A Generic Health Index Approach for Multisensor Degradation Modeling and Sensor Selection

Minhee Kim, Changyue Song, and Kaibo Liu, *Member, IEEE*

***Abstract*— With recent development in sensor technology, multiple sensors have been widely adopted to monitor the degradation of a single unit simultaneously. The challenge of multisensor degradation modeling lies in that the sensor signals are often correlated and may contain only partial or even no information on the degradation status of a unit. To address these issues, this paper proposes a novel data fusion method that constructs a 1-D health index (HI) via automatically selecting and combining multiple sensor signals to better characterize the degradation process. In particular, this paper develops a new latent linear model that constructs the HI and selects informative sensors in a unified manner. Compared to the existing literature, the proposed method enjoys several unique advantages: 1) being able to derive the best linear unbiased estimator of the fusion coefficients; 2) offering high computational efficiency; 3) not requiring to know the exact value of the failure threshold; and 4) exhibiting general applicability in practice by not imposing restrictive assumptions on the degradation process. Simulation studies are presented to illustrate the effectiveness and evaluate the sensitivity of the proposed method. A case study on the degradation of aircraft gas turbine engines is also performed which shows a better prognostic performance of the proposed method compared with existing approaches.**

***Note to Practitioners*—This paper is motivated by the practical issue of degradation modeling and prognostics when multiple sensors simultaneously monitor the degradation status of a unit. Specifically, there are two fundamental questions involved, including: 1) how to screen out noninformative sensors and 2) how to properly combine the information from the selected sensor signals to accurately estimate the underlying degradation status of the unit. The novelty of this paper lies in developing an innovative latent model that tackles these two challenging questions in an integrated manner. There are four main steps involved when implementing the proposed method: 1) collecting multiple sensor signals and failure time of historical units; 2) selecting the informative sensors and deriving the optimal weight for each selected sensor; 3) constructing the health indices (HIs) of in-service units; and 4) predicting the remaining useful life of the in-service units using the constructed HIs. The proposed method is very useful when the degradation is under a single failure mode in a single environmental condition. In the future research, we will study the extension of the proposed model when sensor signals have a nonlinear relationship, as well as when the degradation process is under more complex scenarios such as multiple failure modes and multiple operation conditions.**

***Index Terms*— Data fusion, degradation modeling, health index (HI), multisensor, predictive data analysis, prognostics.**

Manuscript received October 20, 2018; accepted December 27, 2018. Date of publication January 15, 2019; date of current version July 1, 2019. This paper was recommended for publication by Associate Editor S. Yin and Editor F.-T. Cheng upon evaluation of the reviewers' comments. This work was supported in part by the Office of Naval Research under Grant N00014-17-1-2261 and in part by the Department of Energy under Award DE-NE0008805. *(Corresponding author: Kaibo Liu.)*

The authors are with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: mkim555@wsic.edu; csong39@wisc.edu; kliu8@wsic.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2018.2890608

I. Introduction

DEGRADATION is quite common in engineering systems and will eventually lead to failures. Unexpected failures can cause production downtime, poor customer satisfaction, safety issues, etc. To avoid such losses, sensors have been widely used to monitor the degradation process of a unit. The collected sensor signals contain useful information about the degradation status of the unit, which if properly used, can lead to accurate prediction of the remaining useful life (RUL).

Most of the existing literature focuses on analyzing a single sensor signal [1], and there are two commonly used approaches, including general path models [2], [3] and stochastic process models [4]–[9]. General path models formulate the sensor signal using a random-effect model, where the random-effect parameters are used to capture the unit-to-unit variability. On the other hand, stochastic process models characterize the evolution of a sensor signal as a stochastic process, e.g., Wiener process [4]–[6], inverse Gaussian process [7], and gamma process [8], [9], to account for the temporal variation of sensor signals.

Unfortunately, these approaches are only effective under the assumptions that the physical degradation mechanism of a monitored unit is well understood, and thus a single sensor is sufficient to fully characterize the underlying degradation process. However, in reality, it is common that a single sensor only contains partial information on the degradation process.

In order to overcome this issue, much attention has been recently focused on using multiple sensors to monitor a single unit simultaneously. In this way, different aspects of the degradation process can be captured [10]. Therefore, there is a growing need to develop efficient multisensor degradation modeling approaches. However, different sensor signals usually have different levels of relevance to the degradation process. In many real-world applications, it is even possible that some sensors are unrelated to the underlying degradation process, which compromises the accuracy of RUL prediction by acting as noise. In addition, a collection of these noninformative sensor signals may incur unnecessary costs. As a result, there are two key challenging questions involved in the multisensor degradation modeling: 1) how to

screen out noninformative sensors and 2) how to properly combine the information from the selected sensor signals to accurately estimate the underlying degradation status of a unit.

To address these challenges, this paper presents a novel health index (HI)-based data fusion model for multisensor degradation modeling and sensor selection. In particular, we combine the observable data, i.e., the failure time and the multiple sensor signals, via a novel latent linear model to accurately characterize the unobservable underlying degradation status of the unit. Consequently, the contributions of this paper are summarized as follows. First, unlike the previous HI-based methods which were heuristic in nature, the proposed method ensures to discover the optimal combination of sensor signals to better understand the underlying degradation mechanism. In fact, by solving the latent linear model, our method is able to derive the best linear unbiased estimator (BLUE) of the fusion coefficients. To the best of our knowledge, this is the first work in the context of multisensor degradation modeling that has this nice property. Second, the proposed method significantly reduces the computational time. This is due to the analytical solution of the fusion coefficients that we obtain from the latent linear model. Third, the proposed method is more generic since it does not require restrictive assumptions, such as the specific form of the degradation process, which were imposed in the previous HI-based methods. Thus, it can be widely applied to a variety of situations. Fourth, the proposed method does not require to know the exact value of the failure threshold to predict the RUL, which is usually unknown in practice. As a comparison, most of the existing studies need to assume that the failure threshold is known as *a priori*. Last but not least, variable selection methods for linear regression models such as adaptive lasso can be directly incorporated in our proposed method to achieve a systematic sensor selection. This would lead to more accurate prediction results and reduce unnecessary costs.

