}(\cdot)$, $S_{l,(k,3)}(\cdot)$ respectively. The adequacy of the imperfect maintenance model can be assessed by pooling the various resulting u values into a u-space within a range of $[0, 1] \times [0, 1]$ as shown in Fig. 1(b). Under the assumption that all the inter-arrival times $T_{l,(k,i)}$ are distributed according to their respective conditional survival distributions $S_{l,(k,i)}(\cdot)$, all the $u_{l,(k,i)}$ constitute a standard uniform distribution. This fact is well known as the probability integral transform theorem in statistics [38]. The adequacy of the imperfect maintenance model can be validated by checking whether or not the $u_{l,(k,i)}$ are uniformly distributed.

To validate the uniformity of the u values, some well-established GOF test methods, including Kolmogorov-Smirnov (K-S), Anderson-Darling, and Cramér-von Mises, can be used here to detect evidence of significant disagreement between the distribution of u values and the standard uniform distribution. The K-S test, which is applicable to all types of distributions, is adopted in this paper, and its corresponding test statistic is defined as the maximum distance between the empirical Cumulative Density Function (CDF) of u values and a standard uniform distribution

$$
D_n^{\max} = \max_{1 \le i \le ns} \left| \hat{F}_u(u_i) - U_n(u_i) \right|, \quad (10)
$$

where $ns$ is the total amount of u values. $U_n(\cdot)$ is the CDF of a standard uniform distribution. The $p$-value, defined as the probability of obtaining a test statistic at least as extreme as the one that was actually observed, is expressed by

$$
p = \Pr\left\{ D_n^{\max} \le D_n^\alpha \right\} = 1 - \alpha. \quad (11)
$$

If $D_n^{\max}$ is smaller than a critical value $D_n^\alpha$ which can be found in the standard table [39], the null hypothesis that the u values

are consistent with the standard uniform distribution cannot be rejected; otherwise, it is rejected with $100 \times (1 - \alpha)\%$ confidence.

Instead of the K-S test, other GOF test methods, like the Anderson-Darling test which is more sensitive to the tails of a distribution, can be used to validate the uniformity of u values. The results might be slightly different from that of the K-S test because the test statistics of GOF test methods are different. However, this paper focuses on using one of the existing GOF test methods to check the uniformity of u values, and does not intend to compare the GOF test methods in terms of validating uniformity.

Through checking the uniformity of u values, the adequacy of the pre-assumed imperfect maintenance model can be validated. One of the most important merits of the u-pooling method is that it allows for validating the model adequacy by pooling all the failure data from multiple copies of the repairable system into the u-space. Another merit worth mentioning is that, by only replacing the mathematical formulation of the conditional survival distribution accordingly, the proposed GOF test can be applied to check the adequacy for any kind of imperfect maintenance model.

## IV. BAYESIAN MODEL SELECTION

The proposed GOF test in the preceding section can only answer the question of whether or not the pre-assumed imperfect maintenance model is adequate enough. However, there might be several imperfect maintenance models available to characterize the maintenance efficiency, and more than one model might not be rejected by the proposed GOF test, because they all are able to describe the observed failure data in an adequate manner. It is, therefore, necessary to develop a model selection method to help analysts choose the most adequate imperfect maintenance model among the available competitive candidates. On the other hand, the uncertainty due to the lack of sufficient failure data is inevitable in practical applications, and the conventional model selection based on asymptotic results might yield an untrustworthy conclusion when ignoring the potential uncertainty resulting from the parameters estimation [34]. The $p$-value computed from the K-S test, by itself, is not a direct quantitative measure of the probability that the null hypothesis is true, i.e. a small $p$-value does not mean that there is a small probability that the pre-assumed imperfect maintenance model is the true model [40], [41]. In addition, the uncertainty associated with parameter estimation cannot be taken into account by the $p$-value. Thereby, the $p$-value is not an appropriate measure for model selection. To overcome this issue in imperfect maintenance model selection, the Bayesian model selection method is tailored in the present paper to identify the most adequate model among candidates under the uncertainty of parameter estimation.

Suppose a finite set $\mathbf{M}$ consists of all the candidate imperfect maintenance models, e.g. $\mathbf{M} = \{M_1, M_2, \ldots, M_c\}$ ($c$ is number of candidates), which are not rejected by the proposed GOF test. The Bayesian model selection consists of defining $c$ hypotheses as follows.

$H_k$: The failure data are from the $M_k$ ($k = 1, 2, \ldots, c$) imperfect maintenance model.

Based on the basis of the Bayesian model selection method [34], the corresponding posterior probability for the hypothesis $H_k$, given the failure data set, is defined as

$$\Pr\{H_k | data\} = \frac{\Pr\{data | H_k\} \cdot \Pr\{H_k\}}{\sum\limits_{i=1}^{c} \Pr\{data | H_i\} \cdot \Pr\{H_i\}}, \quad (12)$$

where $\Pr\{H_i\}$ is the prior preference of each hypothesis. If there is no prior information, one can assign $\Pr\{H_i\} = 1/c$, meaning every candidate has an equal chance to be favored. $\Pr\{data | H_i\}$ is the likelihood function under the hypothesis $H_i$, given by

$$\Pr\{data | H_i\} = \int \Pr\{data | \hat{\boldsymbol{a}}, H_i\} \Pr\{\hat{\boldsymbol{a}} | H_i\} d\hat{\boldsymbol{a}}, \quad (13)$$

where $\Pr\{data | \hat{\boldsymbol{a}}, H_i\}$ is the likelihood function of the $i$th imperfect maintenance model with the estimated parameter $\hat{\boldsymbol{a}}$, equal to (7) if the parameter $\boldsymbol{\theta}$ of the initial failure intensity function is exactly known. $\Pr\{\hat{\boldsymbol{a}} | H_i\}$ is the distribution of estimated parameter $\hat{\boldsymbol{a}}$ of the $i$th imperfect maintenance model, and it is usually provided by analysts to quantify the uncertainty of parameter estimation due to the limited data. In the case where multiple copies of the identical repairable system exist, the bootstrap method can be used to gain the distribution of the estimator $\hat{\boldsymbol{a}}$ by using a re-sampling technique.

The bootstrap method was introduced by Efron [42] as a computational method to estimate the distribution of the estimated parameters, and its general idea is to create many sets of bootstrapping samples through re-sampling with replacement from the original data [43]. By gathering the estimators calculated from each sample set, the empirical distribution of estimators can be utilized to approximate the true distribution of estimators. A very nice property of the bootstrap method is that the distribution of the estimator can be objectively and straightforwardly determined by the available data. The basic procedures of applying the bootstrap method to derive $\Pr\{\hat{\boldsymbol{a}} | H_i\}$ are listed as follows.

