IISE Transactions

Quality & Reliability Engineering

IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach

Changyue Song & Kaibo Liu

To cite this article: Changyue Song & Kaibo Liu (2018) Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach, IISE Transactions, 50:10, 853-867, DOI: 10.1080/24725854.2018.1440673

To link to this article: https://doi.org/10.1080/24725854.2018.1440673

Published online: 17 May 2018.

IISE TRANSACTIONS
2018, VOL. 50, NO. 10, 853–867
https://doi.org/10.1080/24725854.2018.1440673

Taylor & Francis
Taylor & Francis Group

Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach

Changyue Song and Kaibo Liu

Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA

ABSTRACT

Nowadays multiple sensors are widely used to simultaneously monitor the degradation status of a unit. Because those sensor signals are often correlated and measure different characteristics of the same unit, effective fusion of such a diverse "gene pool" is an important step to better understanding the degradation process and producing a more accurate prediction of the remaining useful life. To address this issue, this article proposes a novel data fusion method that constructs a composite Health Index (HI) via the combination of multiple sensor signals for better characterizing the degradation process. In particular, we formulate the problem as indirect supervised learning and leverage the quantile regression to derive the optimal fusion coefficient. In this way, the prognostic performance of the proposed method is guaranteed. To the best of our knowledge, this is the first article that provides the theoretical analysis of the data fusion method for degradation modeling and prognostics. Simulation studies are conducted to evaluate the proposed method in different scenarios. A case study on the degradation of aircraft engines is also performed, which shows the superior performance of our method over existing HI-based methods.

ARTICLE HISTORY
Received 13 August 2017
Accepted 4 February 2018

KEYWORDS
Indirect supervised learning; multisensor fusion; quantile regression

1. Introduction

Sensors have been developed that are now widely used to monitor the degradation of critical units such as machines, automotive batteries, and aircraft engines. The collected sensor signals contain useful information with respect to the underlying degradation process, and thus modeling and analyzing the sensor signals play an important role in estimating the health status and preventing unexpected failures of units (Nelson, 1990; Meeker and Escobar, 1998). Although the degradation modeling of a single signal from a single sensor has received considerable attention in the literature (Jardine *et al.*, 2006; Si *et al.*, 2011; Ye and Xie, 2015), fewer studies have focused on the modeling of multiple signals from multiple sensors. The need to consider multiple signals from multiple sensors arises when a single sensor signal cannot fully characterize the complex and stochastic nature of the underlying degradation mechanism (Brotherton *et al.*, 2002; Saxena *et al.*, 2008). As a result, it is a common practice to deploy multiple sensors to simultaneously monitor different degradation characteristics of the same unit; however, the number of studies on the degradation modeling of multiple sensor signals is still limited.

As different sensor signals are often correlated and may contain only partial information regarding the health status of a unit, effective fusion of such a diverse *"gene pool"* from multiple sensors is critically important to better characterize the underlying degradation process and produce a more accurate prediction of the Remaining Useful Life (RUL; Liu and Shi (2015)). In this article, we consider one of the popular data fusion approaches to construct a composite Health Index (HI) by directly combining the multiple signals from multiple sensors. Mathematically, let $\boldsymbol{L}_i(t) = [L_{i,1}(t), \ldots, L_{i,s}(t)] \in \mathbb{R}^{1\times s}$ be the collected sensor measurements for unit i at time t, where $L_{i,j}(t)$ is the sensor measurement for unit i, sensor j at time t, and s is the number of sensors. Then, the HI is constructed as $h_i(t) = z(\boldsymbol{L}_i(t), \boldsymbol{w}_0)$, where $h_i(t)$ is the HI for unit i at time t, $z(\cdot)$ is an appropriate fusion function to combine the multiple signals from multiple sensors, and $\boldsymbol{w}_0$ is the fusion coefficient. Specifically, we will estimate $\boldsymbol{w}_0$ based on the collected signals of a set of historical units. Then, we leverage the derived fusion coefficient $\hat{\boldsymbol{w}}$ to construct the HIs for in-field units and predict their RULs in real time. There are several significant reasons why the HI-based approach is of great interest. First, by constructing a one-dimensional HI, the problem of analyzing the multiple signals from multiple sensors is transformed into analyzing a single sensor signal (i.e., the HI). As a result, the existing rich literature on the degradation modeling, prognostics, and maintenance scheduling based on a single sensor signal can be directly applied. Second, different degradation models for a single sensor signal can be incorporated into the HI-based method to adapt to different situations. Consequently, the HI-based method gains flexibility and the ability to deal with many practical issues, such as variations among different units (Liu and Huang, 2016). Third, the constructed HI provides a continuous visualization of the degradation progression of a unit, which is easy to understand and thus highly desired in industrial practice. In contrast, other data fusion methods (e.g., see Hu *et al.* (2012) and Tian (2012)) behave like a "black box" that directly produces an output of the RUL prediction without a clear visualization of the degradation process. For these methods, it would be difficult for practitioners to fully understand the current health status of a unit

CONTACT Kaibo Liu kliu8@wisc.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiie.

and thus confidently use the predicted RUL to make better decisions.

Despite the promise of the HI-based approaches, constructing an effective HI is indeed a major challenge. First, although the goal of constructing the HI is to better characterize the degradation process, the true degradation status is always unknown and cannot be directly observed from the multiple sensor signals. Thus, the construction of the HI is significantly different from the classic supervised learning problem (Hastie *et al.*, 2009), where the response variable is directly observable. Likewise, the construction of the HI should be considered as neither an unsupervised learning problem that contains no response variable nor a semi-supervised learning problem (Chapelle *et al.*, 2006), in which the response variable is directly available for only a fraction of samples. Second, the degradation process represented by the HI has to comply with the engineering principles that govern the underlying failure mechanism. Specifically, considering that the degradation is inherently an irreversible process, the degradation path should be monotonic, especially when the degradation is in the form of wear and cumulative damage (Park and Bae, 2010; Ye and Xie, 2015). In addition, the degradation process should mirror the physical fact that a failure occurs once the degradation path of a unit passes the pre-defined failure threshold (Lu and Meeker, 1993; Gebraeel *et al.*, 2005). We call this kind of problem indirect supervised learning. Currently, there is no commonly accepted definition of indirect supervised learning in the literature. Here, we use the term "indirect supervised learning" to refer to a group of problems, in which the response variable (i.e., the underlying degradation status) is not directly observable but information related to the response variable (i.e., failure events and domain knowledge about the degradation process) is available, which thus can be used to indirectly infer the response variable.

In current practice, the HI is mainly constructed by combining the data of a few sensors or Key Performance Indicators based on simplified physical laws and empirical knowledge (Vaisnys *et al.*, 2006). However, such approaches rely on the special skills and experience of personnel and are limited to systems with simple structures and limited sensor data. Although some attempts have been recently made to develop techniques to construct the HI via data-driven approaches (Liu *et al.*, 2013; Liu and Huang, 2016), these methods are heuristic in nature, which cannot guarantee that the constructed HI will lead to an improved prognostic performance. The main reason for this problem is that these existing HI-based models treat the construction of the HI and the prognosis as two separate tasks. Specifically, the HI is constructed using unsupervised learning based on multiple sensor signals and then the prognosis is formulated as a supervised learning problem after the HI is derived. Therefore, the intrinsic connection between the two tasks is missing, and the effectiveness of the constructed HI remains to be tested and verified. Currently, the existing literature still lacks a solid analytic foundation that supports the HI-based approaches. In this article, we aim to fill this gap by developing a novel HI-based method with theoretical justifications.

By combining data fusion methods with the engineering knowledge extracted from the degradation process, we propose to minimize the difference between the predicted failure time and the true failure time of each historical unit. Specifically, our innovative idea is to formulate the construction of an HI via the quantile regression technique and then the engineering knowledge on degradation, such as monotonicity and failure threshold, is realized as constraints in the proposed model. Unlike previous HI-based methods, which contain two separate steps including an unsupervised learning (i.e., construction of HI) step and a supervised learning (i.e., prognosis) step, the proposed method integrates the two steps in a unified manner and formulates the problem of degradation modeling based on multiple sensor signals as indirect supervised learning.

One of the main contributions of this work is that our proposed method guarantees to derive the optimal combination of sensor signals (i.e., the estimated fusion coefficient $\hat{\boldsymbol{w}}$ will converge to the true value $\boldsymbol{w}_0$) and thus it ensures the effective performance of the constructed HI. Currently, there are no other HI-based approaches that guarantee this desirable property. The rest of this article is organized as follows. Section 2 reviews some degradation models that are commonly used to analyze a single sensor signal and also existing data fusion methods for degradation modeling and prognostics. Section 3 describes the details of the proposed method and investigates its asymptotic properties. Section 4 and Section 5 evaluate the proposed method using a number of simulation studies and a case study that involves the degradation of aircraft engines, respectively. Section 6 draws conclusions and discusses future work.

2. Literature review

In this section, we review the related works in the literature. Since the constructed HI can be viewed as a new single sensor signal, we first provide a quick review on the degradation models for a single sensor signal. Then we discuss the existing data fusion methods for degradation modeling and prognostics of multiple sensor signals.

2.1. Degradation models for a single sensor signal

Degradation models for a single sensor signal can be classified into three broad categories: general path models, stochastic process models, and other models. Recent reviews of these degradation models can be found in Si *et al.* (2011) and Ye and Xie (2015).

The general path model was first introduced by Lu and Meeker (1993), where the sensor signal was modeled by a parametric degradation path plus a measurement error with fixed-effect and random-effect parameters. Gebraeel *et al.* (2005) extended the general path model and proposed to update the posterior distribution of the random-effect parameters based on *in situ* signal measurements, which can be further used to predict the RUL of each individual unit. Some other extensions include exploration of different distributions of the random-effect parameters (Yu, 2006; Bae *et al.*, 2007), investigation of different parametric and nonparametric forms of the degradation path (Bae and Kvam, 2004; Zhou, Serban, Gebraeel, and Muller, 2014), and incorporation of dynamic covariates into the general path model (Hong *et al.*, 2015).

The uncertainty of general path models originates from the random-effect parameters and measurement errors. Thus, given the random-effect parameters, the underlying degradation path

is assumed to be deterministic. On the other hand, stochastic process models assume the degradation path to be a stochastic process, such as Wiener process (Whitmore and Schenkelberg, 1997; Si *et al.*, 2012; Ye *et al.,* 2013; Zhai and Ye, 2017), gamma process (Bagdonavicius and Nikulin, 2001; Lawless and Crowder, 2004), and inverse Gaussian process (Wang and Xu, 2010; Ye and Chen, 2014). These models are often used to account for the influence of unobserved environmental factors on the degradation path.

In addition to general path models and stochastic process models, other models such as Proportional Hazard Model (PHM) and state-space models have been developed for modeling of a single degradation signal. For example, Zhou *et al.* (2014) investigated the integration of the general path model and PHM to address hard failures of critical units. Christer *et al.* (1997) applied a Kalman filter to the prognosis and maintenance of furnaces.

One common assumption of the above models is that the single sensor signal can directly reflect the underlying degradation path; i.e., the single sensor signal is either directly treated as the degradation path or assumed to be the degradation path plus a measurement error. However, this assumption may not be true in practical applications, especially for complex systems. In this case, multiple sensors can be used to collect data related to different characteristics of the degradation process and provide more accurate estimation on the degradation status.

2.2. Existing data fusion methods for degradation modeling and prognostics

As previously mentioned, data fusion is critical for dealing with multiple sensor signals. In the literature, data fusion methods for degradation modeling and prognostics can be mainly classified into three categories based on the level at which the fusion operation is performed (Hall and Llinas, 1997; Jardine *et al.,* 2006): decision-level fusion, feature-level fusion, and data-level fusion.

Decision-level fusion methods combine different diagnostic or prognostic results. For example, Hu *et al.* (2012) aggregated the RUL predictions from multiple machine learning algorithms by weighted average, where the weights were determined by cross-validation. Baraldi *et al.* (2012) also used a weighted average to combine prognostic results, but the weights were dynamically derived using a Kalman filter. However, decision-level fusion methods are heuristic and can only obtain a point estimator of the RUL for a unit in general.