The rest of this paper is organized as follows. Section II provides a literature review of the data fusion methods for prognostics. Section III describes the details of the proposed data fusion methods to construct a composite HI of a degraded unit and to predict its RUL. Section IV conducts a simulation study to illustrate the effectiveness and evaluate the sensitivity of the proposed method. Section V further tests the proposed method using the degradation data set of aircraft gas turbine engines and compares the results with the existing benchmark method. Section VI provides a conclusion and a discussion of future research directions.

II. Literature Review

In the literature, several efforts have been made to tackle multisensor degradation modeling using data fusion. In general, data fusion methods can be classified into two main categories based on the implementation level of the fusion operation: decision-level fusion and data-level fusion [10], [11].

Decision-level fusion integrates multiple results derived from different diagnostic/prognostic approaches. For example, Hu *et al.* [12] combined the RUL prediction results from different member algorithms by weighted average, where k-fold cross validation was used to determine the weights. One of the main drawbacks of decision-level fusion approaches is that they are postprocessing techniques, and thus, the performance highly relies on the quality of the raw data and the data preprocessing procedure. In addition, most of these methods only produce a point estimate of the RUL.

In contrast, data-level fusion methods directly combine the sensor signals or the extracted features. In the literature, a number of data-level fusion methods have been proposed including machine learning approaches [13]–[15], state-space models [16], [17], principal component analysis (PCA) [18], and HI-based approaches [19]–[24]. In particular, machine learning approaches such as artificial neural network directly take the most recent sensor signals or features as the inputs and provide the predicted RUL as the outputs. However, sensor signals are time-series data and conventional machine learning approaches fail to effectively capture the autocorrelation of sensor signals in the context of the degradation process. To overcome this drawback, Guo *et al.* [13] recently applied a recurrent neural network (RNN) to fuse multiple features of bearings. Though RNNs are known to be useful for handling time-series data, the constructed RNNs behave like a black box which makes it less explainable and hard to incorporate domain knowledge into the models. In addition, RNNs need to be trained by very large amounts of historical data, which is costly and often inapplicable in degradation systems. Another commonly used approach is to utilize the state-space models and discretize the degradation status into a finite state space. For example, in Yu [17], the state-space model was used to model the degradation of lithium-ion battery and to predict RUL. However, this approach relies on the memoryless assumption, i.e., the future degradation depends only on the current degradation status of a unit rather than the past, which does not always hold in real-world applications [1], [25]. The PCA has also been used for data-level fusion. Recently, in Fang *et al.* [18], functional principal components analysis was used to select the informative sensors, and multivariate functional principal components analysis was used to extract features from multiple sensor signals for prognosis. Unfortunately, the extracted features are quite difficult to interpret in practice.

In this paper, we focus on HI-based methods. The key idea of the HI-based method is to construct a 1-D HI by directly combining multiple sensor signals to characterize the underlying degradation process. Compared with the aforementioned data fusion approaches, HI-based methods are highly desired in practice due to three main reasons. First, the rich literature based on a single sensor signal for degradation modeling and prognostics can be directly applied based on the constructed HI as the HI can be regarded as another single sensor signal but with more information. Second, the constructed HI shows a real-time characterization of the degradation process of a unit, which results in better interpretation than most other data fusion models that behave like a black box by providing only a final prediction result. Third, the 1-D HI can be easily visualized to help practitioners make better decisions. In fact, most of the existing prescriptive models, such

as maintenance scheduling and spare parts logistics have already assumed such a real-time HI is available when making decisions.

Despite these advantages, great challenges also exist in HI-based methods. One major challenge is that the underlying degradation status is unobservable. To address this issue, Yang *et al.* [24] explicitly expressed the HI of a unit as a deterministic function of time and regressed the multiple sensor signals against the function values. However, this approach failed to capture the stochastic nature of the degradation process. Alternatively, [20]–[22] identified the desired properties of a good degradation signal and constructed the HI in the way such that these desired properties were optimized. Although these methods showed a promising prognostic performance, they were heuristic and could not guarantee to find the optimal combination of sensor signals.

Recently, Song and Liu [23] developed a new approach that solved the HI construction by the quantile regression technique. While [23] showed that it was theoretically possible to find the best combination of sensor signals for HI construction by solving the quantile regression problem, restrictive assumptions were made to ensure the theoretical properties. For example, [23] modeled the HI by a mixed effect model with the random-effect parameter assumed to be multivariate normally distributed, which thus limited its applications. Also, [23] required to solve a large-scale quantile regression problem, which was time-consuming and might not be able to numerically find the global optimal solution in practice.

Since some sensors may not be related to the underlying degradation status, a sensor selection algorithm is necessary to ensure the effectiveness of the constructed HI and prognostic performance. However, there is still a lack of a systematic approach to identify the informative sensors signals in the current literature of multisensor degradation modeling. Very few studies attempted to provide systematic sensor selection procedures [18]. Nevertheless, these procedures are not generic enough, i.e., they are designed for specific data-level fusion models, and still cannot guarantee to select out the optimal subsets of sensors to recover the underlying degradation status of a unit.

To fill this literature gap, this paper aims at developing a more generic HI-based method that allows deriving the optimal combination of sensor signals with greater applicability and also the incorporation of a unified sensor selection procedure.

III. Methodology

In this section, we will introduce the proposed data-level fusion method in detail. In Sections III-A and III-B, we describe the formulation of our problem and present the parameter estimation method. Section III-C elaborates the latent linear model involving the multiple sensor signals and the failure time. In Section III-D, the adaptive lasso technique is incorporated for sensor selection. Section III-D discusses several considerations in implementing the proposed method in practice. Finally, in Section III-F, we discuss RUL prediction using the constructed HI.