Step 1: Obtain initial $N$ sets of failure data from $N$ identical repairable systems.

Step 2: Randomly select $N$ sets of failure data through re-sampling with replacement.

Step 3: Estimate the unknown parameter $\boldsymbol{a}$ of the $i$th imperfect maintenance model by the MLE method introduced in Section II-B.

Step 4: Repeat Steps 2 and 3, $p$ times (typically 500 to 5000).

Step 5: Derive the empirical distribution of estimator $\hat{\boldsymbol{a}}$ based on the $p$ sets of MLE results.

After knowing the distribution of the estimated parameter $\hat{\boldsymbol{a}}$, the posterior probability of the hypothesis $H_k$ can be computed using (12). The imperfect maintenance model which has the highest posterior probability among the candidates is selected as the most adequate model to describe the failure data.

In the case of one set of data from one copy of a repairable system, deriving $\Pr\{\hat{\boldsymbol{a}} | H_i\}$ is not straightforward. Doyen [44] developed the asymptotic confidence intervals for two imperfect repair models called the Arithmetic Reduction of Intensity model (ARI), and the Arithmetic Reduction of Age model

(ARA). Nevertheless, very limited work has been found to deal with the confidence interval of the estimated parameters for many other imperfect maintenance models [45], especially the observations from only one repairable system. This is an important topic to be explored in our future work.

## V. IMPLEMENTATION IN MAINTENANCE DECISION

In this section, we form a periodic PM policy to demonstrate how to incorporate the conclusion from imperfect maintenance model selection into the maintenance decision framework considering the uncertainty due to the parameter estimation and model selection. The basic assumptions are listed as follows.

1. The failure intensity function has a strictly increasing trend if no PM intervention is involved.
2. PM activities are executed at fixed time interval $T_{PM}$.
3. The PM activities are imperfect, whereas the minimal repairs are performed for the failures between adjacent PMs.
4. The system is replaced by a new, identical one at time $N_{PM} \cdot T_{PM}$.
5. Compared to the operation time, the time for PM, minimal repair, and replacement are negligible.
6. The planning horizon is infinite.
7. The objective is to minimize the expected average maintenance cost per unit time.

Due to the limited data, the uncertainty associated with the parameter estimation and model selection need to be quantified in the maintenance decision model. Let $\mathbf{M} = \{M_1, M_2, \ldots, M_c\}$ represent the imperfect maintenance models that are not rejected by the proposed GOF test, and $\mathbf{p}^{pst} = \{p_1^{pst}, p_2^{pst} \ldots, p_c^{pst}\}$ denote the posterior probability computed by (12). If $p_l^{pst} \gg p_i^{pst}$ $i \in \{1, 2, \ldots, c\}, i \neq l$), one of the models greatly dominates the others. The expected average maintenance cost per unit time can be formulated as, (14), shown at the bottom of the page, where $\lambda_k^l(t, \hat{\mathbf{a}})$ represents the failure intensity function in the $k$th PM cycle derived from the $l$th imperfect maintenance model. The uncertainty of estimated parameters is quantified by $\Pr\{\hat{\mathbf{a}}|H_i\}$ from the proposed bootstrap method. Hence, $C(T_{PM}, N_{PM})$ is a random quantity. If $p_l^{pst} \gg p_i^{pst}$ doesn't hold, which means that at least two candidate models are competitive, it is necessary to account for the uncertainty related to the model selection.

Thereby, the expected average maintenance cost per unit time can be written as (15), shown at the bottom of the page, where $C(T_{PM}, N_{PM})$ has uncertainty propagated from the uncertainty of both estimated parameters, and model selection. In the circumstance where $C(T_{PM}, N_{PM})$ is a random quantity, the optimal PM policy could be chosen from various perspectives. Three decision-making criterions, presented next, can be used.

1  ***The expectation criterion***

$$C^*(T_{PM}^*, N_{PM}^*) = \min\{E[C(T_{PM}, N_{PM})]\}. \quad (16)$$

Thereby, the PM policy with minimal expectation will be chosen as the optimal policy.

2  ***The robustness criterion***

$$C^*(T_{PM}^*, N_{PM}^*) = \min\left\{w \cdot \frac{E[C(T_{PM}, N_{PM})]}{E^*} + (1-w) \cdot \frac{\sigma[C(T_{PM}, N_{PM})]}{\sigma^*}\right\}. \quad (17)$$

$E^*$, and $\sigma^*$ correspond to the minimal values of the mean, and the standard deviation of random quantity $C(T_{PM}, N_{PM})$ respectively. $w$, pre-determined by the decision maker, is the relative weight between the mean and the standard deviation. In this criterion, the decision maker is seeking a compromise between the mean performance and the robustness. The smaller $w$, the more important is the robustness.

3  ***The confidence criterion***

$$C^*(T_{PM}^*, N_{PM}^*) = \min\{C_\alpha(T_{PM}, N_{PM})\}, \quad (18)$$

where $C_\alpha(T_{PM}, N_{PM})$ represents the $\alpha \times 100\%$ one-sided confidence bound of the random quantity $C(T_{PM}, N_{PM})$. For example, if $\alpha = 0.9$, we are to seek the optimal solution whose 90% one-sided confidence bound is minimal. The optimal solution can guarantee that 90% of the expected average maintenance cost is lower than the minimal value $C^*(T_{PM}^*, N_{PM}^*)$ when uncertainties are present in both estimated parameters and model selection.

Note that, instead of the aforementioned three criterions, other case-dependent criterions can be defined to identify the most suitable policy considering the uncertainty of parameter estimation and model selection.