The other two categories, feature-level fusion and data-level fusion, combine the extracted features and the raw sensor signals, respectively. The methods of these two categories are generally interchangeable if the extracted features are time-series and can be treated as sensor signals. For example, machine learning algorithms, such as neural networks, are commonly used in the literature and directly produce an output of the RUL prediction with the inputs of the most recent features or signal measurements (Tian, 2012; Loutas *et al.*, 2013). However, these methods fail to realize the unique characteristics of the degradation process. State-space models have also been considered to directly process multiple sensor signals (Xu *et al.,* 2008; Saha *et al.*, 2009). However, state-space models often assume that future degradation depends only on the current degradation status instead of the past—i.e., following the Markov property—which may not be valid in practice (Bae and Kvam, 2004; Chen and Tsui, 2013). Also, it is nontrivial to incorporate variations among different units into state-space models. Recently, Fang *et al.* (2017) proposed to use Functional Principal Component Analysis (FPCA) to extract features from multiple sensor signals for RUL prediction. Unfortunately, the extracted features are difficult to interpret in practice, and their relations to the underlying degradation status are unknown.

Another type of feature-level or data-level fusion method is to construct a one-dimensional HI for characterizing the underlying degradation process, which is the focus of this study. As we introduced in Section 1, the main task of HI-based methods is to estimate the fusion function with the corresponding fusion coefficient. One simple approach is to explicitly express the underlying degradation process as a deterministic function of time (Wang *et al.*, 2008; Yang *et al.*, 2016). Then the fusion coefficient can be estimated by a regression of the multiple sensor measurements against the function values. However, this method oversimplifies the problem and cannot capture the randomness of each unit. To address this issue, Liu *et al.* (2013), Liu and Huang (2016), and Liu *et al.* (2017) proposed to construct the HI by optimizing the desired properties such as a small model fitting error and small variation of failure threshold. However, these methods are still heuristic and lack theoretical justifications. In other words, the constructed HI is not guaranteed to have a better prognostic performance than each single sensor signal.

To summarize, different techniques have been attempted to model multiple sensor signals in degradation analysis and prognostics, and these techniques are commonly formulated into two steps (see, for example, Tian (2012); Liu *et al.* (2013); Loutas *et al.* (2013); Liu and Huang (2016); Fang *et al.* (2017); and Liu *et al.* (2017)). The first step involves feature extraction or HI construction. As reviewed above, since there is no response variable, various unsupervised learning techniques—e.g., FPCA in Fang *et al.* (2017) and optimizing the desired properties of the HI in Liu *et al.* (2017)—are employed in this step. In the second step, the relationship of the output of the first step (treated as the known predictor) to the RUL (treated as the response variable) is derived, which can be regarded as supervised learning. The two steps are considered separately, which thus limits the performance of these methods.

Currently, the existing literature on data fusion models still lacks an effective method tailored for the degradation modeling of multiple sensor signals, which is theoretically justified to ensure the prognostic performance as well as capable to deal with practical issues in degradation applications.

3. Methodology

In this article, we aim to propose a new HI-based method with theoretical justifications. Unlike existing data fusion methods for degradation modeling that consist of two separate steps, the proposed method directly relates the multiple sensor signals and the RUL using indirect supervised learning. With the proposed method, we are able to combine data fusion techniques with engineering knowledge to better satisfy the needs of degradation modeling and prognostics. Specifically, we employ

quantile regression to minimize the difference between the predicted and actual failure times for historical units and then engineering principles extracted from the degradation process are formulated as constraints. We can further show that the estimated fusion coefficient $\hat{\boldsymbol{w}}$ will converge to the true value $\boldsymbol{w}_0$. In the following, we first present a formal definition of the HI in Section 3.1 and then specify the proposed model in Section 3.2. Next, two general cases when implementing our method are discussed in Section 3.3 and Section 3.4, depending on whether some parameters are known or not. Section 3.5 investigates the asymptotic behavior of the proposed method. Section 3.6 describes an iterative procedure to solve the optimization problem involved and Section 3.7 discusses RUL prediction and tuning parameters.

3.1. Definition of HI

Before describing the proposed method, it is important to first present a formal definition of the HI. Previous studies simply considered the single signal constructed from the combination of multiple sensor signals as the HI (Liu *et al.*, 2013; Liu *et al.*, 2017). However, this definition is unclear, as there are many ways to combine multiple sensor signals and, by this definition, the HI is not necessarily related to the underlying degradation process. Therefore, a formal definition of the HI has to be provided first to highlight the basic concept of the HI-based methods.

Following the common settings in the literature (Lu and Meeker, 1993; Gebraeel *et al.*, 2005), we define the failure time of unit i as the time that the underlying degradation process $\eta(t; \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda})$ first crosses the failure threshold l:

$$T_i = \operatorname*{argmin}_t \eta(t; \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}) \geq l, \tag{1}$$

where T_i is the failure time for unit i, $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Lambda}$ are random-effect and fixed-effect parameters respectively, and the degradation process $\eta(\cdot)$ can take any form, such as the general path model and Wiener process model mentioned in Section 2.1.

The main idea of HI-based methods is to unravel the underlying degradation process via the combination of multiple sensor signals. Specifically, we assume the underlying degradation process can be derived from an appropriate fusion function based on multiple sensor signals with the contamination of white noise:

$$\eta(t; \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}) = z(\boldsymbol{L}_i(t), \boldsymbol{w}_0) - \varepsilon_i(t). \tag{2}$$

Here $\varepsilon_i(t) \sim N(0, \sigma_0^2)$ is the independent Gaussian noise with variance σ_0^2. A special fusion function $z(\boldsymbol{L}_i(t), \boldsymbol{w}_0) = L_{i,j}(t)$ selects one signal from the multiple sensor signals, which indicates that there exists a single sensor signal directly reflecting the underlying degradation path. This represents the case where the degradation model for a single signal is applied. From this point of view, the restrictive assumption of degradation models for a single signal is relaxed via the fusion function in our method.

Then, we present a formal definition of HI as follows:

Definition 1. The HI is a combination of sensor signals that characterizes the underlying degradation process

$$h_i(t) = z(\boldsymbol{L}_i(t), \boldsymbol{w}_0), \tag{3}$$

where $h_i(t)$ is the HI for unit i at time t. Based on Equations (2) and (3), we obtain

$$h_i(t) = \eta(t; \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}) + \varepsilon_i(t).$$

3.2. Model formulation

Although using the proposed method makes it possible to incorporate a variety of degradation models for a single signal as reviewed in Section 2.1, we focus on the general path model for the underlying degradation process in this article. In addition, to highlight our main ideas, we consider the following parametric form of the degradation model:

$$\eta(t; \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}) = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i, \tag{4}$$

where $\boldsymbol{\psi}(t) \in \mathbb{R}^{1\times p}$ contains the time covariates with each entry to be a smooth function of t (e.g., for $\boldsymbol{\psi}(t) = [1, t, \ldots, t^{p-1}]$, we have a polynomial model), and $\boldsymbol{\Gamma}_i$ is a p-dimensional random-effect parameter. In this study, we follow the existing HI-based literature, such as Liu *et al.* (2017), and assume that $\boldsymbol{\Gamma}_i$ follows a multivariate normal distribution $\boldsymbol{\Gamma}_i \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. We will consider the relaxation of this normal distribution assumption in a future study. The general path model in Equation (4) is flexible in describing various degradation processes and has been widely utilized in the literature (Lu *et al.*, 1997; Gebraeel, 2006; Liu *et al.*, 2013; Zhou *et al.*, 2014). The advantage of this model is that conditioning on the collected sensor signals, the analytical expression for the posterior distribution of $\boldsymbol{\Gamma}_i$ and the distribution of failure time $F_{T_i}(t)$ can be derived. In practice, the form of $\boldsymbol{\psi}(t)$ can be determined from domain knowledge or historical data. When such knowledge is not available, nonparametric models can be adopted where $\boldsymbol{\psi}(t)$ is composed of a series of functional basis (Zhou, Serban, and Gebraeel, 2014; Zhou, Serban, Gebraeel, and Muller, 2014; Hong *et al.*, 2015).

In addition to the degradation model, we need to specify the form of the fusion function $z(\cdot)$. Without loss of generality, in this article, we focus on the linear fusion function and will consider the extension to nonlinear fusion functions in a future study:

$$z(\boldsymbol{L}_i(t), \boldsymbol{w}_0) = \boldsymbol{L}_i(t)\boldsymbol{w}_0.$$

In fact, a variety of nonlinear fusion functions can be linearized by creating artificial sensor signals. For example, if the fusion function has the form

$$z(\boldsymbol{L}_i(t), \boldsymbol{w}_0) = \sum_{k=1}^{K} u_k(\boldsymbol{L}_i(t))w_k, \tag{5}$$

where $\boldsymbol{w}_0 = (w_1, \ldots, w_K)^T \in \mathbb{R}^{K\times 1}$, and $u_k(\cdot)$ is a function of $\boldsymbol{L}_i(t)$, then it is equivalent to $z(\boldsymbol{L}_i'(t), \boldsymbol{w}_0) = \boldsymbol{L}_i'(t)\boldsymbol{w}_0$, where $\boldsymbol{L}_i'(t) = [u_1(\boldsymbol{L}_i(t)), \ldots, u_K(\boldsymbol{L}_i(t))] \in \mathbb{R}^{1\times K}$. In this way, a more general fusion function can be approximated in the form of Equation (5) using a set of functional basis functions $\{u_k(\cdot), k = 1, \ldots, K\}$.

As a result, we have the following relation among the HIs $h_i(t)$, the multiple sensor signals $\boldsymbol{L}_i(t)$, and the underlying degradation process

$$h_i(t) = \boldsymbol{L}_i(t)\boldsymbol{w}_0 = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i + \varepsilon_i(t). \tag{6}$$

If we let $\boldsymbol{h}_i = [h_i(t_{i,1}), \ldots, h_i(t_{i,n_i})]^T \in \mathbb{R}^{n_i \times 1}$ and

$$\boldsymbol{L}_i = \begin{bmatrix} \boldsymbol{L}_i(t_{i,1}) \\ \vdots \\ \boldsymbol{L}_i(t_{i,n_i}) \end{bmatrix} \in \mathbb{R}^{n_i \times s}$$

be the HI and all available multiple sensor signals for historical unit i, respectively, and let

$$\boldsymbol{\Psi}_i = \begin{bmatrix} \boldsymbol{\psi}(t_{i,1}) \\ \vdots \\ \boldsymbol{\psi}(t_{i,n_i}) \end{bmatrix} \in \mathbb{R}^{n_i \times p}$$

be the design matrix, and $\boldsymbol{\varepsilon}_i = [\varepsilon_i(t_{i,1}), \ldots, \varepsilon_i(t_{i,n_i})]^T \in \mathbb{R}^{n_i \times 1}$, then Equation (6) can be written in matrix form as

$$\boldsymbol{h}_i = \boldsymbol{L}_i \boldsymbol{w}_0 = \boldsymbol{\Psi}_i \boldsymbol{\Gamma}_i + \boldsymbol{\varepsilon}_i. \tag{7}$$

The parameters related to this model include the fusion coefficient $\boldsymbol{w}_0$, the failure threshold l, and the set of parameters $\boldsymbol{\Omega}_0 = \{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \sigma_0^2\}$. There are two general cases that we need to consider depending on whether $\boldsymbol{\Omega}_0$ and l are known or not in practice. Next, we discuss the parameter estimation for both cases in detail.