A. Problem Formulation

Following most of the existing studies [3], [7], [26], we first provide a definition of failure as the result of degradation. Specifically, let $\eta_i(t)$ denote the underlying degradation status of unit i at time t. Then, the failure time T_i of unit i is defined as the time that the underlying degradation status of unit i first reaches a predefined failure threshold l

$$T_i = \operatorname*{argmin}_t \ \eta_i(t) \geq l. \tag{1}$$

While the specific form of $\eta_i(t)$ is not required, we consider p linearly independent basis functions $\boldsymbol{\psi}(t) = [\psi_1(t), \ldots, \psi_p(t)] \in \mathbb{R}^{1\times p}$ and decompose $\eta_i(t)$ as

$$\eta_i(t) = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i \tag{2}$$

where $\boldsymbol{\Gamma}_i = [\Gamma_{i,1}, \ldots, \Gamma_{i,p}]^T \in \mathbb{R}^{p\times 1}$ are the coefficients of the basis functions for unit i. For example, if $\boldsymbol{\psi}(t) = [1, t, \ldots, t^{p-1}]$, then $\eta_i(t)$ is represented as the $(p-1)$-order polynomial model. The existing literature (see [21]–[23]) often restricted $\eta_i(t)$ to be a special general path model, where $\boldsymbol{\Gamma}_i$ was assumed to follow a p-dimensional multivariate normal distribution. However, the assumption of multivariate normality can be quite limited. First, the symmetry required by the normal distribution may not be satisfied in general. Second, the underlying degradation process should be monotonic [27]; however, the normally distributed $\Gamma_{i,j}$ can have either positive or negative values, which may violate the monotonicity property. In this paper, we do not impose any restriction on the specific form of the degradation process $\eta_i(t)$, nor the normality assumption for $\boldsymbol{\Gamma}_i$. In other words, a wide range of degradation models including the general path models and the stochastic process models can be adopted to describe $\eta_i(t)$. As a result, the proposed method is more generic and can be applied to various situations.

We follow [23] to define the composite HI. In particular, we assume that there exists a fusion function $z(\cdot)$ to recover the underlying degradation status of a unit from the multiple sensor signals with the contamination of a white noise, i.e.,

$$\eta_i(t) = z(\boldsymbol{L}_i(t)) - \varepsilon_i(t) \tag{3}$$

where $\boldsymbol{L}_i(t) = [L_{i,1}(t), \ldots, L_{i,s}(t)] \in \mathbb{R}^{1\times s}$ is a vector of the sensor signals collected from s sensors of unit i at time t, $L_{i,j}(t)$ is the jth sensor signal of unit i at time t, and $\varepsilon_i(t) \sim N(0, \sigma_0^2)$ is the independent and identically distributed noise. Then, the composite HI of unit i at time t, denoted by $h_i(t)$, is defined as

$$h_i(t) = z(\boldsymbol{L}_i(t)). \tag{4}$$

Without loss of generality, in this paper, we consider the linear fusion function, i.e.,

$$z(\boldsymbol{L}_i(t)) = \boldsymbol{L}_i(t)\boldsymbol{w}_0 \tag{5}$$

where $\boldsymbol{w}_0 = [w_1, \ldots, w_s]^T \in \mathbb{R}^{s\times 1}$ is a vector of fusion coefficients to combine multiple sensor signals. In fact, the conventional degradation model for a single sensor signal is only a special case with $z(\boldsymbol{L}_i(t)) = L_{i,j}(t)$, which assumes the jth sensor can fully characterize the degradation process.

In addition, note that nonlinear fusion functions can be approximated in the linear form. In particular, with K basis functions, denoted by $B_k(\cdot)$ $(k = 1, \ldots, K)$, a nonlinear fusion function $z(\boldsymbol{L}_i(t))$ can be approximated as

$$z(\boldsymbol{L}_i(t)) \approx \sum_{k=1}^{K} B_k(\boldsymbol{L}_i(t))w_k = \sum_{k=1}^{K} L'_{i,k}(t)w_k \tag{6}$$

where $L'_{i,k}(t) = B_k(\boldsymbol{L}_i(t))$ is an artificial signal (i.e., transformed features from the original sensor signals). One of the most commonly used methods for the nonlinear mappings are kernel-based methods [28], [29]. However, to limit the scope of this paper, we will consider this extension to nonlinear fusion functions in the future study.

To summarize, the HI $h_i(t)$, the sensor signals $\boldsymbol{L}_i(t)$, and the degradation status $\eta_i(t) = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i$ can be expressed as follows:

$$h_i(t) = \boldsymbol{L}_i(t)\boldsymbol{w}_0 = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i + \varepsilon_i(t). \tag{7}$$

Assume there are m historical units that have failed, and for historical unit i, the sensor signals $\boldsymbol{L}_i(t)$ are measured at time $t = t_{i,1}, t_{i,2}, \ldots t_{i,n_i}$, where n_i is the total number of measurements of unit i. Let $\boldsymbol{h}_i = [h_i(t_{i,1}), \ldots, h_i(t_{i,n_i})]^T \in \mathbb{R}^{n_i \times 1}$ denote a vector of HI for unit i,

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

denote a matrix containing all the sensor signals of unit i,

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

denote a design matrix, and $\boldsymbol{\varepsilon}_i = [\varepsilon_i(t_{i,1}), \ldots, \varepsilon_i(t_{i,n_i})]^T \in \mathbb{R}^{n_i \times 1}$ denote a vector containing errors. Then, (7) can be rewritten in the following matrix form:

$$\boldsymbol{h}_i = \boldsymbol{L}_i\boldsymbol{w}_0 = \boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i + \boldsymbol{\varepsilon}_i. \tag{8}$$

Our goal is to estimate the fusion coefficients $\boldsymbol{w}_0$. Equation (8) looks similar to the conventional linear regression models at the first glance. However, since the response variable $\boldsymbol{h}_i$ is unobservable, we cannot directly derive $\boldsymbol{w}_0$ using the existing linear regression approaches. Next, we propose a novel method to estimate the fusion coefficients $\boldsymbol{w}_0$.