$$C(T_{PM}, N_{PM}) = \frac{c_m \left(\sum_{k=1}^{N_{PM}} \int_0^{T_{PM}} \lambda_k^l(t; \hat{\mathbf{a}})dt\right) \cdot \Pr\{\hat{\mathbf{a}}|H_i\} + c_p(N_{PM} - 1) + c_r}{N_{PM} \cdot T_{PM}}, \quad (14)$$

$$C(T_{PM}, N_{PM}) = \frac{c_m \sum_{i=1}^{c} p_i^{pst} \cdot \left(\sum_{k=1}^{N_{PM}} \int_0^{T_{PM}} \lambda_k^i(t; \hat{\mathbf{a}})dt\right) \cdot \Pr\{\hat{\mathbf{a}}|H_i\} + c_p(N_{PM} - 1) + c_r}{N_{PM} \cdot T_{PM}}, \quad (15)$$

| System ID $l$ | $t^f_{l,(2,i)}$ (The 2$^{nd}$ PM Cycle) | $t^f_{l,(3,i)}$ (The 3$^{rd}$ PM Cycle) |
|---|---|---|
| 1 | 1.68, 18.05, 18.80, 39.22, 62.00 | 23.88, 32.76, 46.69, 53.73, 66.16 |
| 2 | 2.95, 9.78, 20.98, 31.79, 35.18 | 39.50, 39.76, 41.65, 48.71, 50.62, 55.66, 55.85, 67.22 |
| 3 | 54.28, 59.57 | 17.74, 30.58, 30.98, 45.16, 55.63, 62.18 |
| 4 | 23.65, 32.64, 66.60 | 21.38, 21.54, 30.26, 52.99, 68.82 |
| 5 | 10.76, 11.19, 33.74, 40.73, 52.03, 66.38, 69.33 | 9.15, 14.85, 47.80, 54.69, 59.98, 62.68 |

## VI. CASE STUDIES

*Designed Test Problem I:* We use a designed test problem to demonstrate the effectiveness of the proposed GOF test, and Bayesian model selection method. The initial failure intensity of the repairable system is assumed to follow a log-linear process

$$\lambda_1(t) = e^{\alpha + \beta t},$$

where $\alpha = -4.0$, and $\beta = 0.015$. The four imperfect preventive maintenance models introduced in Section II-A are used here as the candidate models. We assume that the failure data sets are collected from five copies of identical repairable systems which are governed by the imperfect maintenance Model I. For the purpose of estimating the unknown parameters, and selecting a model rather than optimizing the interval of periodic PM, the failure data sets are artificially generated with the pre-specified setting that the PM is periodically executed with time interval equal to 70 hours. The system will be discarded at the end of the third PM cycle. Each PM is imperfect, with identical parameter value $a_i = 0.5$. The failures between any two consecutive PMs are minimally repaired. An example of the artificially generated failure data set is tabulated in Table I. Because the imperfect PM activities are not involved in the first PM cycle, the failure data in the first PM cycle do not have any contribution to the parameter estimation of the imperfect PM model. Table I lists 52 failure data observed in the second and third PM cycles, and these failure data will be used in the proposed method to estimate the model parameter.

To apply the proposed GOF test to examine the adequacy of the models, the unknown parameters of the pre-assumed imperfect maintenance models are first estimated via MLE method. Based on the estimated imperfect maintenance model, all the failure data are pooled into the u-space according to the corresponding conditional survival density. Fig. 2 plots the empirical distributions of u values under different pre-assumed imperfect maintenance models (the stepwise line) versus the standard uniform distribution (the diagonal line). The empirical distributions of u values from Model I, Model II, and Model III are closer to the standard uniform distribution than that of Model IV, indicating Model I, Model II, and Model III are more adequate than Model IV. By applying the K-S test, we conclude that Model IV cannot pass the test under a 95% confidence level setting. The corresponding $p$-values of the K-S test are also in Fig. 2.

The most adequate model among the possible candidates (Model I, Model II, and Model III) can be further identified by the proposed Bayesian model selection method. Without any knowledge, the prior preference $\Pr\{H_i\}$ in (12) is set to 1/3 for all three potential models, i.e. $\Pr\{H_1\} = \Pr\{H_2\} = \Pr\{H_3\} = 1/3$. The distributions of the estimator $\hat{a}_i$ of Model I, Model II, and Model III are gained by using the proposed bootstrap method with $p = 500$ (See Fig. 3). The posterior probabilities computed using (12) are 39.5%, 30.0%, and 30.5% for Model I, Model II, and Model III, respectively. Hence, Model I is identified as the most adequate model.

The results from model selection are further considered in the periodic PM policy by (15). The estimated parameter in the imperfect maintenance model is treated as a random quantity with the distribution from the bootstrap method. Let $c_m = 2.0$, $c_p = 8.0$, and $c_r = 25.0$. The optimal periodic PM policies under the three proposed criterions are listed in Table II.