3.3. Case I: Ω_0 and l are known

To begin with, we start from the case where $\boldsymbol{\Omega}_0$ and the failure threshold l are known. Accordingly, the only unknown parameter is the fusion coefficient $\boldsymbol{w}_0$. To predict the failure time of unit i, we adopt the Bayesian updating approach proposed by Gebraeel *et al.* (2005). In particular, conditioning on the constructed HI $\boldsymbol{h}_i = \boldsymbol{L}_i \boldsymbol{w}_0$, the posterior distribution of $\boldsymbol{\Gamma}_i$ is

$$P(\boldsymbol{\Gamma}_i | \boldsymbol{L}_i \boldsymbol{w}_0) \propto P(\boldsymbol{L}_i \boldsymbol{w}_0 | \boldsymbol{\Gamma}_i) P(\boldsymbol{\Gamma}_i).$$

According to Equation (7), the conditional distribution $\boldsymbol{L}_i \boldsymbol{w}_0 | \boldsymbol{\Gamma}_i \sim N_{n_i}(\boldsymbol{\Psi}_i \boldsymbol{\Gamma}_i, \sigma_0^2 \boldsymbol{I}_{n_i \times n_i})$, where $\boldsymbol{I}_{n_i \times n_i} \in \mathbb{R}^{n_i \times n_i}$ is the identity matrix. Recall the random-effect parameter $\boldsymbol{\Gamma}_i \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Let $\boldsymbol{X}_{i,1} = \boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i \in \mathbb{R}^{p \times p}$, $\boldsymbol{X}_{i,2} = \boldsymbol{\Psi}_i^T \boldsymbol{L}_i \in \mathbb{R}^{p \times s}$, and $\boldsymbol{X}_i = [\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}]$, then we can obtain that the posterior $\boldsymbol{\Gamma}_i | \boldsymbol{L}_i \boldsymbol{w}_0$ is normally distributed:

$$\boldsymbol{\Gamma}_i | \boldsymbol{L}_i \boldsymbol{w}_0 \sim N_p(\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0), \boldsymbol{\Sigma}_i(\boldsymbol{X}_i; \boldsymbol{\Omega}_0)),$$

where

$$\begin{aligned} \boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0) &= \left(\frac{\boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} \left(\frac{\boldsymbol{\Psi}_i^T \boldsymbol{L}_i \boldsymbol{w}_0}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 \right) \\ &= \left(\frac{\boldsymbol{X}_{i,1}}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} \left(\frac{\boldsymbol{X}_{i,2} \boldsymbol{w}_0}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 \right), \end{aligned} \tag{8}$$

$$\boldsymbol{\Sigma}_i(\boldsymbol{X}_i; \boldsymbol{\Omega}_0) = \left(\frac{\boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} = \left(\frac{\boldsymbol{X}_{i,1}}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1}. \tag{9}$$

Thus, the posterior $\boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i | \boldsymbol{L}_i \boldsymbol{w}_0$ also follows a normal distribution:

$$\begin{aligned} &\boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i | \boldsymbol{L}_i \boldsymbol{w}_0 \\ &\quad \sim N(\boldsymbol{\psi}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0), \boldsymbol{\psi}(t)\boldsymbol{\Sigma}_i(\boldsymbol{X}_i; \boldsymbol{\Omega}_0)\boldsymbol{\psi}(t)^T). \end{aligned} \tag{10}$$

According to Equation (1), the Cumulative Distribution Function (CDF) of the failure time T_i is

$$\begin{aligned} F_{T_i}(t | \boldsymbol{L}_i \boldsymbol{w}_0) &= P(T_i \le t | \boldsymbol{L}_i \boldsymbol{w}_0) = P(\boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i \ge l | \boldsymbol{L}_i \boldsymbol{w}_0) \\ &= \Phi\left(\frac{\boldsymbol{\psi}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0) - l}{\sqrt{\boldsymbol{\psi}(t)\boldsymbol{\Sigma}_i(\boldsymbol{X}_i; \boldsymbol{\Omega}_0)\boldsymbol{\psi}(t)^T}} \right) \\ &= \Phi(g(t, \boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0)). \end{aligned} \tag{11}$$

Here Φ is the CDF of a standard normal distribution. Since the distribution is skewed, we follow the existing literature and consider the median value $\hat{T}_i$ as the estimated failure time (Gebraeel *et al.*, 2005). This means that $\hat{T}_i$ satisfies $F_{T_i}(\hat{T}_i) = 0.5$ and thus

$$\boldsymbol{\psi}(\hat{T}_i)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0) - l = 0. \tag{12}$$

Please note that in the procedure described above, we utilize all available sensor measurements of each historical unit to accurately estimate the failure time. Let $r(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0) = \hat{T}_i$ be the real positive root of the equation $\boldsymbol{\psi}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}_0; \boldsymbol{\Omega}_0) - l = 0$. Then, we propose to estimate $\boldsymbol{w}_0$ by using quantile regression to minimize the difference between the estimated failure time $\hat{T}_i$ and the true failure time τ_i for all m historical units:

$$\begin{aligned} \hat{\boldsymbol{w}} &= \underset{\boldsymbol{w}}{\operatorname{argmin}} \sum_{i=1}^{m} |r(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}_0) - \tau_i|, \\ \text{s.t.} \quad & \boldsymbol{w} \in \mathcal{W}_{\boldsymbol{\Omega}_0}. \end{aligned} \tag{13}$$

In this formulation, we constrain $\boldsymbol{w}$ to be within the set $\mathcal{W}_{\boldsymbol{\Omega}_0}$, which is required to ensure that the estimated process $\boldsymbol{\psi}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}_0)$ complies with the engineering knowledge extracted from the underlying degradation process. Specifically, we consider the following constraints.

Constraint 1: $||\boldsymbol{w}|| \le C_0$, *where* $||\boldsymbol{w}||$ *is the* l_2*-norm of* $||\boldsymbol{w}||$ *and* C_0 *is a positive number. This constraint ensures that* $\mathcal{W}_{\boldsymbol{\Omega}_0}$ *is bounded, based on the fact that* $\boldsymbol{w}$ *should be finite in practice.*

Constraint 2: $\boldsymbol{\psi}(0)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}_0) \le l - C_1$, *where* C_1 *is a positive number. This constraint ensures that at time 0, the degradation process is strictly below the failure threshold* l.

Constraint 3: $\dot{\boldsymbol{\psi}}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}_0) \ge C_2$, *where* $\dot{\boldsymbol{\psi}}(t) = \partial \boldsymbol{\psi}(t)/\partial t$ *and* C_2 *is a positive number. This constraint ensures that the degradation process is strictly increasing.* ▲

Due to the $|\cdot|$ operator, the objective function of Equation 13 is not smooth, which increases the complexity in numerically solving the problem. A common practice to address this issue is to transform the objective function to a smooth one with additional constraints. Specifically, we can transform Equation 13 into

$$\begin{aligned} &\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \sum_{i=1}^{m} (\xi_{i,1} + \xi_{i,2}), \\ &\text{s.t. } r(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}_0) - \tau_i = \xi_{i,1} - \xi_{i,2}, \xi_{i,1} \ge 0, \xi_{i,2} \ge 0, \ \boldsymbol{w} \in \mathcal{W}_{\boldsymbol{\Omega}_0}. \end{aligned}$$

Then a variety of algorithms, such as the interior point algorithm, and sequential quadratic programming can be implemented to solve this constrained nonlinear optimization problem (Nocedal and Wright, 2006).

3.4. Case II: Ω_0 and l are unknown

In practice, the parameters $\mathbf{\Omega}_0$ of the general path model and the failure threshold l are often unknown and thus we need to estimate $\mathbf{\Omega}_0, l, \boldsymbol{w}_0$ simultaneously from the historical units. In this subsection, we will discuss how to address this challenge by extending the model of Case I and incorporating the relations between $\mathbf{\Omega}_0$ and $\boldsymbol{w}_0$.

First, we observe that if we multiply $\boldsymbol{w}, \boldsymbol{\mu}, l$ by the same scalar $\gamma > 0$ and multiply $\mathbf{\Sigma}, \sigma^2$ by γ^2, then the resulting $r(\boldsymbol{X}_i, \boldsymbol{w}; \mathbf{\Omega})$ remains unchanged, where $\mathbf{\Omega} = \{\boldsymbol{\mu}, \mathbf{\Sigma}, \sigma^2\}$. In other words, $\boldsymbol{w}, \boldsymbol{\mu}, l, \mathbf{\Sigma}, \sigma^2$ and $\gamma\boldsymbol{w}, \gamma\boldsymbol{\mu}, \gamma l, \gamma^2\mathbf{\Sigma}, \gamma^2\sigma^2$ have the same objective function value $\sum_{i=1}^{m} |r(\boldsymbol{X}_i, \boldsymbol{w}; \mathbf{\Omega}) - \tau_i|$ and thus the optimal solution is not unique. To avoid this issue, we can set l to be any positive number—e.g., $l = 1$—and then we estimate $\mathbf{\Omega}_0$ and $\boldsymbol{w}_0$ simultaneously given l. Guided by Lu and Meeker (1993), the relation between $\mathbf{\Omega}_0$ and $\boldsymbol{w}_0$ can be formulated as follows.

Given $\boldsymbol{w}_0$, we estimate the true value of the random-effect parameter $\mathbf{\Gamma}_i$ as

$$\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0) = \left(\mathbf{\Psi}_i^T\mathbf{\Psi}_i\right)^{-1}\mathbf{\Psi}_i^T\boldsymbol{L}_i\boldsymbol{w}_0 = \boldsymbol{X}_{i,1}^{-1}\boldsymbol{X}_{i,2}\boldsymbol{w}_0.$$

According to Equation (7), $\boldsymbol{L}_i\boldsymbol{w}_0$ has a normal distribution with mean $E[\boldsymbol{L}_i\boldsymbol{w}_0] = \mathbf{\Psi}_i\boldsymbol{\mu}_0$ and variance $Var[\boldsymbol{L}_i\boldsymbol{w}_0] = \mathbf{\Psi}_i\mathbf{\Sigma}_0\mathbf{\Psi}_i^T + \sigma_0^2\boldsymbol{I}$ and thus the distribution of $\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)$ is also normal with mean and variance as follows:

$$\begin{aligned} E[\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)] &= \left(\mathbf{\Psi}_i^T\mathbf{\Psi}_i\right)^{-1}\mathbf{\Psi}_i^T E[\boldsymbol{L}_i\boldsymbol{w}_0] = \boldsymbol{\mu}_0, \\ Var[\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)] &= \left(\mathbf{\Psi}_i^T\mathbf{\Psi}_i\right)^{-1}\mathbf{\Psi}_i^T Var[\boldsymbol{L}_i\boldsymbol{w}_0]\mathbf{\Psi}_i\left(\mathbf{\Psi}_i^T\mathbf{\Psi}_i\right)^{-1} \\ &= \mathbf{\Sigma}_0 + \sigma_0^2\left(\mathbf{\Psi}_i^T\mathbf{\Psi}_i\right)^{-1}. \end{aligned}$$

Therefore, the unbiased estimators for $\boldsymbol{\mu}_0, \sigma_0^2$, and $\mathbf{\Sigma}_0$ are

$$\hat{\boldsymbol{\mu}}(\boldsymbol{w}_0) = \frac{1}{m}\sum_{i=1}^{m}\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0), \tag{14}$$

$$\hat{\sigma}^2(\boldsymbol{w}_0) = \frac{1}{m}\sum_{i=1}^{m}\frac{\left(\boldsymbol{L}_i\boldsymbol{w}_0 - \mathbf{\Psi}_i\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)\right)^T\left(\boldsymbol{L}_i\boldsymbol{w}_0 - \mathbf{\Psi}_i\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)\right)}{n_i - p}, \tag{15}$$

$$\begin{aligned} \hat{\mathbf{\Sigma}}(\boldsymbol{w}_0) = &\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0) - \hat{\boldsymbol{\mu}}(\boldsymbol{w}_0))(\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0) - \hat{\boldsymbol{\mu}}(\boldsymbol{w}_0))^T \\ &- \frac{1}{m}\sum_{i=1}^{m}\frac{\left(\boldsymbol{L}_i\boldsymbol{w}_0 - \mathbf{\Psi}_i\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)\right)^T\left(\boldsymbol{L}_i\boldsymbol{w}_0 - \mathbf{\Psi}_i\hat{\mathbf{\Gamma}}_i(\boldsymbol{w}_0)\right)}{n_i - p} \\ &\times \left(\mathbf{\Psi}_i^T\mathbf{\Psi}_i\right)^{-1}. \end{aligned} \tag{16}$$

If $\mathbf{\Omega}_0$ is unknown, we can estimate $\mathbf{\Omega}_0$ by $\hat{\mathbf{\Omega}}(\boldsymbol{w}_0) = \{\hat{\boldsymbol{\mu}}(\boldsymbol{w}_0), \hat{\mathbf{\Sigma}}(\boldsymbol{w}_0), \hat{\sigma}^2(\boldsymbol{w}_0)\}$ based on $\boldsymbol{w}_0$, and we can prove the following proposition (see Appendix for details).