B. Estimation of Fusion Coefficients

At first, we regard $\boldsymbol{w}_0$ as known and obtain the least squares estimation of $\boldsymbol{\Gamma}_i$ based on (8) as

$$\hat{\boldsymbol{\Gamma}}_i = \left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i\boldsymbol{w}_0. \tag{9}$$

Since $\boldsymbol{L}_i\boldsymbol{w}_0$ is normally distributed given $\boldsymbol{\Gamma}_i$ according to (8), i.e., $\boldsymbol{L}_i\boldsymbol{w}_0|\boldsymbol{\Gamma}_i \sim N_{n_i}(\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i, \sigma_0^2\mathbf{I})$, the conditional distribution of $\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i$ also follows a p-dimensional multivariate normal distribution with mean and variance as

$$\begin{aligned} E(\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i) &= \left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T E(\boldsymbol{L}_i\boldsymbol{w}_0|\boldsymbol{\Gamma}_i) \\ &= \left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_i \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i) &= \left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\,\mathrm{Var}(\boldsymbol{L}_i\boldsymbol{w}_0|\boldsymbol{\Gamma}_i)\boldsymbol{\Psi}_i\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1} \\ &= \sigma_0^2\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}. \end{aligned} \tag{10}$$

This distribution relies on the unknown variable $\boldsymbol{\Gamma}_i$ and thus cannot be directly used. To address this challenge, our new idea is to utilize the observable failure time T_i from historical units to characterize the unobservable $\boldsymbol{\Gamma}_i$ according to (1). Specifically, recalls that the degradation status is $\eta_i(t) = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i$, and thus we can write $\boldsymbol{\psi}(T_i)\boldsymbol{\Gamma}_i = l$. This motivates us to investigate the distribution of $\boldsymbol{\psi}(T_i)\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i$. Since the failure time T_i can be regarded as a function of $\boldsymbol{\Gamma}_i$, T_i will be a constant given $\boldsymbol{\Gamma}_i$ (later on we will show that the failure threshold l can be set as any positive number and thus l can be treated as known here already). This indicates that $\boldsymbol{\psi}(T_i)\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i$ also follows a multivariate normal distribution with mean and variance as

$$E[\boldsymbol{\psi}(T_i)\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i] = \boldsymbol{\psi}(T_i)\boldsymbol{\Gamma}_i = l$$

and

$$\mathrm{Var}[\boldsymbol{\psi}(T_i)\hat{\boldsymbol{\Gamma}}_i\big|\boldsymbol{\Gamma}_i] = \sigma_0^2\boldsymbol{\psi}(T_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(T_i)^T.$$

Therefore

$$\boldsymbol{\psi}(T_i)\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i \sim N_p\left(l, \sigma_0^2\boldsymbol{\psi}(T_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(T_i)^T\right). \tag{11}$$

Interestingly, this distribution does not require $\boldsymbol{\Gamma}_i$ to be known. Therefore, we can pretend the realizations of $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_m$ are known and use maximum likelihood estimation to estimate $\boldsymbol{w}_0$. Let $\boldsymbol{\Gamma}_1^*, \ldots, \boldsymbol{\Gamma}_m^*$ denote the realizations of $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_m$ for the historical units, and τ_i denote the observed failure time of unit i. The conditional likelihood is

$$\begin{aligned} L_p &= P\left(\boldsymbol{\psi}(\tau_1)\hat{\boldsymbol{\Gamma}}_1, \ldots, \boldsymbol{\psi}(\tau_m)\hat{\boldsymbol{\Gamma}}_m|\boldsymbol{\Gamma}_1^*, \ldots, \boldsymbol{\Gamma}_m^*\right) \\ &= \prod_{i=1}^{m} P\left(\boldsymbol{\psi}(\tau_i)\hat{\boldsymbol{\Gamma}}_i|\boldsymbol{\Gamma}_i^*\right). \end{aligned} \tag{12}$$

It is straightforward to obtain the log-likelihood function as

$$\log L_p = -\frac{1}{2\sigma_0^2}\sum_{i=1}^{m}\frac{(\boldsymbol{\psi}(\tau_i)\hat{\boldsymbol{\Gamma}}_i - l)^2}{\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(\tau_i)^T} + C \tag{13}$$

where C is a constant. Therefore, we can estimate $\boldsymbol{w}_0$ by maximizing $\log L_p$, i.e.,

$$\begin{aligned} \hat{\boldsymbol{w}} &= \underset{\boldsymbol{w}}{\mathrm{argmin}}\sum_{i=1}^{m}\frac{(\boldsymbol{\psi}(\tau_i)\hat{\boldsymbol{\Gamma}}_i - l)^2}{\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(\tau_i)^T} \\ &= \underset{\boldsymbol{w}}{\mathrm{argmin}}\sum_{i=1}^{m}\frac{\left(\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i\boldsymbol{w} - l\right)^2}{\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(\tau_i)^T}. \end{aligned} \tag{14}$$

Please note that although the likelihood function is conditioned on $\boldsymbol{\Gamma}_1^*, \ldots, \boldsymbol{\Gamma}_m^*$, the true realizations of $\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_m$ are not required in the above optimization problem. As a result, we can get the analytical solution of (14) as

$$\hat{\boldsymbol{w}} = l\left(\sum_{i=1}^{m}\boldsymbol{a}_i\boldsymbol{a}_i^T\right)^{-1}\left(\sum_{i=1}^{m}\boldsymbol{a}_i b_i\right) \tag{15}$$

where

$$\boldsymbol{a}_i^T = \frac{\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i}{\sqrt{\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(\tau_i)^T}} \in \mathbb{R}^{1\times s}$$

and

$$b_i = \frac{1}{\sqrt{\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\psi}(\tau_i)^T}}.$$

Since the failure threshold l in (15) only acts as a scale factor, we can arbitrarily set l to any positive number if it is unknown, e.g., $l = 1$. This is a particularly useful result in that the failure threshold is often unknown and is hard to obtain its exact value in practice. We will further explain in Section III-F that setting l to any positive number does not affect the RUL prediction result. Since the closed-form solution in (15) only requires the computation of the inverse of a $s \times s$ matrix, the proposed approach is very computationally efficient, and thus can be easily applied even to massive data, i.e., with many historical units.