As one observes from Table II, the optimal periodic PM policies for criterions 1 and 2 are the same, but slightly different from that of criterion 3. When one sets $N_{PM} = 2$, the expected average maintenance cost $C(T_{PM}, N_{PM} = 2)$ with respect to different settings of the periodic PM interval $T_{PM}$ are plotted in Fig. 4. Because the true value of the parameter in the imperfect maintenance model where the failure data are generated is known, the corresponding $C(T_{PM}, N_{PM} = 2)$ can be used as a benchmark, and denoted by the black solid line as shown in Fig. 4. Due to the uncertainty from parameter estimation, $C(T_{PM}, N_{PM} = 2)$ under different assumptions of underlying imperfect maintenance models is also represented by a random quantity. The 90% confidence bounds (CB) are indicated by the dashed lines, and means are denoted by the solid lines with distinct markers. As shown in Fig. 4, if one assumes the imperfect maintenance model is Model II or Model III, the corresponding $C(T_{PM}, N_{PM} = 2)$ is far away from the true value, and such discrepancy increases as the periodic PM interval $T_{PM}$ increases. The 90% confidence bounds of $C(T_{PM}, N_{PM} = 2)$ from Model I contain the true values. However, compared to the results computed by the proposed maintenance decision model (15), the proposed model is slightly closer to the true values, but also has narrower uncertainty bounds. Therefore, considering the uncertainty of imperfect maintenance model selection in the maintenance decision will reduce the potential risk of specifying an imperfect maintenance model without model selection, es-
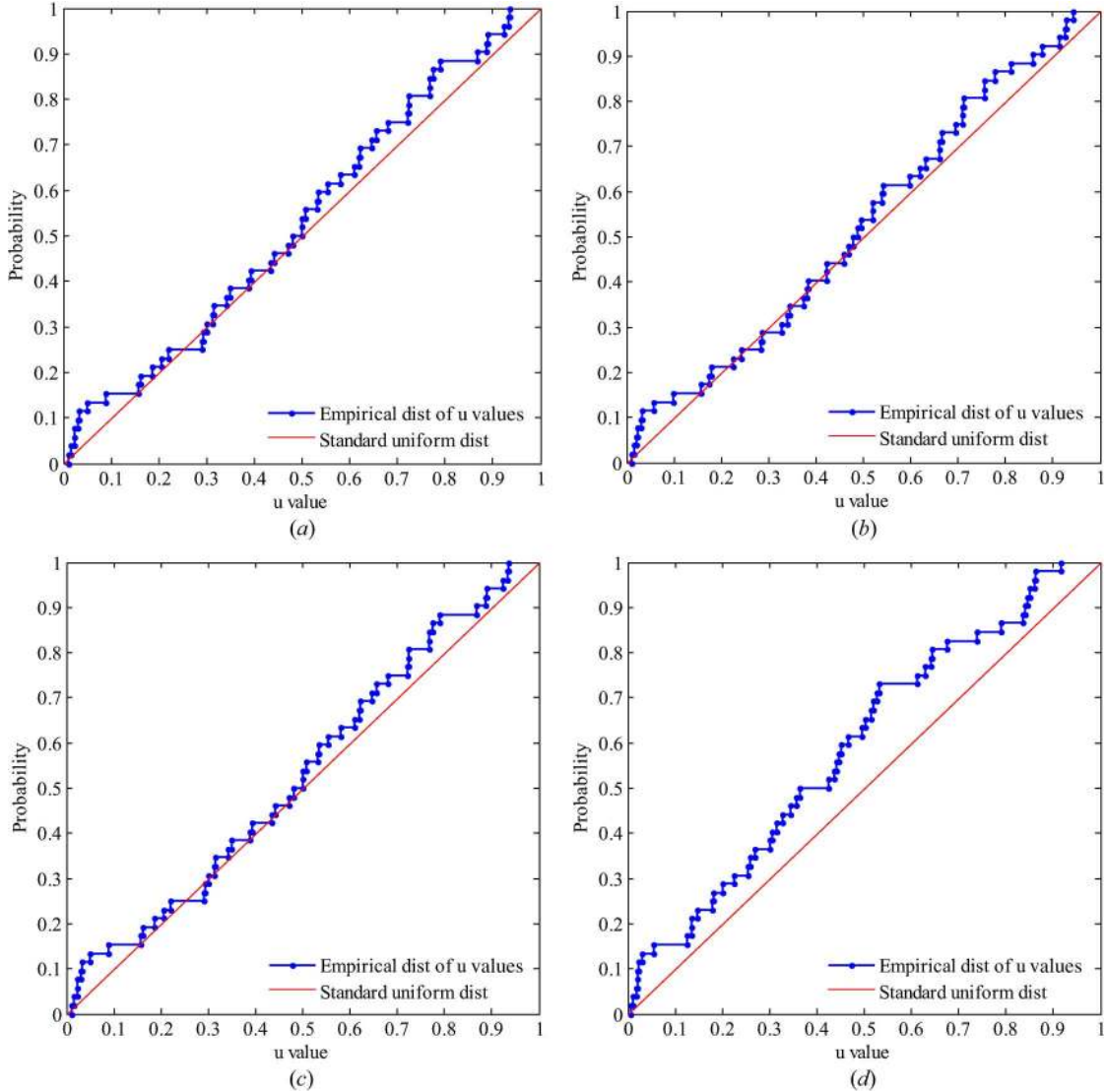
Fig. 2.   Empirical distribution of u values vs. the standard uniform distribution. (a) Model I ($\hat{a}_i = 0.49$ and $p-\text{value} = 0.72$). (b) Model II ($\hat{a}_i = 0.59$ and $p-\text{value} = 0.70$). (c) Model III ($\hat{a}_i = 1.67$ and $p-\text{value} = 0.71$). (d) Model IV ($\hat{a}_i = 1.53$ and $p-\text{value} = 0.03$).

pecially in the case where the pre-determined imperfect maintenance model is not the true model.

*Designed Test Problem II:* To further demonstrate the benefit of the proposed models in maintenance decisions when none of the candidate models is the true model, we design test problem II in which the failure data sets are artificially generated using hybrid imperfect maintenance models [9]. The hybrid model combines the concepts from both Model I and Model III, and the failure intensity function of the $k$th PM cycle is defined as

$$\lambda_k(t; a_1, \ldots, a_{k-1}, b_1, \ldots, b_{k-1}) = \prod_{i=1}^{k-1} b_i \cdot \lambda_1 \left( t + \sum_{i=1}^{k-1} a_i t_i^p \right)$$

where $0 \leq a_i \leq 1$, and $b_i \geq 1$. The PMs not only reduce the system's virtual age, but also change the slope of the failure intensity function. The failure data sets are collected from ten copies of identical, repairable systems with the parameters of imperfect preventive maintenance equal to 0.25, and 1.3 for $a_i$, and $b_i$, respectively; and the parameters of the initial failure

intensity, as well as the pre-specified interval of periodic PM, are the same as those used in the designed test problem I.