Proposition 1. *As $m \to \infty$, the estimated values $\hat{\mathbf{\Omega}}(\boldsymbol{w}_0)$ converge to the true values $\mathbf{\Omega}_0$.*

Therefore, Equation (11) can also be regarded as the true CDF of the failure time T_i with $\mathbf{\Omega}_0$ replaced by $\hat{\mathbf{\Omega}}(\boldsymbol{w}_0)$ when m is large. To address the challenge that $\boldsymbol{w}_0$ is also unknown, we formulate $\hat{\mathbf{\Omega}}(\boldsymbol{w})$ as a function of $\boldsymbol{w}$ and estimate $\boldsymbol{w}$ by solving the following quantile regression problem with constraints

$$\begin{aligned} \hat{\boldsymbol{w}} = &\operatorname*{argmin}_{\boldsymbol{w}} \sum_{i=1}^{m} |r(\boldsymbol{X}_i, \boldsymbol{w}; \hat{\mathbf{\Omega}}(\boldsymbol{w})) - \tau_i| \\ \text{s.t.} \quad &\boldsymbol{w} \in \mathcal{W}_{\hat{\mathbf{\Omega}}(\boldsymbol{w})}. \end{aligned} \tag{17}$$

In this formulation, we require $\hat{\mathbf{\Sigma}}(\boldsymbol{w})$ to be well conditioned, since we need to calculate the inverse as in Equation 8 with $\mathbf{\Sigma}_0$ replaced by $\hat{\mathbf{\Sigma}}(\boldsymbol{w})$. Other constraints are similar to Case I with $\mathbf{\Omega}_0$ replaced by $\hat{\mathbf{\Omega}}(\boldsymbol{w})$. Accordingly, we list the constraints as follows:

Constraint 1: $||\boldsymbol{w}|| \le C_0$.

Constraint 2: $\boldsymbol{\psi}(0)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \hat{\mathbf{\Omega}}(\boldsymbol{w})) \le l - C_1$.

Constraint 3: $\dot{\boldsymbol{\psi}}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \hat{\mathbf{\Omega}}(\boldsymbol{w})) \ge C_2$.

Constraint 4: $\hat{\mathbf{\Sigma}}(\boldsymbol{w})$ is well conditioned.

Solving this optimization problem, however, is challenging. One of the reasons is that the constraints are highly nonlinear. In Section 3.6, we propose an iterative procedure to search for the optimal solution for Case II.

3.5. Asymptotic property

In this subsection, we will focus on studying the asymptotic property of $\hat{\boldsymbol{w}}$ that is estimated by using the proposed method. In particular, we can show that the estimated $\hat{\boldsymbol{w}}$ indeed converges to the true $\boldsymbol{w}_0$:

Theorem 1. *If $\boldsymbol{w}_0$ is unique and is an inner point of the search space $\mathcal{W}$, then for both Equation* (13) *in Case I and Equation* (17) *in Case II, there exists positive definite matrices $\boldsymbol{G}_0$ and $\boldsymbol{G}_1$ such that*

$$\sqrt{m}(\hat{\boldsymbol{w}} - \boldsymbol{w}_0) \to N_s\left(0, \frac{1}{4}\boldsymbol{G}_1^{-1}\boldsymbol{G}_0\boldsymbol{G}_1^{-1}\right) \tag{18}$$

as $m \to \infty$.

The proof of the theorem is based on the convergence theory of nonlinear quantile regression (see Oberhofer (1982) and Koenker (2005 Chap. 4.4)). The details of the proof are omitted here due to space limitations. We will further demonstrate the convergence of the estimated $\hat{\boldsymbol{w}}$ in the simulation study in Section 4.

3.6. An iterative procedure for case II

As previously mentioned, solving the optimization problem in Equation (17) for Case II is challenging due to the highly nonlinear constraints. In this section, we propose an iterative procedure to search for the optimal solution in Case II when the prior parameter $\mathbf{\Omega}_0$ is unknown. The iterative procedure is described as follows.

Iterative Procedure for Solving Case II

Input: initial guess $\boldsymbol{w}^{(0)}$, failure threshold l, parameters C_0, C_1, C_2, sensor signals $\boldsymbol{L}_i$, failure time τ_i for all historical units ($i = 1, \ldots, m$), and maximum number of iterations $k_{\max}$.

Output: estimated fusion coefficient $\hat{\boldsymbol{w}}$.

1: **Procedure**:
2: Set $k = 0$.
3: **do**
4: Calculate $\boldsymbol{\Omega}^{(k+1)} = \hat{\boldsymbol{\Omega}}(\boldsymbol{w}^{(k)})$ from (14)–(16) with $\boldsymbol{w}_0$ replaced by $\boldsymbol{w}^{(k)}$. If $\hat{\boldsymbol{\Sigma}}(\boldsymbol{w}^{(k)})$ is ill conditioned, let $\hat{\boldsymbol{\Sigma}}(\boldsymbol{w}^{(k)}) = \hat{\boldsymbol{\Sigma}}(\boldsymbol{w}^{(k)}) + \gamma \boldsymbol{I}_{p\times p}$, where γ is a small positive scalar such that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{w}^{(k)})$ is well conditioned.
5: If $\boldsymbol{w}^{(k)} \in \mathcal{W}_{\boldsymbol{\Omega}^{(k+1)}}$, calculate

$$\boldsymbol{w}^{(k+1)} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \sum_{i=1}^{m} |r(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}^{(k+1)}) - \tau_i| \quad \text{s.t. } \boldsymbol{w} \in \mathcal{W}_{\boldsymbol{\Omega}^{(k+1)}}. \tag{19}$$

Otherwise, generate $\boldsymbol{w}^{(k+1)}$ such that it satisfies the constraint $\boldsymbol{w}^{(k+1)} \in \mathcal{W}_{\hat{\boldsymbol{\Omega}}(\boldsymbol{w}^{(k+1)})}$.
6: Set $k = k + 1$
7: **until** $k = k_{\max}$ or $\boldsymbol{w}^{(k)}$ converges
8: Set $\hat{\boldsymbol{w}} = \boldsymbol{w}^{(k)}$.

The idea behind the iterative procedure is to iteratively approximate $\boldsymbol{\Omega}_0$ by $\boldsymbol{\Omega}^{(k+1)}$. If $\boldsymbol{\Omega}^{(k+1)}$ is a good approximation, then the corresponding $\boldsymbol{w}^{(k+1)}$ calculated in Equation (19) will be close to $\hat{\boldsymbol{w}}$ estimated from Equation (13); i.e., a consistent estimator of $\boldsymbol{w}_0$ according to Theorem 1. There are two assumptions associated with the iterative procedure. The first assumption is that there exists a $\boldsymbol{w}^*$ within a *small* neighborhood $\mathcal{N}_1(\boldsymbol{w}_0)$ of $\boldsymbol{w}_0$ such that $\boldsymbol{w}^* = \operatorname{argmin}_{\boldsymbol{w}} \sum_{i=1}^{m} |r(\boldsymbol{X}_i, \boldsymbol{w}; \hat{\boldsymbol{\Omega}}(\boldsymbol{w}^*)) - \tau_i|$ where $\boldsymbol{w} \in \mathcal{W}_{\hat{\boldsymbol{\Omega}}(\boldsymbol{w}^*)}$. In other words, if we use the prior parameters $\hat{\boldsymbol{\Omega}}(\boldsymbol{w}^*)$ estimated based on $\boldsymbol{w}^*$ to replace $\boldsymbol{\Omega}_0$ in Equation (13), we will obtain $\boldsymbol{w}^*$ again. We call this property of $\boldsymbol{w}^*$ *self-producing*. If $m \to \infty$, we have shown that $\hat{\boldsymbol{\Omega}}(\boldsymbol{w}_0) \to \boldsymbol{\Omega}_0$ and $\hat{\boldsymbol{w}} \to \boldsymbol{w}_0$ where $\hat{\boldsymbol{w}}$ is estimated by Equation (13) and thus $\boldsymbol{w}_0$ is self-producing by neglecting the (possible) small difference between $\hat{\boldsymbol{w}}$ and $\boldsymbol{w}_0$. Practically, m is finite and, in such a case, we assume another point $\boldsymbol{w}^* \in \mathcal{N}_1(\boldsymbol{w}_0)$ to be self-producing, which is a necessary condition for the iterative procedure to converge to $\boldsymbol{w}^*$. As m increases, the neighborhood $\mathcal{N}_1(\boldsymbol{w}_0)$ becomes smaller and $\boldsymbol{w}^*$ becomes closer to $\boldsymbol{w}_0$.

The second assumption is that for $\boldsymbol{w}^{(k)}$ within a neighborhood $\mathcal{N}_2(\boldsymbol{w}^*)$ of $\boldsymbol{w}^*$, the updated $\boldsymbol{w}^{(k+1)}$ after one iteration gets closer to $\boldsymbol{w}^*$ from $\boldsymbol{w}^{(k)}$. This assumption relies on the variance of the estimated prior parameters and the sensitivity of Equation (19) to the mis-specification of the prior parameters. Specifically, if m is large, the variance of the estimated prior parameter $\boldsymbol{\Omega}^{(k+1)} = \hat{\boldsymbol{\Omega}}(\boldsymbol{w}^{(k)})$ is small. Thus, if $\boldsymbol{w}^{(k)} \in \mathcal{N}_2(\boldsymbol{w}^*)$, the difference between $\boldsymbol{\Omega}^{(k+1)}$ and $\hat{\boldsymbol{\Omega}}(\boldsymbol{w}^*)$ is bounded. If Equation (19) is not sensitive to the mis-specification of the prior parameters, then the resulting $\boldsymbol{w}^{(k+1)}$ will be closer to $\boldsymbol{w}^*$. This conclusion requires n_i to be relatively large; i.e., a large number of historical observations are available for unit i. This is often valid in practice, due to the rapid development of sensor and cloud technologies. In this way, as $n_i \to \infty$, the posterior mean $\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}) \to \hat{\boldsymbol{\Gamma}}(\boldsymbol{w})$, which only has a small dependence on the prior parameter $\boldsymbol{\Omega}$. As a result, the estimated failure time $r(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega})$ will be insensitive to the value of $\boldsymbol{\Omega}$ but will be sensitive to the value of $\boldsymbol{w}$, and then our second assumption will be valid. To summarize, with a large m and a large n_i, the iterative procedure is expected to deliver a better convergence performance and a more accurate estimation of $\boldsymbol{w}_0$. This point will be further investigated in Sections 4.1 and 4.2.

Consequently, an important issue here is that the initial guess $\boldsymbol{w}^{(0)}$ should be close to $\boldsymbol{w}^*$. Practically, there are several options for choosing $\boldsymbol{w}^{(0)}$. For example, we can use the estimation of $\boldsymbol{w}_0$ from existing data-level fusion methods (Liu and Huang, 2016; Liu *et al.*, 2017) as an initial guess $\boldsymbol{w}^{(0)}$. Another option is to calculate $\boldsymbol{w}^{(0)}$ by solving Equation (19) given $k = 0$ and noninformative prior parameters $[\boldsymbol{\Sigma}^{(1)}]^{-1} = \boldsymbol{0}$, in which case $\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}; \boldsymbol{\Omega}^{(1)}) = \hat{\boldsymbol{\Gamma}}(\boldsymbol{w})$ does not depend on $\boldsymbol{\Omega}^{(1)}$. Meanwhile, to ensure the performance of the proposed iterative procedure in practice, we can also randomly generate a number of $\boldsymbol{w}^{(0)} \in \mathcal{W}_{\boldsymbol{\Omega}^{(0)}}$, repeat the iterative procedure multiple times with different initial guesses of $\boldsymbol{w}^{(0)}$, and finally select the best solution based on the values of the objective function.