C. Latent Linear Model

As we can see, the estimation of fusion coefficients is very similar to the least squares estimation of a linear regression model. The following proposition provides more insights on the proposed approach and presents the latent linear model for HI construction.

Proposition 1: The fusion coefficients $\boldsymbol{w}_0$ satisfies the weighted linear model as

$$\boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i \cdot \boldsymbol{w}_0 + \tilde{\varepsilon}_i = l \quad \forall i = 1, \ldots, m \tag{16}$$

where $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i \in \mathbb{R}^{1\times s}$ can be regarded as covariates, and $\tilde{\varepsilon}_i \sim N(0, \boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\psi}(\tau_i)^T\sigma_0^2)$ are the mutually independent noises.

To prove this proposition, we can write

$$\begin{aligned} l &= \boldsymbol{\psi}(\tau_i)\boldsymbol{\Gamma}_i = \boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i \\ &= \boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^T(\boldsymbol{L}_i\boldsymbol{w}_0 - \boldsymbol{\varepsilon}_i) \\ &= \boldsymbol{\psi}(\tau_i)\left(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_\mathbf{i}\right)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i \cdot \boldsymbol{w}_0 + \tilde{\varepsilon}_i. \end{aligned}$$

The first equality is due to the definition of failure time in (1), and the third equality results from (8). Then, it is straightforward to obtain the distribution of $\tilde{\varepsilon}_i$ that is $\tilde{\varepsilon}_i \sim N(0, \boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\psi}(\tau_i)^T\sigma_0^2)$.

This latent linear model provides a meaningful physical interpretation of the HI construction. Let $\boldsymbol{L}_{i,j}$ denote a vector containing the jth sensor signal of unit i for all measurements. We can write the jth entry of the covariates $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i$ as $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_{i,j}$. This entry can be interpreted as the fitted jth sensor signal at the observed failure time τ_i. Thus, we can consider $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i \cdot \boldsymbol{w}_0$ as the fitted HI of unit i at the observed failure time τ_i. This implies that the latent linear model connects the HI at the failure time and the failure threshold. Compared with the original linear model in (8), we can see that the latent linear model does not require any unobservable variables to estimate $\boldsymbol{w}_0$.

We can easily transform the weighted linear model in (16) to an unweighted linear model by multiplying both sides with $\{\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\psi}(\tau_i)^T\}^{-(1/2)}$ and obtain

$$\boldsymbol{a}_i^T\boldsymbol{w}_0 + \varepsilon_i^* = b_i l$$

with $\boldsymbol{a}_i$ and b_i as defined in (15), and $\varepsilon_i^* \sim N(0, \sigma_0^2)$ are mutually independent noises.

Recall that Gauss–Markov theorem says that under certain conditions, the ordinary least squares (OLS) estimator of the coefficients of a linear regression model is the BLUE, that is, the estimator that has the smallest variance among those that are unbiased and linear in the observed output variables [30]. In the case that $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i$ has full rank, all Gauss–Markov assumptions are met in the latent linear model. Thus, this finding indicates that $\hat{\boldsymbol{w}}$ is the BLUE in such cases. To the best of our knowledge, this is the first and only study that provides the BLUE of the fusion coefficients for HI-based approaches.

D. Practical Considerations

In practice, it is possible that there is multicollinearity among the entries of $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i$; then, the minimizer of (14) may not converge to $\boldsymbol{w}_0$. Specifically, in such a case, there may exist $\tilde{\boldsymbol{w}} \neq \boldsymbol{w}_0$ that satisfies $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_i\tilde{\boldsymbol{w}} + \tilde{\varepsilon}_i' = l$ for some residuals $\tilde{\varepsilon}_i'$ with smaller variance than $\tilde{\varepsilon}_i$, and thus the estimation will converge to $\tilde{\boldsymbol{w}}$ rather than $\boldsymbol{w}_0$. For example, let us consider the case where the signal of the first sensor is constant, e.g., $L_{i,1}(t) = 1$ for $\forall t$ and $\forall i = 1, \ldots, m$, and all entries of the first column of $\boldsymbol{\Psi}_i$ are 1, i.e., $\boldsymbol{\Psi}_i = [\mathbf{1}, \tilde{\boldsymbol{\Psi}}_i]$. Then, the fitted sensor signal of the first sensor $\boldsymbol{\psi}(\tau_i)(\boldsymbol{\Psi}_i^T\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{L}_{i,1} = 1$ is also a constant. In this case, the minimizer of (14) will be $\tilde{\boldsymbol{w}} = [l, 0, \ldots, 0]^T$, which perfectly fits the linear model. However, the constructed HI based on $\tilde{\boldsymbol{w}}$ does not show any trends, and thus fails to provide any meaningful information for prognostics. Similarly, if there exists a linear combination of sensor signals that is almost flat (i.e., no clear degradation trend), the minimizer of (14) may converge to the wrong vector $\tilde{\boldsymbol{w}}$ as well.

If the constructed HI does not show any clear trend, there are three possible strategies to address this issue. First, we can conduct a preselection of sensors to make sure that there is no sensor signal nor a linear combination of different sensor signals which does not show any clear degradation trend. Second, we can preselect sensors that only show consistent increasing/decreasing trend across all units and add sign constraints to $\boldsymbol{w}$ according to the trend information of each sensor signal. In particular, we constrain that the sensors with increasing (decreasing) trends only have positive (negative) fusion coefficients when solving (14). In this way, we can avoid the situations where an increasing sensor signal and a decreasing sensor signal cancel out and result in a constant value. The third strategy is inspired by PCA. In particular, we may repeat solving (14) as long as the solution constructs an HI without trend, where for the Kth iteration ($K \geq 2$), we add the constraints $\tilde{\boldsymbol{w}}_k^T\boldsymbol{w} = 0$, $\forall k = 1, 2, \ldots, K-1$. Here, $\tilde{\boldsymbol{w}}_k$ is the optimal solution in the kth iteration which results in a constant HI. In other words, we seek the minimizer of (14) only in the space that is orthogonal to the