By using the proposed GOF test, it has been found that Model I, Model II, and Model III are adequate enough to characterize the failure date sets. The posterior probability of these three candidate models can be computed via the proposed Bayesian model selection method, and the results are 1%, 20%, and 79% for Model I, Model II, and Model III, respectively. Knowing the true values of the parameters of the hybrid imperfect maintenance model, the optimal periodic PM policy can be solved, and it is $T_{PM}^* = 107$, $N_{PM}^* = 2$. The corresponding expected average maintenance cost per unit time is plotted in Fig. 5. Even if one chooses a model among Model I, Model II, and Model III, and considers the uncertainty associated with parameter estimations, the corresponding values of the expected average maintenance cost per unit time are quite different from the true values, as observed from Fig. 5. Model III is very slightly closer to the true values than Model II, but also has lower uncertainty bound. Although the expected average maintenance cost per unit time
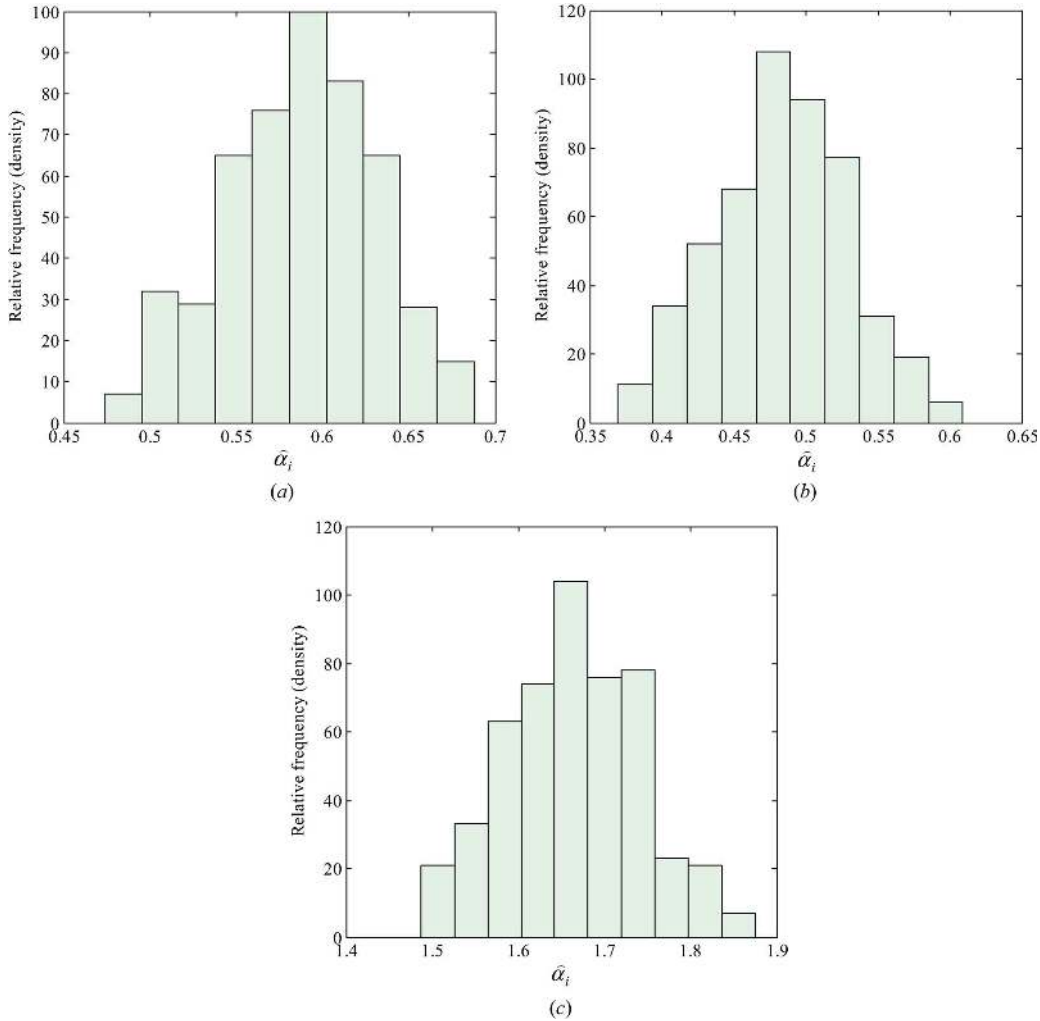
Fig. 3. Empirical distributions of the estimator $\hat{a}_i$ for candidate models. (a) $\Pr\{\hat{a}_i|H_1\}$ (Model I). (b) $\Pr\{\hat{a}_i|H_2\}$ (Model II). (c) $\Pr\{\hat{a}_i|H_3\}$ (Model III).

TABLE II
OPTIMAL PERIODIC PM POLICIES UNDER THE PROPOSED CRITERIONS

| Criterion | $N_{PM}^*$ | $T_{PM}^*$ | $C^*(T_{PM}^*, N_{PM}^*)$ |
|---|---|---|---|
| (1) Expectation | 2 | 102 | 0.3210 |
| (2) Robustness | 2 | 102 | 0.5316 |
| (3) Confidence | 2 | 100 | 0.3259 |

of the proposed maintenance decision model (15) has some discrepancy with the true values, it is the most accurate.

Because the failure data sets are not generated from the four candidate imperfect maintenance models, it is impossible to find one model that has exactly the same failure behavior with the true model. However, based on the proposed model selection method, Model III is identified as the most accurate model among the others because it has the largest posterior probability. Considering the uncertainty from model selection, the proposed maintenance decision model can further provide more accurate results than Model III because the error from selecting an incorrect model can be reduced.

*Designed Test Problem III:* To study how sensitive the model selection approach is with respect to the amount of data, the
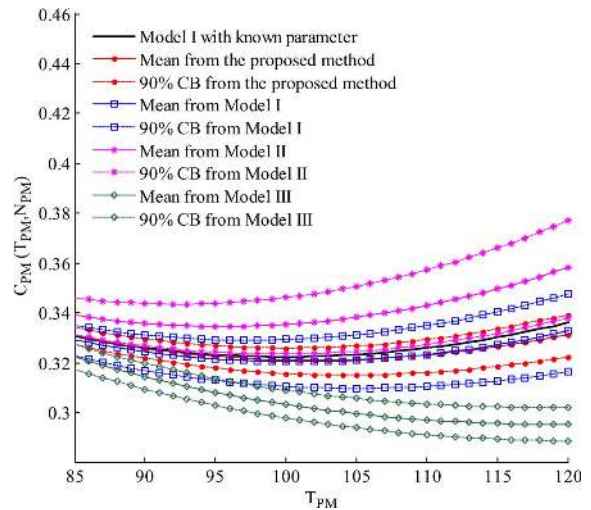


Fig. 4. Expected average maintenance cost per unit time with respect to $T_{PM}$ when $N_{PM} = 2$.

number of identical repairable systems is specified to 10, 20, 40, and 60, respectively. The parameters of the initial failure intensity, as well as the pre-specified interval of periodic PM, are the
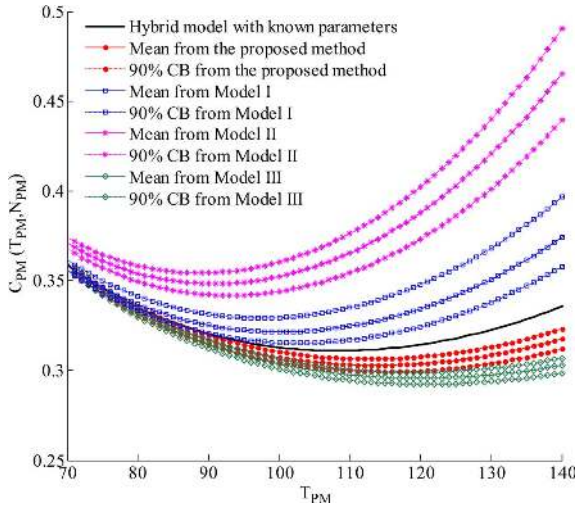
Fig. 5.   Expected average maintenance cost per unit time with respect to $T_{PM}$ in the case of the hybrid imperfect model and $N_{PM} = 2$.