3.7. RUL prediction and discussion

In this subsection, we discuss how to predict the RUL based on the estimated $\hat{\boldsymbol{w}}$ for in-field unit i when this unit has partially degraded but has not yet failed. The procedure is exactly the same as Equations (8) to (11) except that $\boldsymbol{w}_0$ is replaced by $\hat{\boldsymbol{w}}$. In particular, the CDF of the failure time can be updated in real time given that the RUL > 0. For example, in Case I, the CDF can be calculated as

$$F_{T_i}(t|T_i > t_{i,n}) = \frac{\Phi(g(t, \boldsymbol{X}_i, \hat{\boldsymbol{w}}; \boldsymbol{\Omega}_0)) - \Phi(g(t_{i,n_i}, \boldsymbol{X}_i, \hat{\boldsymbol{w}}; \boldsymbol{\Omega}_0))}{1 - \Phi(g(t_{i,n_i}, \boldsymbol{X}_i, \hat{\boldsymbol{w}}; \boldsymbol{\Omega}_0))}.$$

Here n_i is the number of available observations $\boldsymbol{L}_i(t)$ for in-field unit i, and t_{i,n_i} is the time when the last measurement was collected. Then the predicted failure time $\hat{T}_i$ can be calculated as the median value of the updated CDF as in Equation (12).

The prognostic performance of our proposed method can be well explained by the similarity of the procedures in parameter estimation and RUL prediction. In parameter estimation, we choose $\hat{\boldsymbol{w}}$ to minimize the prediction error in historical units and thus if historical units and in-field units are drawn from the same population, the prognostic performance for the in-field units should also be satisfactory. This is one of the main differences between our proposed method and existing data fusion approaches.

In the model formulation, there are several parameters C_0, C_1, and C_2. A large C_0 and a small C_1 and C_2 will enlarge the search space $\mathcal{W}_{\boldsymbol{\Omega}_0}$ or $\mathcal{W}_{\hat{\boldsymbol{\Omega}}(\boldsymbol{w})}$ and thus $\boldsymbol{w}_0$ is more likely to be within the search space. On the other hand, a small C_0 and a large C_1 and C_2 will reduce the search space and thus likely decrease the computational cost. Therefore, the setting of these parameters should be based on the practical application. If $\boldsymbol{w}_0$ is known to be within a smaller search space, then a small C_0 and a large C_1 and C_2 are preferred; otherwise, the search space should be large enough to avoid excluding $\boldsymbol{w}_0$.

4. Simulation study

We conduct a number of simulation studies to evaluate the estimation and prognostic performance of our proposed method for both Cases I and II. In particular, we consider the following three scenarios: (i) the ideal scenario where all assumptions are satisfied; (ii) the available sensor measurements are sparse; and (iii) the degradation path model is mis-specified.

4.1. Ideal scenario

First, we consider the ideal scenario when all assumptions are satisfied. Specifically, we aim to verify the asymptotic property in Sections 3.5 using a simulated dataset.

In this simulation, we randomly generate 1000 units with a linear degradation path; i.e., the underlying degradation process for each unit is

$$\eta(t; \mathbf{\Gamma}_i, \mathbf{\Lambda}) = \Gamma_{i,0} + \Gamma_{i,1} t,$$

where the random-effect parameter $\mathbf{\Gamma}_i$ of unit i follows the prior distribution

$$\mathbf{\Gamma}_i = \begin{pmatrix} \Gamma_{i,0} \\ \Gamma_{i,1} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 100 & 1 \\ 1 & 0.5 \end{pmatrix} \right).$$

Note that the probability of $\Gamma_{i,1} \leq 0$ is less than 10^{-4}, which is thus negligible. If any $\mathbf{\Gamma}_i$ with $\Gamma_{i,1} \leq 0$ is generated, we reject the sample and generate a new one to ensure that the underlying degradation process of any unit is increasing. The failure threshold is set to be $l = 400$ and we record the true failure time as τ_i according to Equation (1). The underlying true HI is generated by adding a random noise to the underlying degradation process:

$$h_i(t) = \eta(t; \mathbf{\Gamma}_i, \mathbf{\Lambda}) + \varepsilon_i(t) = \Gamma_{i,0} + \Gamma_{i,1} t + \varepsilon_i(t),$$

where $\varepsilon_i(t) \sim N(0, 20^2)$. Furthermore, we consider four sensors with the true value of fusion coefficient $\boldsymbol{w}_0 = (w_1, w_2, w_3, w_4) = (0.6, 0.2, -0.5, 0)^T$. Signal 1 is randomly generated by

$$L_{i,1}(t) = U_{i,1}^{(1)} \sqrt{t} - U_{i,1}^{(2)} \sin(0.05t) + \varepsilon_{i,1}(t), \tag{20}$$

and Signal 2 is randomly generated by

$$L_{i,2}(t) = U_{i,2}^{(1)} t + U_{i,2}^{(2)} \sin(0.1t) + \varepsilon_{i,2}(t). \tag{21}$$

Signal 3 is calculated based on $\boldsymbol{w}_0$ using the HI $h_i(t)$ and the first two signals:

$$L_{i,3}(t) = (h_i(t) - w_1 L_{i,1}(t) - w_2 L_{i,2}(t))/w_3. \tag{22}$$

Since $w_4 = 0$, we are considering the case that Signal 4 is not related to the underlying degradation process, which is likely to happen in practice. Signal 4 is randomly generated by

$$L_{i,4}(t) = U_{i,4}^{(1)} t + U_{i,4}^{(2)} + \varepsilon_{i,4}(t). \tag{23}$$

In Equations (20) to (23), $U_{i,1}^{(1)}, U_{i,1}^{(2)}, U_{i,2}^{(2)}, U_{i,4}^{(2)} \sim \text{Uniform}(0, 30)$, $U_{i,2}^{(1)}, U_{i,4}^{(1)} \sim \text{Uniform}(0, 2)$, and $\varepsilon_{i,1}(t), \varepsilon_{i,2}(t), \varepsilon_{i,4}(t) \sim N(0, 20^2)$. All of the signals are sampled at time $t = 1, 2, \ldots, n_i$, where $n_i = \lfloor \tau_i \rfloor$ is the largest integer less or equal to τ_i. Figure 1 shows the true HI and four signals for three randomly generated units. In Figure 1, significant differences are observed among these units.

To verify the convergence property of the proposed method, we randomly select m historical units and estimate the fusion coefficient $\boldsymbol{w}_0$ in both Case I ($\mathbf{\Omega}_0$ known) and Case II ($\mathbf{\Omega}_0$ unknown) based on the sensor signals and failure time while the true HI is regarded as unknown. We repeat the above procedure 50 times to obtain the mean and standard deviation of our estimations. Figure 2 shows the results for Case I, where the x-axis is the number of historical units sampled for estimation, the solid horizontal lines represent the true value for each entry of $\boldsymbol{w}_0$, and the solid and dashed curves represent the mean and one standard deviation of the estimated values, respectively. As we can see, the estimation becomes more accurate with more historical units, since there is more information available that can reduce the variance of the estimated fusion coefficient as implied by Equation (18). Moreover, the performance of our proposed method is satisfactory, as all true values are within one standard deviation from the mean of the estimated values.

For Case II, the iterative procedure proposed in Sections 3.6 is used to estimate $\boldsymbol{w}_0$. For each random sample set of historical

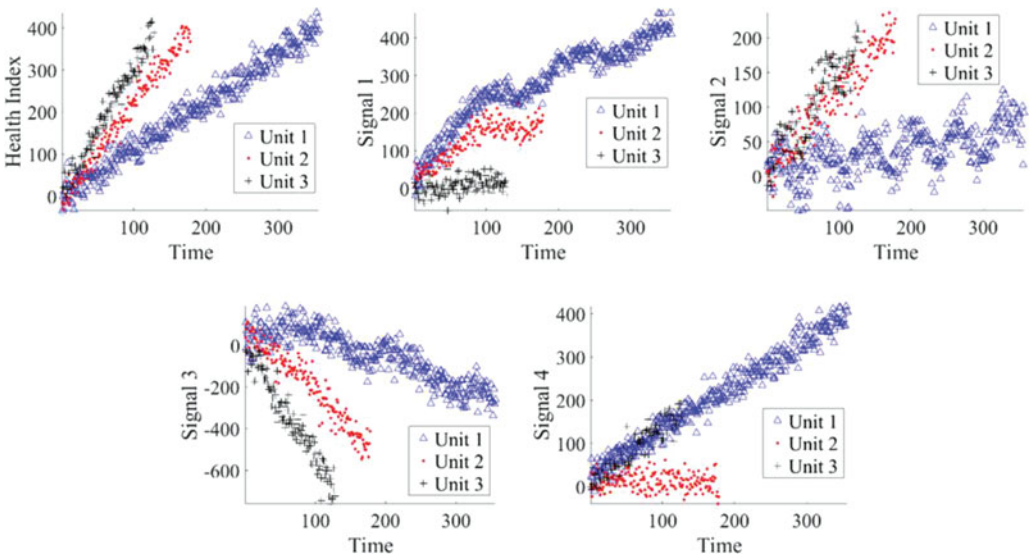

Figure 1. True HI and four signals for three randomly generated units.

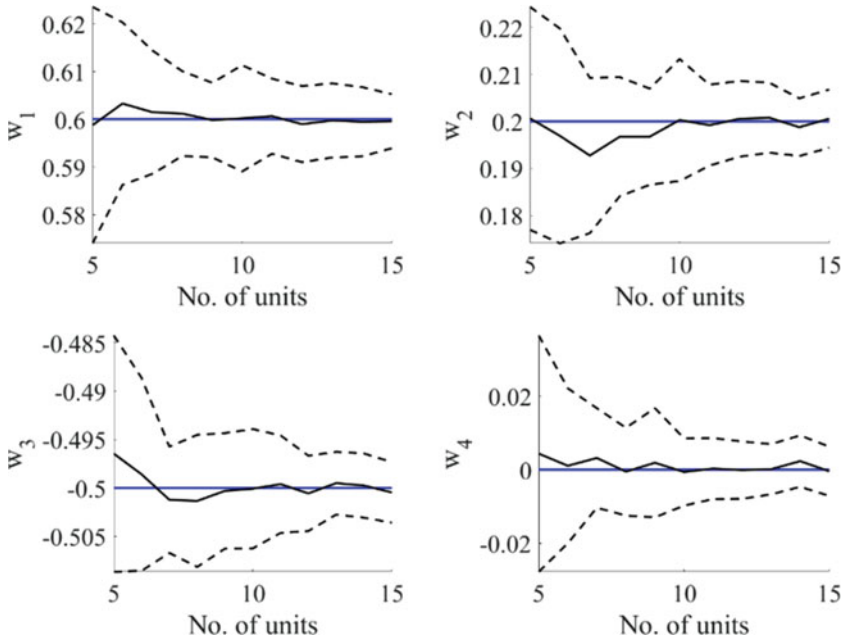

Figure 2. Estimation results for Case I. The solid line and dashed lines are the mean and one standard deviation for each entry of $\hat{\boldsymbol{w}}$, respectively. The solid horizontal line is the true value of $\boldsymbol{w}_0$.

units, we repeat the iterative procedure 50 times with 50 different initial guesses of $\boldsymbol{w}^{(0)}$. The first initial guess is obtained by solving Equation (19) given $k = 0$ and noninformative prior parameters $[\boldsymbol{\Sigma}^{(1)}]^{-1} = 0$, and the remaining 49 initial guesses are randomly generated and satisfy the constraint $\boldsymbol{w}^{(0)} \in \mathcal{W}_{\boldsymbol{\Omega}^{(0)}}$. We call the first initial guess the *derived initial guess* and the remaining 49 as *random initial guess*. The final estimation is obtained by selecting the optimal solution with the minimum objective value among the 50 repetitions. The results of Case II are shown in Figure 3. Similar to Case I, the horizontal lines represent the true value for each entry of $\boldsymbol{w}_0$, and the solid and dashed curves are the mean and one standard deviation of the estimations, respectively. As we can see from this figure, the performance of our proposed method is also satisfactory in Case II. The results of Figures 2 and 3 further validate the benefit of indirect supervised learning, as in our proposed method the estimation converges to the true value. On the contrary, for existing data fusion methods that formulate the problem in two steps, there is no guarantee of such a convergence.