previous solutions. The idea behind this strategy is that $\boldsymbol{w}_0$ should be orthogonal to any $\tilde{\boldsymbol{w}}$ which constructs an HI without a clear trend. This is because if they are not orthogonal, then we can decompose $\boldsymbol{w}_0$ as $\boldsymbol{w}_0 = \alpha_1\tilde{\boldsymbol{w}} + \alpha_2\boldsymbol{w}'$ for some scalars α_1 and α_2, where $\boldsymbol{w}'$ is a vector orthogonal to $\tilde{\boldsymbol{w}}$. Thus, the constructed HI $\boldsymbol{L}_i\boldsymbol{w}_0 = \alpha_1\boldsymbol{L}_i\tilde{\boldsymbol{w}} + \alpha_2\boldsymbol{L}_i\boldsymbol{w}'$ contains a component $\boldsymbol{L}_i\tilde{\boldsymbol{w}}$ that does not show any information, and only the component $\boldsymbol{L}_i\boldsymbol{w}'$ is informative. We will consider more systematic investigations to avoid the constant HI construction in the future study.

E. Sensor Selection

Note that if a sensor does not relate to the underlying degradation process, it should be assigned a fusion coefficient of 0. Thanks to the latent linear model developed in Section III-C, a variety of existing variable selection approaches based on a linear model can be directly applied to the multisensor degradation modeling problems. In this section, we apply the well-known adaptive lasso proposed by Zou [31] to this latent linear model for sensor selection.

Similar to the popular lasso method [32], the adaptive lasso also uses the l_1-norm of the coefficient as a penalty. The difference is that the adaptive lasso imposes different penalty weights for different regression coefficients. In other words, larger penalty weights are applied to less important covariates. In this way, the adaptive lasso enjoys the oracle property, i.e., it performs as well as if the true underlying model was given in advance.

Recall that since the failure threshold l only acts as a scale factor, we simply set l to 1. After introducing the adaptive lasso to the latent linear model, we estimate $\boldsymbol{w}_0$ by

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \sum_{i=1}^{m} \left(\boldsymbol{a}_i^T\boldsymbol{w} - b_i\right)^2 + \lambda\sum_{j=1}^{s}\delta_j|w_j| \tag{17}$$

where λ is a regularization parameter and δ_j is a penalty parameter for the jth sensor. Following the existing literature [31], the penalty parameter is set to $\delta_j = 1/|\hat{w}_j^{LS}|^\gamma$, where $\hat{w}_j^{LS}$ is the OLS estimate of $\boldsymbol{w}_0$ and γ is the some positive constant. If a sensor is less related to the underlying degradation process (i.e., smaller $|\hat{w}_j^{LS}|$), it is assigned a larger penalty weight (i.e., larger δ_j) which results in a smaller fusion coefficient (i.e., smaller w_j). As a result, the fusion coefficients of noninformative sensors will be forced to 0.

F. Remaining Useful Life Prediction

Once the estimated fusion coefficients $\hat{\boldsymbol{w}}$ is derived, we can construct the HI, $\boldsymbol{h}_i = \boldsymbol{L}_i\hat{\boldsymbol{w}}$, for each historical unit i. Similarly, for an in-service unit r which is partially degraded with the collected signals $\boldsymbol{L}_r$ in real time, we can construct its HI as $\boldsymbol{h}_r = \boldsymbol{L}_r\hat{\boldsymbol{w}}$. Then, a degradation model for a single sensor signal can be used to analyze the constructed HI and predict the RUL of the in-service unit.

As mentioned earlier, since we do not make any restrictions on the degradation process $\eta_i(t)$, a variety of degradation models including the general path models and stochastic process models can be used to analyze the constructed HI. As an example, here we show how to analyze the HI based on a popular general path model [2]. Specifically, we consider the general path model as $h_i(t) = \eta_i(t) + \varepsilon_i(t) = \boldsymbol{\psi}(t)\boldsymbol{\Gamma}_i + \varepsilon_i(t)$ where $\boldsymbol{\Gamma}_i$ is a random-effect parameter with prior distribution $\boldsymbol{\Gamma}_i \sim G(\cdot)$, and the prior distribution $G(\cdot)$ can be estimated based on historical units. We can then update the posterior distribution of $\boldsymbol{\Gamma}_r$ for the in-service unit r as $P(\boldsymbol{\Gamma}_r|\boldsymbol{h}_r) \propto P(\boldsymbol{h}_r|\boldsymbol{\Gamma}_r)P(\boldsymbol{\Gamma}_r)$. If there is no analytical solution for the posterior distribution, numerical methods such as Monte Carlo Markov Chain can be employed. Therefore, the cumulative distribution function (CDF) of the failure time T_r of unit r is $F_{T_r}(t|\boldsymbol{h}_r) = P(T_r \le t|\boldsymbol{h}_r) = P(\boldsymbol{\psi}(t)\boldsymbol{\Gamma}_r \ge l|\boldsymbol{h}_r)$ according to the definition in (1).

From this CDF and (15), we can confirm that the failure threshold l only acts as a scale factor and does not affect the RUL prediction results. Specifically, if we replace l with $l' = \xi l$ in (15), where ξ is a positive constant, the estimated fusion coefficient will change from $\hat{\boldsymbol{w}}$ to $\xi\hat{\boldsymbol{w}}$. Then, the constructed HI is also scaled by a factor ξ, i.e., $\xi\boldsymbol{h}_i$. According to the RUL prediction procedure described above, it is straightforward to see that the new HI $\xi\boldsymbol{h}_i$ coupled with the new failure threshold ξl will lead to the same CDF of the estimated failure time as the HI $\boldsymbol{h}_i$ coupled with the failure threshold l. This verifies that when the true failure threshold is unknown, we can arbitrarily set l to any positive number.