TABLE III
VALUES OF THE PARAMETERS OF THE ORIGINAL
IMPERFECT MAINTENANCE MODELS

|  | Model I | Model II | Model III | Model IV |
|---|---|---|---|---|
| Low effect $a_i$ | 0.8 | 0.8 | 1.1 | 1.1 |
| High effect $a_i$ | 0.3 | 0.3 | 1.6 | 1.6 |

same as those used in the designed test problem I. The failure data set of each individual system is randomly generated by assuming the system follows one of the imperfect maintenance models (called the original model in the following context) with pre-set values for the model parameter as listed in Table III. The "low effect" row in Table III is the situation where the preventive maintenance has a lower impact on the failure intensity function, whereas the "high effect" row corresponds to the case having a larger impact. Due to the randomness of the failure data, the statistical performance of the proposed GOF test method and Bayesian model selection method is examined by the sum of $p$-values computed from the proposed GOF test, and the sum of the posterior probabilities of the candidate models over 100 trials of simulation. The proposed method should likely find the original model from which the failure data are artificially generated.

The sum of the $p$-values computed from the proposed GOF test is shown in Fig. 6, and a greater value indicates that the imperfect maintenance model has a lower chance to be rejected in the proposed GOF test. The original imperfect maintenance model where the failure data sets are artificially generated is indicated by the square brackets. In our study, both Model I and Model II are capable of representing the failure data generated from the other models, and in a majority of cases the sum of $p$-values of these two models are greater than 50.0. Most of the time, Model III and Model IV are rejected by the GOF test when the failure data are from Model I and Model II, and the corresponding $p$-values are very small (less than 5.0), as shown in Fig. 6. However, when the original model is Model III or Model IV, all the candidate models have a large chance not to be rejected by the GOF test, and the sum of the $p$-values from the four
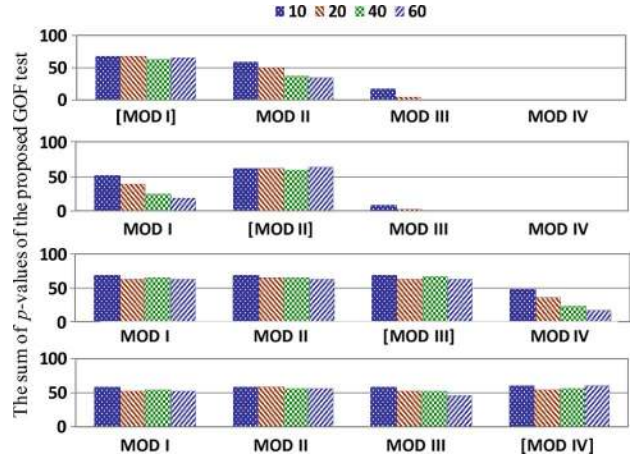


Fig. 6.   Sum of $p$-values of the proposal GOF test in the lower effect case.
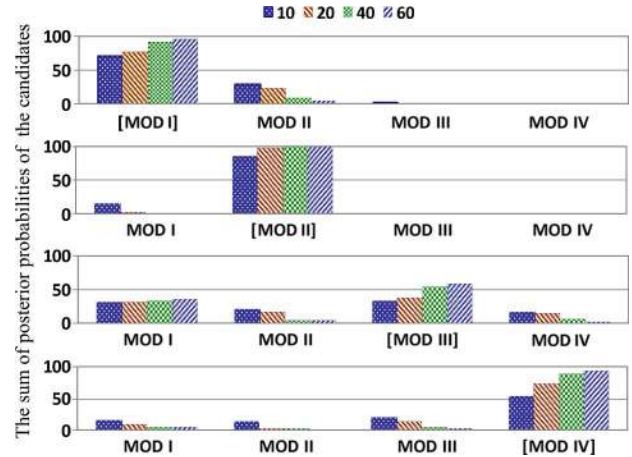


Fig. 7.   Sum of posterior probabilities of the candidate models in the lower effect case.

candidate models are very high (greater than 50.0). Thereby, it is necessary to identify the most adequate model by the proposed Bayesian model selection method.

The sum of the posterior probabilities of the candidate models over 100 trials are shown in Fig. 7. Apparently, the original model has the highest chance to be identified via the proposed method. Only when the original model is Model III, Model I will become a competitive model with Model III if the data are very limited. However, by adding more failure data, the chance of identifying Model I as the original model decreases.

The same testing is also performed on the scenario where the impact from imperfect maintenance is large. Both Model I and Model II have a large chance not to be rejected by the GOF test even if the failure data come from the other models, indicating Model I and Model II have a greater capability to fit failure data than the other two models. However, by comparing Fig. 8 with Fig. 6, we find that adding more failure data will reduce the sum of $p$-values of Model II when the failure date originated from the other models. Again, if the failure data come from Model IV, most of the time, the other three models are adequate to describe the failure data. After applying the proposed Bayesian model selection method, the most adequate model can be clearly identified even if limited data are available (only 10 sets of failure
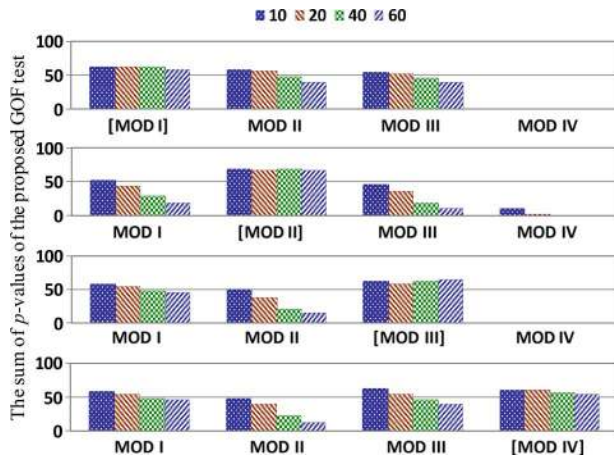
Fig. 8. Sum of $p$-values of the proposal GOF test in the higher effect case.
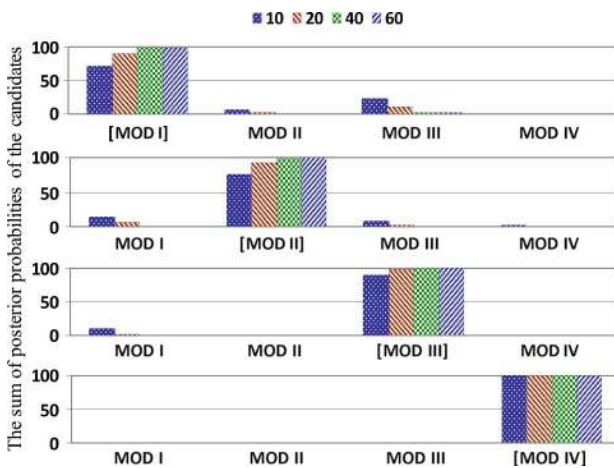


Fig. 9. Sum of posterior probabilities of the candidate models in the higher effect case.

data), as shown in Fig. 9. By comparing Fig. 9 with Fig. 7, one can also conclude that the posterior probabilities of Model III and Model IV will be increased when the failure data are originally from Model III and Model IV with higher maintenance efficiencies.