We are also interested in understanding the convergence of the iterative procedure. Specifically, we want to answer (i) how many repetitions does the iterative procedure fail to converge; and (ii) how many iterations have to be performed for the iterative procedure to converge. Figure 4 illustrates the proportion of repetitions that the iterative procedure with a random initial guess $\boldsymbol{w}^{(0)}$ fails to converge. Generally, the proportion of all

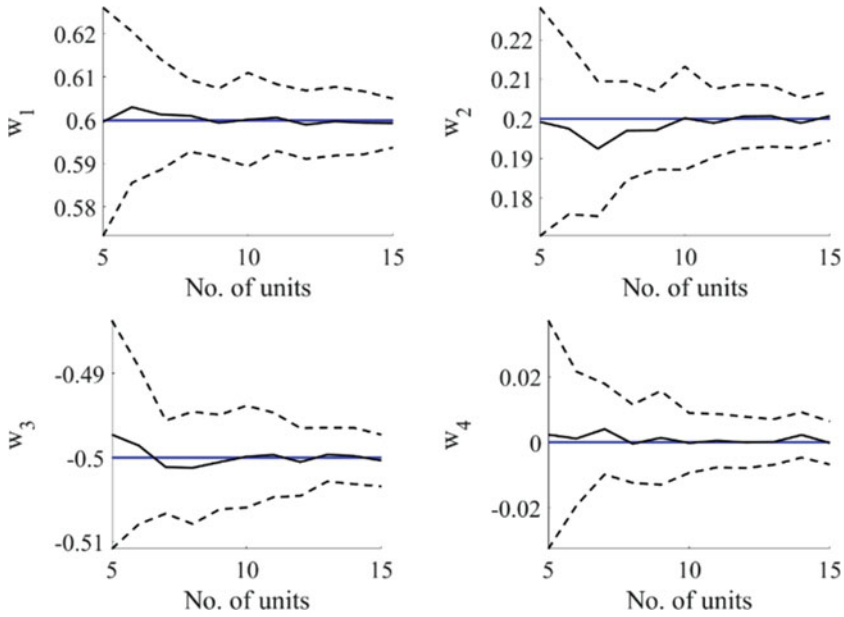

Figure 3. Estimation results for Case II. The solid lines and dashed lines are the mean and one standard deviation for each entry of $\hat{\boldsymbol{w}}$, respectively. The solid horizontal line is the true value of $\boldsymbol{w}_0$.

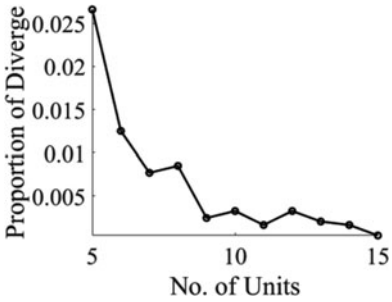

Figure 4. Proportion of the repetitions of the iterative procedure with random initial guesses that fail to converge.

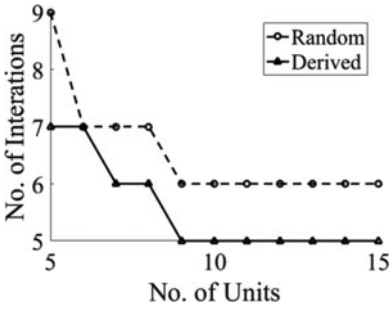

Figure 5. The 90% quantile of the number of iterations it takes the iterative procedure with either a random or a derived initial guess to converge.

repetitions failing to converge decreases as more historical units are available, which agrees with our expectation as discussed in Sections 3.6. Even with only six historical units, less than 1.5% of the repetitions fail to converge. Moreover, note that in Figure 4, the initial guesses of $\boldsymbol{w}^{(0)}$ in these repetitions are all randomly generated. In fact, in this study, we observe that the repetitions with derived initial guesses all converge, indicating that a good initial guess can increase the probability of convergence for the iterative procedure.

Figure 5 shows the 90% quantile of the number of iterations that the iterative procedure with random or derived initial guesses takes to converge. With as few as five historical units, 90% of repetitions of the iterative procedure converge within nine iterations, indicating that the iterative procedure is able to quickly converge. Furthermore, the derived initial guesses are observed to be a little better than the random initial guesses, with one or two fewer iterations.

4.2. Sensitivity to sparse data

The second scenario considered is sparse data, the case where the number of available sensor observations for each unit is small. As discussed in Sections 3.6, the iterative procedure for Case II performs better with more sensor observations for each unit. However, in practice, there may be only sparse data available due to missing data or limited resources. Thus, it is necessary to understand the level of the required number of sensor observations to ensure a satisfactory performance for the iterative procedure. In this simulation, we use the same dataset as generated in Sections 4.1 and randomly choose 10 historical units as the training set. Then for each unit, we randomly sample a number of sensor observations $\boldsymbol{L}_i(t)$ at different t. Our proposed method is used to estimate the fusion coefficient $\boldsymbol{w}_0$ based on the selected sensor observations. By changing the number of selected sensor observations for each unit, we are able to assess our proposed method with different levels of data availability. The procedure described above is repeated 50 times and the results for Case II are shown in Figure 6, where the x-axis represents the number of available observations for each historical unit, the solid horizontal lines denote the true values for each entry of $\boldsymbol{w}_0$, and the solid and dashed curves are the mean and one standard deviation of the estimation, respectively. The results obtained for Case I are very similar and thus not shown here. It is clear that the estimation becomes more accurate as more observations become available.

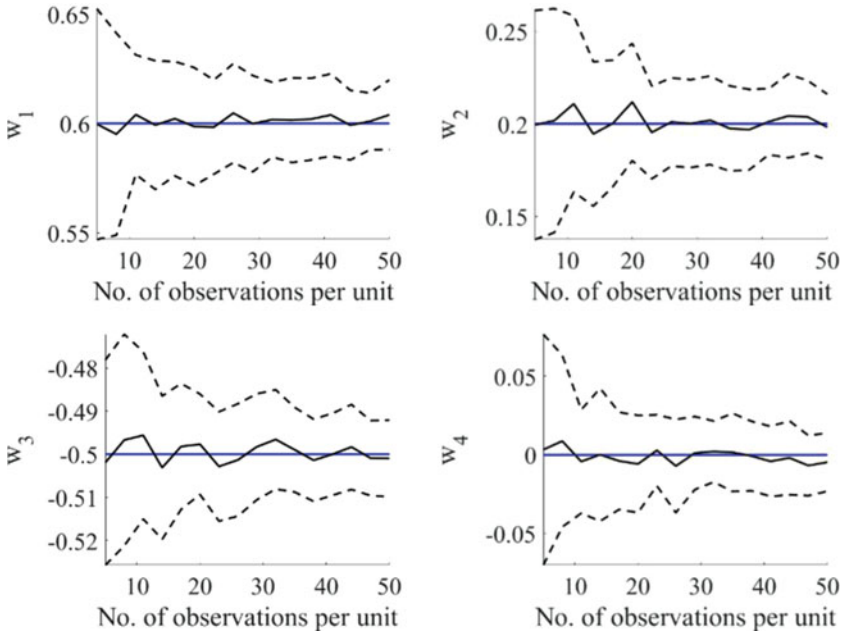

Figure 6. Estimation results with sparse data for Case II. The solid lines and dashed curves are the mean and one standard deviation for each entry of the estimated $\boldsymbol{w}_0$, respectively. The solid horizontal line is the true value of $\boldsymbol{w}_0$.

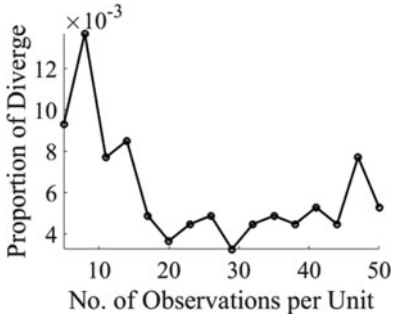

Figure 7. Proportion of the repetitions of the iterative procedure with random initial guesses that fail to converge.

In addition to the estimation performance, we are interested in the convergence of the iterative procedure with sparse data. Figure 7 shows the proportion of repetitions that fail to converge for the iterative procedure with random initial guesses. As Figure 7 shows, even with only eight observations per unit, the proportion of repetitions that fail to converge is only around 1.4%. Figure 8 shows the 90% quantile of the number of iterations that the iterative procedure with random or derived initial guesses takes to converge. The repetitions with random and derived initial guesses are close in the number of iterations. Furthermore, with only eight observations per unit, 90% of the repetitions converge within 13 iterations. Figures 7 and 8 show that the iterative procedure is able to converge quickly with only a small or moderate number of available sensor observations.

4.3. Sensitivity to mis-specification of the general path model

The last scenario we consider concerns the mis-specification of the general path model. In Sections 3.2, we assumed that the true degradation path model was known. However, in some applications, the general path model may be mis-specified, due to the lack of prior knowledge and historical data. If this happens, our proposed method is not guaranteed to converge to the true fusion coefficient. In this simulation, we focus on the prognostic performance of the constructed HI when the general path model is mis-specified. Specifically, when generating the simulated dataset, we set the underlying degradation path to be

$$\eta(t; \mathbf{\Gamma}_i, \mathbf{\Lambda}) = \Gamma_{i,0} + \Gamma_{i,1} t^{1.6},$$

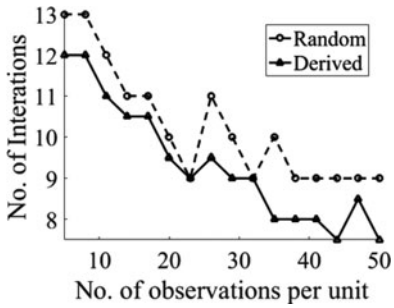

Figure 8. The 90% quantile of the number of iterations it takes the iterative procedure with either a random or a derived initial guess to converge.

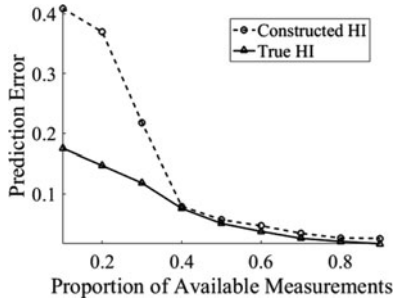

Figure 9. Average prediction error with different proportions of available measurements.

but when estimating the fusion coefficient and predicting the RUL, we still assume the underlying degradation path to be linear. Specifically, we randomly draw 50 units as the training set to estimate the optimal fusion coefficient. Then we randomly draw another 50 units as the testing set. The sensor signals in the testing set are truncated before the actual failure time. Based on the constructed HI for the testing set, we predict the RUL for each testing unit and compare with the actual RUL. For testing unit i, the prediction error is defined as

$$\text{err}_i = \frac{|\hat{T}_i - \tau_i|}{\tau_i}, \tag{24}$$

where $\hat{T}_i$ is the predicted failure time of unit i, and τ_i is the actual failure time of unit i. The procedure is repeated 50 times and the result is shown in Figure 9. The x-axis represents different proportions of available measurements; e.g., "0.4" means that the available signals for each testing unit i are truncated at time $0.4\tau_i$. The curves are the average prediction errors where for "True HI," the true HI and the true degradation path model are used for prognosis, whereas for "Constructed HI," the mis-specified degradation path model is used and the fusion coefficient is derived as Case II with unknown prior parameters. It is interesting to see that the prognostic performance of the constructed HI is very close to the true HI when the proportion of available measurements is moderate or large. One possible reason for this observed behavior is that although the degradation path model is mis-specified, our proposed method is able to derive a biased fusion coefficient such that the constructed HI best fits the mis-specified degradation path model. Since all measurements of historical units are utilized in parameter estimation, the constructed HI will provide a better fit when a moderate or large proportion of measurements have been collected. This study shows that the proposed method could be very useful as the true degradation model is usually unknown and the estimated degradation model may be mis-specified in practice.