Since the in-service unit has not failed yet, the CDF needs to be updated in real time given the latest measurement time t_{r,n_r}

$$F_{T_r}(t|\boldsymbol{h}_r, T_r > t_{r,n_r}) = \frac{P(\boldsymbol{\psi}(t)\boldsymbol{\Gamma}_r \ge l|\boldsymbol{h}_r) - P(\boldsymbol{\psi}(t_{r,n_r})\boldsymbol{\Gamma}_r \ge l|\boldsymbol{h}_r)}{1 - P(\boldsymbol{\psi}(t_{r,n_r})\boldsymbol{\Gamma}_r \ge l|\boldsymbol{h}_r)}.$$

Since the truncated CDF is skewed, we estimate the failure time $\hat{T}_r$ as the median of $F_{T_r}(t|\boldsymbol{h}_r, T_r > t_{r,n_r})$, i.e., $F_{T_r}(\hat{T}_r|\boldsymbol{h}_r, T_r > t_{r,n_r}) = 0.5$. Thus, the estimated RUL is $\hat{T}_r - t_{r,n_r}$.

As a special case, if $G(\cdot)$ is a multivariate normal distribution, i.e., $\boldsymbol{\Gamma}_i \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, the posterior distribution $P(\boldsymbol{\Gamma}_r|\boldsymbol{h}_r)$ has a close-form expression as

$$\boldsymbol{\Gamma}_r|\boldsymbol{h}_r \sim N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \tag{18}$$

where

$$\boldsymbol{\mu}_r = \left(\frac{\boldsymbol{\Psi}_r^T\boldsymbol{\Psi}_r}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\frac{\boldsymbol{\Psi}_r^T\boldsymbol{h}_r}{\sigma_0^2} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)$$

and $\boldsymbol{\Sigma}_r = ((\boldsymbol{\Psi}_r^T\boldsymbol{\Psi}_r/\sigma_0^2) + \boldsymbol{\Sigma}_0^{-1})^{-1}$. It is then straightforward to obtain the conditional CDF of the failure time as

$$F_{T_r}(t|\boldsymbol{h}_r, T_r > t_{r,n_r}) = \frac{\Phi(g(t)) - \Phi(g(t_{r,n_r}))}{1 - \Phi(g(t_{r,n_r}))}$$

where $\Phi(\cdot)$ is CDF of the standard normal distribution, and $g(t) = (\boldsymbol{\psi}(t)\boldsymbol{\mu}_r - l)/(\boldsymbol{\psi}(t)\boldsymbol{\Sigma}_r\boldsymbol{\psi}(t)^T)^{1/2}$ (the detailed proof can be referred to [20]).

IV. Simulation Studies

In this section, a series of numerical studies are conducted to demonstrate the effectiveness and evaluate the sensitivity of the

proposed method using simulated degradation signals. Specifically, we investigate the performance of the proposed method in three different scenarios. Section IV-A introduces how we generate the simulated degradation signals. Section IV-B studies the parameter estimation accuracy, sensor selection accuracy and computational time of the proposed method under an ideal scenario. In Section IV-C, we consider the scenario when the unknown failure threshold is a random variable rather than a fixed value. Finally, in Section IV-D, the simulation is carried out when only sparse data is available to realize the data challenge in practice.

A. Data Generation

Without loss of generality, we generate units with a linear degradation process according to

$$\eta_i(t) = \Gamma_{i,0} + \Gamma_{i,1} t \tag{19}$$

where we draw the random-effect parameter from a bivariate normal distribution

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

As mentioned earlier, since $\Gamma_{i,1}$ follows a normal distribution, it is possible to generate a sample with $\Gamma_{i,1} \leq 0$. In such cases, we discard the sample and generate a new one to ensure the monotonicity, i.e., the underlying degradation processes of all units are increasing. The true failure threshold is set to be $l = 400$. Then, we record the true failure time of unit i, denoted by τ_i, according to (1). True HI is generated by adding a random noise as defined in (3) to the underlying degradation process in (19), i.e., $h_i(t) = \eta_i(t) + \varepsilon_i(t)$ where $\varepsilon_i(t) \sim N(0, 20^2)$.

Each unit has four sensors (i.e., $s = 4$) with the true value of fusion coefficients $\boldsymbol{w}_0 = [w_1, w_2, w_3, w_4]^T = [0.6, 0.2, -0.5, 0]^T$. Four sensor signals are randomly generated as

$$\begin{aligned} L_{i,1}(t) &= U_{i,1}^{(1)} \sqrt{t} - U_{i,1}^{(2)} \sin(0.05t) + \varepsilon_{i,1}(t), \\ L_{i,2}(t) &= U_{i,2}^{(1)} t + U_{i,2}^{(2)} \sin(0.1t) + \varepsilon_{i,2}(t), \\ L_{i,3}(t) &= (h_i(t) - w_1 L_{i,1}(t) - w_2 L_{i,2}(t))/w_3 \end{aligned}$$

and

$$L_{i,4}(t) = U_{i,4}^{(1)} t + U_{i,4}^{(2)} + \varepsilon_{i,4}(t) \tag{20}$$

where $U_{i,1}^{(1)}, U_{i,1}^{(2)}, U_{i,2}^{(1)}, U_{i,4}^{(2)} \sim \text{uniform}(0, 30)$, $U_{i,2}^{(1)}, U_{i,4}^{(1)} \sim \text{uniform}(0, 2)$, and $\varepsilon_{i,1}(t)$, $\varepsilon_{i,2}(t)$, $\varepsilon_{i,4}(t) \sim N(0, 20^2)$. Note that the signal of sensor 3 is calculated according to $\boldsymbol{w}_0$ using $h_i(t)$ and the first two sensors to satisfy (7). Sensor 4 is not related to the underlying degradation process, and thus the corresponding fusion coefficient is 0.

All the signals of unit i are recorded at equidistant times $t = 1, \ldots, n_i$, where $n_i = \lfloor \tau_i \rfloor$ is the largest integer less or equal to the failure time τ_i. Fig. 1 shows the true HI and four signals of three randomly generated units.