## VII. CONCLUSION AND DISCUSSION

In this paper, a data-driven approach for imperfect maintenance selection is proposed to address two important issues in choosing the imperfect maintenance models, especially when the knowledge of potential physics of failure, and the mechanism of improvement from maintenance, are absent. To examine whether the pre-assumed imperfect maintenance model is adequate to characterize the stochastic failure behavior, a novel GOF test is proposed by pooling all the failure data into a u-space via the conditional survival distribution. The validity of the imperfect maintenance model is examined by checking the uniformity of the resulting u values. One of the advantages of the u-pool method is that it allows validating all the failure data sets from multiple copies of identical repairable systems. In the case where more than one candidate model is valid, a Bayesian model selection method is proposed to identify the most adequate imperfect maintenance

model among the candidates. The associated uncertainty of parameter estimation is quantified via the bootstrap method. The periodic PM policy optimization problem is presented to demonstrate how to apply the results from model selection to the maintenance planning, and three decision-making criterions are introduced to account for the uncertainty associated with the parameter estimation and model selection. Three designed test problems are presented to demonstrate the effectiveness of the proposed methods. The correct model from which the artificial failure data are generated has the largest chance to be identified from the competitive candidates by the proposed model selection approach even with limited data, and such identification capability will become greater with the increase of the amount of data. Meanwhile, the proposed method can better identify the correct model if maintenance activities have high effect on the system deterioration process. On the other hand, by incorporating the uncertainty of model selection into the maintenance decision model, we find that the accuracy of computing the quantity of interest can be further improved due to the reduction of the error of model selection.

Although we only applied the proposed method to the circumstance where PMs are imperfect, and the failed systems are minimally repaired, the proposed methods are applicable to the case in which corrective and preventive maintenance are both imperfect. In addition, the candidate imperfect maintenance models are not limited to just the four models illustrated in this paper. However, in the case of no failure data, the proposed method is not applicable because selecting imperfect maintenance purely depends on the available failure data. It is therefore necessary to incorporate information from experts' judgment to facilitate model selection. On the other hand, in the present paper, it is assumed that all the imperfect maintenance models have only one parameter to be estimated, and the efficiency of each PM is equal. We will explore more complicated situations in our future work, like the case where the imperfect maintenance model has more than one parameter, or the improvement from maintenance may vary within the lifecycle. In addition, the identifiability issue might arise if both the parameters of the initial failure intensity and the candidate imperfect maintenance models are unknown. That is a challenging research issue worth exploring in the future. Finally, this paper supposes that the data sets come from multiple identical systems. It is also necessary to develop methods to quantify the uncertainty associated with parameter estimation when only one system is observed, but the time of observation could be very long.

## REFERENCES

[1] S. Wu, "Preventive maintenance models: A review," in *Replacement Models with Minimal Repair*. London: Springer, 2011.

[2] H. Pham and H. Z. Wang, "Imperfect maintenance," *European Journal of Operational Research*, vol. 94, no. 3, pp. 425–438, 1996.

[3] M. Yañez, F. Joglar, and M. Modarres, "Generalized renewal process for analysis of repairable systems with limited failure experience," *Reliability Engineering and System Safety*, vol. 77, no. 2, pp. 167–180, 2002.

[4] M. Brown and F. Proschan, "Imperfect repair," *Journal of Applied Probability*, vol. 20, no. 4, pp. 851–859, 1983.

[5] M. Kijima, H. Morimura, and Y. Suzuki, "Periodical replacement problem without assuming minimal repair," *European Journal of Operational Research*, vol. 37, no. 2, pp. 194–203, 1988.

[6] M. Kijima, "Some results for repairable systems with general repair," *Journal of Applied Probability*, vol. 26, no. 1, pp. 89–102, 1989.

[7] M. A. K. Malik, "Reliable preventive maintenance policy," *AIIE Transactions*, vol. 11, no. 3, pp. 221–228, 1979.

[8] T. Nakagawa, "Sequential imperfect preventive maintenance policies," *IEEE Trans. Rel.*, vol. 37, no. 3, pp. 295–298, 1988.

[9] D. Lin, M. J. Zuo, and R. C. M. Yam, "General sequential imperfect preventive maintenance models," *International Journal of Reliability, Quality and Safety Engineering*, vol. 7, no. 3, pp. 253–266, 2000.

[10] Y. Lam, "A note on the optimal replacement problem," *Advances in Applied Probability*, vol. 20, no. 2, pp. 479–482, 1988.

[11] H. Z. Wang and H. Pham, "A quasi renewal process and its application in the imperfect maintenance," *International Journal of Systems Science*, vol. 27, no. 10, pp. 1055–1062, 1996.

[12] H. Z. Wang and H. Pham, *Reliability and Optimal Maintenance*. London: Springer, 2006.

[13] M. J. Kallen, "Modelling imperfect maintenance and the reliability of complex systems using superposed renewal processes," *Reliability Engineering and System Safety*, vol. 96, no. 6, pp. 636–641, 2011.

[14] S. Gasmi, C. E. Love, and W. Kahle, "A general repair, proportional-hazards, framework to model complex repairable systems," *IEEE Trans. Rel.*, vol. 52, no. 1, pp. 26–32, 2003.

[15] L. Doyen and O. Gaudoin, "Classes of imperfect repair models based on reduction of failure intensity or virtual age," *Reliability Engineering and System Safety*, vol. 84, no. 1, pp. 45–56, 2004.

[16] R. Calabria and G. Pulcini, "Inference and test in modeling the failure/repair process of repairable mechanical equipments," *Reliability Engineering and System Safety*, vol. 67, no. 1, pp. 41–53, 2000.

[17] A. Mettas and W. Zhao, "Modeling and analysis of repairable systems with general repair," in *Proceedings of Annual Reliability and Maintainability Symposium*, 2005, pp. 176–182.

[18] M. Y. You, H. Li, and G. Meng, "Control-limit preventive maintenance policies for components subject to imperfect preventive maintenance and variable operational conditions," *Reliability Engineering and System Safety*, vol. 96, no. 5, pp. 590–598, 2011.

[19] T. J. Lim, "Estimating system reliability with fully masked data under Brown-Proschan imperfect repair model," *Reliability Engineering and System Safety*, vol. 59, no. 3, pp. 277–289, 1998.

[20] J. H. Lim, K. L. Lu, and D. H. Park, "Bayesian imperfect repair model," *Communications in Statistics-Theory and Methods*, vol. 27, no. 4, pp. 965–984, 1998.