5. Application

A case study is also conducted to evaluate the proposed method in situations where the basic truth such as the true fusion function and the true general path model are unknown and the assumptions of the proposed method may be violated. This represents a typical practical situation when implementing the proposed method in real applications. In this case study, we focus

Table 1. Detailed description of the 21 sensors (Saxena *et al.*, 2008).

Symbol	*Description*	*Units*
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	psia
P15	Total pressure in bypass duct	psia
P30	Total pressure at HPC outlet	psia
Nf	Physical fan speed	rpm
Nc	Physical core speed	rpm
epr	Engine pressure ratio (P50/P2)	—
Ps30	Static pressure at HPC outlet	psia
phi	Ratio of fuel flow to Ps30	pps/psi
NRf	Corrected fan speed	rpm
NRc	Corrected core speed	rpm
BPR	Bypass ratio	—
farB	Burner fuel–air ratio	—
htBleed	Bleed enthalpy	—
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s

on modeling and predicting the degradation of aircraft turbofan engines.

5.1. Dataset description

The dataset was generated by C-MAPSS, which is a widely used simulator for studying the degradation process of large commercial turbofan engines (Saxena *et al.*, 2008; Sarkar *et al.*, 2011). In particular, our dataset consists of 21 sensor signals of turbofan engines under a single failure mode and a single environmental condition. The signals are collected simultaneously and continuously at time $t = 1, 2, \ldots, n_i$, and contain detailed information such as the temperature and pressure of each engine i. A detailed description of these sensors is given in Table 1. There are 100 historical units (i.e., $m = 100$) with a total of 20 631 observations (i.e., $\sum_{i=1}^{m} n_i = 20\,631$) and 100 in-field units with 13 096 observations. For the historical units, the dataset contains all of the measurements from run to failure, whereas for in-field units, measurements are only available up to a time point before failure and the goal is to predict the RULs of these in-field units based on the available measurements. The true RULs are provided in a separated file; in this way, we are able to evaluate the performance of our proposed method by comparing the predicted RULs with the true ones. The dataset is available online (Saxena and Goebel, 2008).

5.2. Results and comparison

To the best of our knowledge, among the HI-based approaches, Liu *et al.* (2017) studied the same dataset and achieved one of the best currently existing prognostic results. Specifically, Liu *et al.* (2017) proposed a metric SNR[d] (signal-to-noise ratio) for measuring the quality of the degradation signals. Then they formulated the problem in two steps as described in Section 2.2, where in the first step, they combined the multiple sensor signals in such a way that the "quality" of the constructed HI was maximized and then, in the second step, the RUL predictions were conducted based on the constructed HI. They also showed that their method outperformed the prognostic result based on any single original sensor signal. Therefore, in this study, we choose to compare the proposed method with the SNR-based method. We follow the same data preprocessing procedures for sensor selection and data transformation and use the same general path model as in Liu *et al.* (2017) to provide a basis for a fair comparison. Specifically, we select 11 sensors out of the 21 sensors based on the criterion that the sensor signals should consistently exhibit an increasing or decreasing trend in all historical units. In other words, a sensor is selected if its last measurement is consistently larger (increasing trend) or smaller (decreasing trend) than the first measurement in all historical units. As a result, the selected sensors are T24, T50, P30, Nf, Ps30, phi, NRf, BPR, htBleed, W31, and W32. Then these sensors are transformed by a logarithm operation and then standardized in the same way as in Liu *et al.* (2017). Likewise, the quadratic general path model (i.e., $\boldsymbol{\psi}(t) = [1, t, t^2]$) is adopted here to describe the underlying degradation process, as it provides a good fit according to Liu *et al.* (2017).

After preprocessing of the dataset, we implement the iterative procedure for Case II with the failure threshold arbitrarily set to a value of two. In addition, we relax the constraints in Equation (19) to test the flexibility of our method and ensure that the search space $\mathcal{W}_{\boldsymbol{\Omega}^{(k+1)}}$ is not empty. Specifically, the original constraint $\dot{\boldsymbol{\psi}}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}^{(k)}; \hat{\boldsymbol{\Omega}}(\boldsymbol{w}^{(k)})) \geq C_2$ requires the degradation path to be strictly increasing, which is equivalent to that the second and third entries of $\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}^{(k)}; \hat{\boldsymbol{\Omega}}(\boldsymbol{w}^{(k)}))$ being strictly positive, as $\boldsymbol{\psi}(t) = [1, t, t^2]$. Instead, we relax this constraint and only require the third entry to be strictly positive in this case study, i.e., $\ddot{\boldsymbol{\psi}}(t)\boldsymbol{\mu}_i(\boldsymbol{X}_i, \boldsymbol{w}^{(k)}; \hat{\boldsymbol{\Omega}}(\boldsymbol{w}^{(k)})) \geq C_2$, where $\ddot{\boldsymbol{\psi}}(t) = \partial^2\boldsymbol{\psi}(t)/\partial t^2$. In other words, we allow the degradation path represented by the HI to be stable at the beginning of the lifecycle of the unit, due to the influence of the measurement errors. The estimated weight fusion coefficient is shown in Table 2.

First, we visually compare the 11 sensor signals with the HI constructed by our proposed method in Figure 10. The dots denote signal measurements and the curve denotes the degradation process fitted by the quadratic model. As this figure shows, the constructed HI provides a clearer degradation trajectory than any other sensor signal.

Figure 11 compares the prognostic results of the in-field units using our proposed method with the SNR-based method (HI-SNR) proposed by Liu *et al.* (2017), where the x-axis represents different levels of actual RUL. For example, "80" means that only in-field units with an actual RUL less or equal to 80 are considered, and "all" means that all in-field units are considered. The curves refer to the average of the prediction errors as defined in Equation (24), with the error bars being equal to one standard deviation of the average predictor error. The curves are not strictly aligned to the labels of the x-axis to facilitate visualization. According to Figure 11, our proposed method consistently creates lower average prediction errors with a lower

Table 2. Estimated fusion coefficient $\hat{\boldsymbol{w}}$ for the selected sensors.

	Sensor										
	T24	*T50*	*P30*	*Nf*	*Ps30*	*phi*	*NRf*	*BPR*	*htBleed*	*W31*	*W32*
Coef.	0.13	0.37	−0.03	−0.05	0.23	−0.21	−0.08	0.16	0.12	−0.05	−0.16

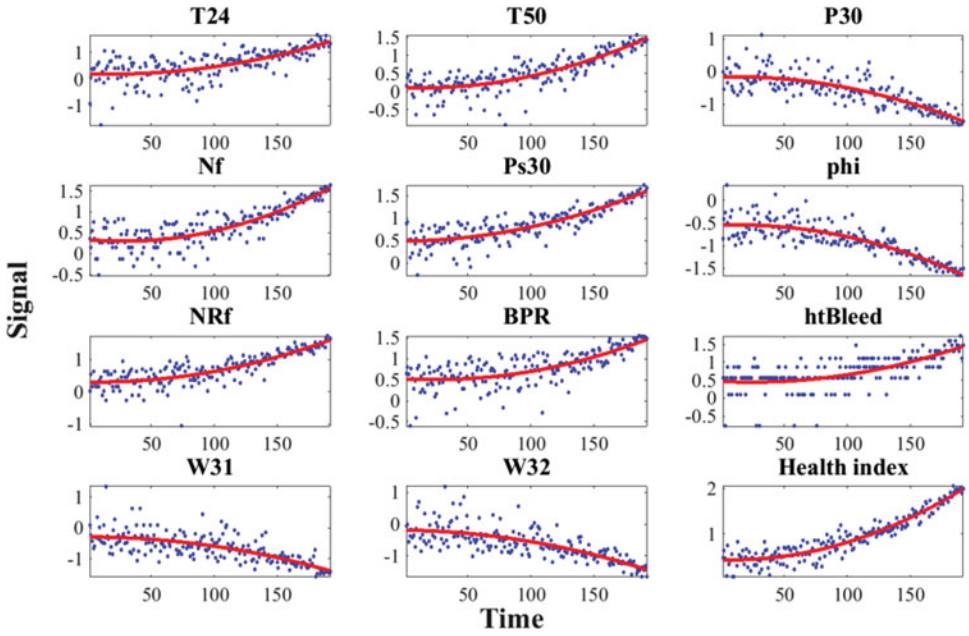

Figure 10. Comparison of the 11 sensor signals and the HI constructed by the proposed method for a historical unit. The dots denote measurements and the curve denotes the degradation process fitted using a quadratic model.

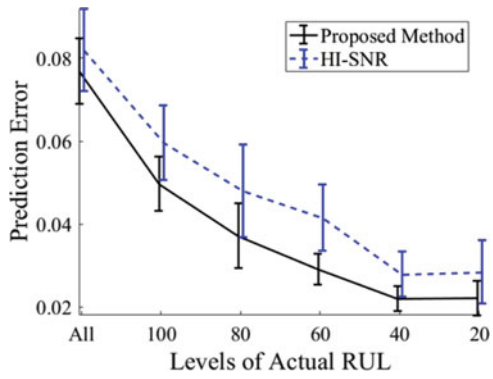

Figure 11. Comparison of the RUL prediction errors of the in-field units using the SNR-based method and our proposed method.

variation at different levels of actual RUL compared with the SNR-based method.

6. Conclusion

To prevent unexpected failures of critical units, sensors are widely used to continuously monitor the underlying degradation process and infer the health status of these units in real time. Although the statistical modeling of a single degradation signal has been extensively studied in the literature, the information contained in a single sensor signal may not be sufficient to characterize the underlying degradation process and produce an accurate RUL prediction, especially for complex systems. As a result, multiple sensors are often used in practice to simultaneously monitor a unit and thus there is a pressing need to develop an effective data fusion method tailored for degradation modeling and prognostics of multiple sensor signals. In this article, we aim to address this issue by proposing a novel HI-based method.

The critical challenges in constructing the HI mainly lie in that the underlying degradation process is unobservable, and the constructed HI must satisfy the engineering knowledge extracted from the underlying degradation process. We call this kind of problem indirect supervised learning. In this article, we assumed that a number of historical units with known failure time have been acquired and then proposed the minimization of the difference between the predicted failure time and the true failure time based on the historical units. Particularly, our innovative idea is to formulate the construction of HI using the quantile regression technique and then realize the extracted engineering knowledge from the underlying degradation process, such as monotonicity and failure threshold, by imposing different constraints in the proposed method. Contrary to existing studies that tackle the problem in two separate steps, we solved the problem as an integrated procedure, which thus ensures the prognostic performance of our constructed HI.

One of the major contributions of this study is that the proposed method ensures the convergence of the estimated fusion coefficient $\hat{\boldsymbol{w}}$ to its true value. To the best of our knowledge, this is the first article that provides such a theoretical analysis, which is expected to stimulate follow-up studies. A number of simulation studies were conducted to evaluate the sensitivity and prognostic performance of the proposed method under different scenarios. In addition, the proposed method was further tested by a case study that involved the degradation of aircraft turbofan engines when the underlying truth is unknown. The results showed that the performance of the proposed method was satisfactory in both the simulation and case studies and also superior to the related existing benchmark method.

There are several topics worthy of further study. First, the asymptotic properties of the proposed method are based on the assumption of a general path model for the degradation process. More studies should be conducted when other degradation models, such as Wiener process or gamma process, are used. Second, when multiple sensors are used for condition monitoring, they may exhibit complex relations and some sensor signals may not be related to the underlying degradation process. How to automatically screen out non-informative sensor signals

and extend the linear fusion function (such as via kernel methods) would be an interesting problem for investigation. Third, this study focuses on degradation modeling under a single failure mode and a single operation condition, whereas in practice, the situation may be more complex. Therefore, more efforts are needed to consider multiple failure modes and also different operation conditions.

Funding

This work was supported in part by the Office of Naval Research under grant N00014-17-1-2261.

Notes on contributors

Changyue Song received a B.S. degree in industrial engineering from Tsinghua University in 2012 and an M.S. degree in industrial engineering from Peking University in 2015. Currently he is a Ph.D. student at the Department of Industrial and Systems Engineering, University of Wisconsin–Madison. His research interests are focused on statistical modeling of complex systems.

Kaibo Liu received a B.S. degree in industrial engineering and engineering management from the Hong Kong University of Science and Technology in 2009 and an M.S. degree in statistics and Ph.D. degree in industrial engineering from the Georgia Institute of Technology in 2011 and 2013, respectively. Currently, he is an assistant professor at the Department of Industrial and Systems Engineering, University of Wisconsin–Madison. His research interests are focused on data fusion for process modeling, monitoring, diagnosis, and prognostics. He is a member of IEEE, ASQ, INFORMS, and IISE.

References

Bae, S.J., Kuo, W. and Kvam, P.H. (2007) Degradation models and implied lifetime distributions. *Reliability Engineering and System Safety*, **92**(5), 601–608.

Bae, S.J. and Kvam, P.H. (2004) A nonlinear random-coefficients model for degradation testing. *Technometrics*, **46**(4), 460–469.

Bagdonavicius, V. and Nikulin, M.S. (2001) Estimation in degradation models with explanatory variables. *Lifetime Data Analysis*, **7**(1), 85–103.

Baraldi, P., Mangili, F. and Zio, E. (2012) A Kalman filter-based ensemble approach with application to turbine creep prognostics. *IEEE Transactions on Reliability*, **61**(4), 966–977.

Brotherton, T., Grabill, P., Wroblewski, D., Friend, R., Sotomayer, B. and Berry, J. (2002) A testbed for data fusion for engine diagnostics and prognostics, in *Proceedings of IEEE Aerospace Conference 2002*, IEEE Press, Piscataway, NJ, pp. 3029–3042.

Chapelle, O., Schölkopf, B. and Zien, A. (2006) *Semi-Supervised Learning*, MIT Press, Cambridge, MA.

Chen, N. and Tsui, K.L. (2013) Condition monitoring and remaining useful life prediction using degradation signals: Revisited. *IIE Transactions*, **45**(9), 939–952.

Christer, A.H., Wang, W. and Sharp, J.M. (1997) A state space condition monitoring model for furnace erosion prediction and replacement. *European Journal of Operation Research*, **101**(1), 1–14.

Fang, X., Gebraeel, N. and Paynabar, K. (2017) Scalable prognostic models for large-scale condition monitoring applications. *IISE Transactions*, **49**(7), 698–710.

Gebraeel, N. Z. (2006) Sensor-updated residual life distributions for components with exponential degradation patterns. *IEEE Transactions on Automation Science and Engineering*, **3**(4), 382–393.

Gebraeel, N.Z., Lawley, M.A., Li, R. and Ryan, J.K. (2005) Resitual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, **37**(6), 543–557.

Hall, D.L. and Llinas, J. (1997) An introduction to multisensor data fusion. *Proceedings of the IEEE*, **85**, 6–23.

Hastie, T., Tibshirani, R. and Friedman, J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition, Springer, New York, NY.

Hong, Y., Duan, Y., Meeker, W.Q., Stanley, D.L. and Gu, X. (2015) Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data. *Technometrics*, **57**(2), 180–193.

Hu, C., Youn, B.D., Wang, P. and Yoon, J.T. (2012) Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering and System Safety*, **103**, 120–135.

Jardine, A.K., Lin, D. and Banjevic, D. (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, **20**(7), 1483–1510.

Koenker, R. (2005) *Quantile Regression*, Cambridge University Press, Cambridge, UK.

Lawless, J. and Crowder, M. (2004) Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, **10**(3), 213–227.

Liu, K., Chehade, A. and Song, C. (2017) Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis. *IEEE Transactions on Automation Science and Engineering*, **14**(3), 1504–1514.

Liu, K., Gebraeel, N.Z. and Shi, J. (2013) A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. *IEEE Transactions on Automation Science and Engineering*, **10**(3), 652–664.

Liu, K. and Huang, S. (2016) Integration of data fusion methodology and degradation modeling process to improve prognostics. *IEEE Transactions on Automation Science and Engineering*, **13**(1), 344–354.

Liu, K. and Shi, J. (2015) Internet of things (IoT)-enabled system informatics for service decision making: Achievements, trends, challenges, and opportunities. *IEEE Intelligent Systems*, **30**(6), 18–21.

Loutas, T.H., Roulias, D. and Georgoulas, G. (2013) Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression. *IEEE Transactions on Reliability*, **62**(4), 821–832.

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

Lu, C.J., Park, J. and Yang, Q. (1997) Statistical inference of a time-to-failure distribution derived from linear degradation data. *Technometrics*, **39**(4), 391–400.

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

Nelson, W. (1990) *Accelerated Testing Statistical Models, Test Plans and Data Analysis*, Wiley, New York, NY.

Nocedal, J. and Wright, S. (2006) *Numerical Optimization*, Springer, New York, NY.

Oberhofer, W. (1982) The consistency of nonlinear regression minimizing the L_1-norm. *The Annals of Statistics*, **10**(1), 316–319.

Park, J.I. and Bae, S.J. (2010) Direct prediction methods on lifetime distribution of organic light-emitting diodes from accelerated degradation tests. *IEEE Transactions on Reliability*, **59**(1), 74–90.

Saha, B., Goebel, K. and Christophersen, J. (2009) Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*, **31**(3–4), 293–308.

Sarkar, S., Jin, X. and Ray, A. (2011) Data-driven fault detection in aircraft engines with noisy sensor measurements. *Journal of Engineering for Gas Turbines and Power*, **133**(8), 081602.

Saxena, A. and Goebel, K. (2008) Turbofan engine degradation simulation data set, NASA Ames Research Center, Moffett Field, CA. Available at: (https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository) [Accessed 30 April 2018].

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

Si, X.S., Wang, W., Hu, C.H. and Zhou, D.H. (2011) Remaining useful life estimation—A review on the statistical data driven approaches. *European Journal of Operation Research*, **213**(1), 1–14.

Si, X.S., Wang, W., Hu, C.H., Zhou, D.H. and Pecht, M.G. (2012) Remaining useful life estimation based on a nonlinear diffusion degradation process. *IEEE Transactions on Reliability*, **61**(1), 50–67.
Tian, Z. (2012) An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring. *Journal of Intelligent Manufacturing*, **23**(2), 227–237.
Vaisnys, P., Contri, P., Rieg, C. and Bieth, M. (2006) Monitoring the effectiveness of maintenance programs through the use of performance indicators. Report EUR 22602, DG JRC – Institute for Energy, Petten, The Netherlands.
Wang, T., Yu, J., Siegel, D. and Lee, J. (2008) A similarity-based prognostics approach for remaining useful life estimation of engineered system, in *Proceedings of the International Conference on Prognostics and Health Management 2008*, IEEE, Denver, CO.
Wang, X. and Xu, D. (2010) An inverse Gaussian process model for degradation data. *Technometrics*, **52**(2), 188–197.
Whitmore, G.A. and Schenkelberg, F. (1997) Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Analysis*, **3**(1), 27–45.
Xu, Z., Ji, Y. and Zhou, D. (2008) Real-time reliability prediction for a dynamic system based on the hidden degradation process identification. *IEEE Transactions on Reliability*, **57**(2), 230–242.
Yang, F., Habihullah, M.S., Zhang, T., Xu, Z., Lim, P. and Nadarajan, S. (2016) Health index-based prognostics for remaining useful life predictions in electrical machines. *IEEE Transactions on Industrial Electronics*, **63**(4), 2633–2644.
Ye, Z.S. and Chen, N. (2014) The inverse Gaussian process as a degradation model. *Technometrics*, **56**(3), 302–311.
Ye, Z.S., Wang, Y., Tsui, K.L. and Pecht, M. (2013) Degradation data analysis using Wiener process with measurement errors. *IEEE Transactions on Reliability*, **62**(4), 772–780.
Ye, Z.S., and Xie, M. (2015) Stochastic modeling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry*, **31**(1), 16–32.
Yu, H. (2006) Designing an accelerated degradation experiment with a reciprocal Weibull degradation rate. *Journal of Statistical Planning and Inference*, **136**(1), 282–297.
Zhai, Q. and Ye, Z.S. (2017) RUL prediction of deteriorating products using an adaptive Wiener process model. *IEEE Transactions on Industrial Informatics*, **13**(6), 2911–2921.
Zhou, Q., Son, J., Zhou, S., Mao, X. and Salman, M. (2014) Remaining useful life prediction of individual units subject to hard failure. *IIE Transactions*, **46**(10), 1017–1030.
Zhou, R., Serban, N. and Gebraeel, N. (2014) Degradation-based residual life prediction under different environments. *The Annals of Applied Statistics*, **8**(3), 1671–1689.
Zhou, R., Serban, N., Gebraeel, N. and Müller, H.G. (2014) A functional time warping approach to modeling and monitoring truncated degradation signals. *Technometrics*, **56**(1), 67–77.

Appendix

Proof of Proposition 1

Since $E[\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)] = \boldsymbol{\mu}_0$, obviously $\hat{\boldsymbol{\mu}}(\boldsymbol{w}_0)$ converges to $\boldsymbol{\mu}_0$ according to the law of large numbers. Conditioned on $\boldsymbol{\Gamma}_i$, it is well known that the expectation

$$E\left[\frac{\left(\boldsymbol{L}_i\boldsymbol{w}_0 - \boldsymbol{\Psi}_i\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)\right)^T\left(\boldsymbol{L}_i\boldsymbol{w}_0 - \boldsymbol{\Psi}_i\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)\right)}{n_i - p}\Big|\boldsymbol{\Gamma}_i\right] = \sigma_0^2,$$

which is a constant, and thus the unconditioned expectation is also σ_0^2. Thus, $\hat{\sigma}^2(\boldsymbol{w}_0)$ converges to σ_0^2. From now on, we treat $\boldsymbol{X}_{i,1} = \boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i$ as random. Since

$$Var[\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)|\boldsymbol{X}_{i,1}] = \boldsymbol{\Sigma}_0 + \sigma_0^2(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1},$$

we get

$$Var[\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)] = \boldsymbol{\Sigma}_0 + E\left[\sigma_0^2(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\right] + Var[E(\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)|\boldsymbol{X}_{i,1})].$$

Since $E(\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)|\boldsymbol{X}_{i,1}) = \boldsymbol{\mu}_0$, which means $Var[E(\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)|\boldsymbol{X}_{i,1})] = 0$, we get

$$\boldsymbol{\Sigma}_0 = Var[\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)] - E\left[\sigma_0^2\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\right].$$

As a result,

$$\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0) - \hat{\boldsymbol{\mu}}(\boldsymbol{w}_0))(\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0) - \hat{\boldsymbol{\mu}}(\boldsymbol{w}_0))^T \to Var[\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)]$$

and

$$\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n_i - p}(\boldsymbol{L}_i\boldsymbol{w}_0 - \boldsymbol{\Psi}_i\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0))^T(\boldsymbol{L}_i\boldsymbol{w}_0 - \boldsymbol{\Psi}_i\hat{\boldsymbol{\Gamma}}_i(\boldsymbol{w}_0)) \times \left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1} \to E\left[\sigma_0^2\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\right],$$

and thus $\hat{\boldsymbol{\Sigma}}(\boldsymbol{w}_0) \to \boldsymbol{\Sigma}_0$. This completes the proof. □