[21] L. Doyen, "On the Brown–Proschan model when repair effects are unknown," *Applied Stochastic Models in Business and Industry*, 2011, DOI: 10.1002/asmb.869.

[22] S. Wu and M. J. Zuo, "Linear and nonlinear preventive maintenance models," *IEEE Trans. Rel.*, vol. 59, no. 1, pp. 242–249, 2010.

[23] H. W. Block, W. S. Borges, and T. H. Savits, "Age dependent minimal repair," *Journal of Applied Probability*, vol. 22, no. 2, pp. 370–385, 1985.

[24] V. Makis and A. K. S. Jardine, "Optimal replacement policy for a general model with imperfect repair," *Journal of the Operational Research Society*, vol. 42, no. 2, pp. 111–120, 1992.

[25] C. M. Tan and N. Raghavan, "A framework to practical predictive maintenance modeling for multi-state systems," *Reliability Engineering and System Safety*, vol. 93, no. 8, pp. 1138–1150, 2008.

[26] Y. Liu and H. Z. Huang, "Optimal selective maintenance strategy for multi-state systems under imperfect maintenance," *IEEE Trans. Rel.*, vol. 59, no. 2, pp. 356–367, 2010.

[27] Y. Liu and H. Z. Huang, "Optimal replacement policy for multi-state system under imperfect maintenance," *IEEE Trans. Rel.*, vol. 59, no. 3, pp. 483–495, 2010.

[28] S. E. Rigdon and A. P. Basu, *Statistical Methods for the Reliability of Repairable Systems*. New York: Wiley, 2000.

[29] B. H. Lindqvist, "On the statistical modeling and analysis of repairable systems," *Stat Sci*, vol. 21, no. 4, pp. 532–551, 2006.

[30] G. Pulcini, "On the overhaul effect for repairable mechanical units: A Bayes approach," *Reliability Engineering and System Safety*, vol. 70, no. 1, pp. 85–94, 2000.

[31] G. Pulcini, "Modeling the failure data of a repairable equipment with bathtub type failure intensity," *Reliability Engineering and System Safety*, vol. 71, no. 2, pp. 209–218, 2001.

[32] W. J. Park and Y. G. Kim, "Goodness-of-fit tests for the power-law process," *IEEE Trans. Rel.*, vol. 41, no. 1, pp. 107–111, 1992.

[33] M. Z. N. Agustin and E. A. Pena, "Order statistics properties, random generation, and goodness-of-fit testing for a minimal repair model," *Journal of American Statistical Association*, vol. 94, no. 445, pp. 266–272, 1999.

[34] G. Pulcini, "Accounting for model uncertainty in reliability analysis of mechanical repairable equipment," *Communications in Statistics-Theory and Methods*, vol. 31, no. 5, pp. 849–869, 2002.

[35] S. Ferson, W. L. Oberkampf, and L. Ginzburg, "Model validation and predictive capability for the thermal challenge problem," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29–32, pp. 2408–2430, 2008.

[36] Y. Xiong, W. Chen, K. L. Tsui, and D. W. Apley, "A better understanding of model updating strategies in validating engineering models," *Computer Methods in Applied Mechanics and Engineering*, vol. 198, no. 15/16, pp. 1327–1337, 2009.

[37] Y. Liu, W. Chen, P. Arendt, and H. Z. Huang, "Toward a better understanding of model validation metrics," *Journal of Mechanical Design, Transaction of the ASME*, vol. 133, no. 7, p. 071005, 2011.

[38] J. E. Angus, "The probability integral transform and related results," *SIAM Review*, vol. 36, no. 4, pp. 652–654, 1994.

[39] A. Haldar and S. Mahadevan, *Probability, Reliability, and Statistical Methods In Engineering Design*. New York: Wiley, 2000.

[40] R. Rebba and S. Mahadevan, "Computational methods for model reliability assessment," *Reliability Engineering and System Safety*, vol. 93, no. 8, pp. 1197–1207, 2008.

[41] D. H. Johnson, "Insignificance of statistical significance testing," *Journal of Wildlife Management*, vol. 63, no. 3, pp. 763–772, 1999.

[42] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[43] V. Picheny, N. H. Kim, and R. T. Haftka, "Application of bootstrap method in conservative estimation of reliability with limited samples," *Structural and Multidisciplinary Optimization*, vol. 41, no. 2, pp. 205–217, 2010.

[44] L. Doyen, "Asymptotic properties of imperfect repair models and estimation of repair efficiency," *Naval Research Logistics*, vol. 57, no. 3, pp. 296–307, 2010.

[45] H. R. Guo, H. Liao, W. Zhao, and A. Mettas, "A new stochastic model for systems under general repairs," *IEEE Trans. Rel.*, vol. 56, no. 1, pp. 40–49, 2007.

**Yu Liu** is an Associate Professor of the School of Mechanical, Electronic, and Industrial Engineering, University of Electronic Science and Technology of China. He received his Ph.D. degree in Mechanical Engineering from the University of Electronic Science and Technology of China in 2010. He was a visiting pre-doctoral fellow in the Department of Mechanical Engineering at Northwestern University, Evanston, U.S.A. from 2008 to 2010. His research interests include design under uncertainty, reliability of multi-state systems, maintenance decisions, and optimization.

**Hong-Zhong Huang** (M'06) is a Professor, and the Dean of the School of Mechanical, Electronic, and Industrial Engineering, University of Electronic Science and Technology of China. He has held visiting appointments at several universities in the USA, Canada, and Asia. He received a Ph.D. degree in Reliability Engineering from Shanghai Jiaotong University, China, and has published 150 journal papers and 5 books in the fields of reliability engineering, optimization design, fuzzy set theory, and product development. He is a Fellow of ISEAM (International Society of Engineering Asset Management), and a member of ESRA (European Safety and Reliability Association) Technical Committee on System Reliability, a Regional Editor of International Journal of Reliability and Applications, an Editorial Board Member of International Journal of Reliability, Quality and Safety Engineering, International Journal of Quality, Statistics, and Reliability, and International Journal of Reliability and Quality Performance. He received the William A. J. Golomski Award from the Institute of Industrial Engineers in 2006, and the Best Paper Award of the 8th International Conference on Frontiers of Design and Manufacturing in 2008. His current research interests include system reliability analysis, warranty, maintenance planning and optimization, and computational intelligence in product design.

**Xiaoling Zhang** is currently pursuing a Ph.D. degree in the School of Mechanical, Electronic, and Industrial Engineering at the University of Electronic Science and Technology of China. Her research interests focus on multilevel system design optimization under uncertainty.