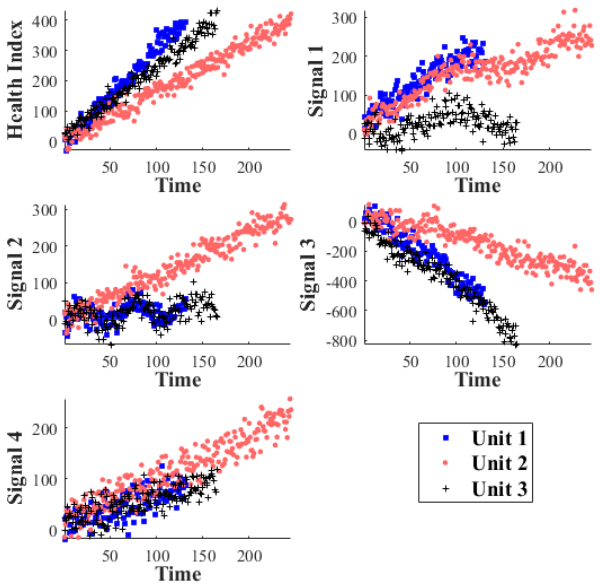


Fig. 1. Degradation signal plots for the constructed HI and four sensor signals of three randomly generated units.

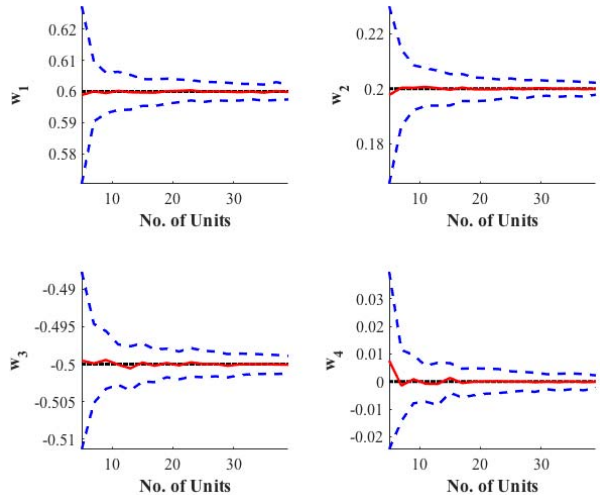


Fig. 2. Fusion coefficients estimation results for the ideal scenario. Solid line: mean estimation for each entry of $\boldsymbol{w}_0$. Dashed lines: one standard deviation of the fusion coefficients estimation. Dotted horizontal line: true value of each entry of $\boldsymbol{w}_0$.

B. Ideal Scenario

The first simulation is conducted to verify the parameter estimation performance of the proposed method in the ideal situation. We randomly select m units as the historical units. Based on the sensor signals and the failure time of the historical units, we estimate the true fusion coefficient of each sensor. The procedure is replicated 100 times for each selected value of m. Recall that since l only acts as a scale factor which does not affect the RUL prediction, here we use $l = 400$ when estimating the fusion coefficients; in this way, we can obtain the correct scale of the fusion coefficients and easily compare our estimation with the true values. Fig. 2 shows the mean and variance of the estimation of the fusion coefficients. The x-axis represents the number of the sampled historical

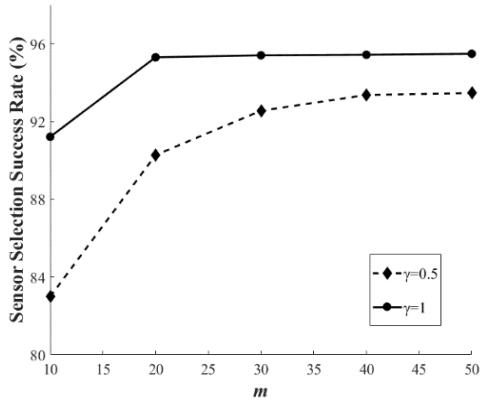


Fig. 3. Proportion of trials that the proposed method selects the right set of sensors in the ideal scenario.

units m. The dotted horizontal line represents the true fusion coefficient of each sensor. The solid and dashed curves represent the mean and one standard deviation of the fusion coefficients estimation, respectively. From Fig. 2, we can see that the estimation is very accurate and improves as the number of the sampled historical units increases.

To verify the sensor selection performance of our method, we again randomly select m historical units and repeat the sensor selection for 1000 times for each selected value of m. When applying the adaptive lasso, we set γ to 0.5 and 1, and use fivefold cross validation to search for the optimal λ for a given γ. Fig. 3 shows the proportion of trials that the proposed method selects the right set of sensors, i.e., only sensors 1, 2, and 3. As can be seen from Fig. 3, the accuracy in finding the right set of sensors increases as more historical units are available.

The computational time of the proposed method for fusion coefficients estimation is also measured and compared with the results of the benchmark method: quantile regression data fusion method in [23]. We use [23] as the benchmark method since the method also ensures that the estimated fusion coefficients converge to the true values under some assumptions. The computational time measurements of the proposed method and the benchmark method are replicated 50 times for each selected value of m. All simulations are implemented in MATLAB and executed on an Intel Core i5-6300U 2.40-GHz processor with 16-GB RAM. The average computational time of the proposed method and the benchmark method is represented in Table I. We can see that the proposed method requires much less computational time than the benchmark method. This is because the proposed method provides an analytical solution of the fusion coefficients, whereas the benchmark method has to solve a large-scale optimization to estimate the fusion coefficients.

As mentioned earlier, in many real-world applications, it is possible that the prior distribution of $\mathbf{\Gamma}_i$ does not follow a multivariate normal distribution. Thus, we conduct additional similar simulations, except that the prior distribution of $\mathbf{\Gamma}_i$ is nonnormal. Specifically, we consider two cases where each entry of $\mathbf{\Gamma}_i$ follows a beta distribution and a gamma distribution, respectively. The result of fusion coefficients estimation is very similar to that with normally distributed $\mathbf{\Gamma}_i$ and thus is omitted here. This further verifies that the proposed method is not limited to the prior distribution of $\mathbf{\Gamma}_i$, which is different from many existing works (see [20]–[23]).

TABLE I

AVERAGE COMPUTATIONAL TIME IN SECONDS FOR FUSION COEFFICIENTS ESTIMATION

m	The Proposed Method (sec)	Benchmark Method (sec)
20	0.020	197.964
40	0.022	274.204
60	0.027	311.350
80	0.034	430.213
100	0.041	492.171

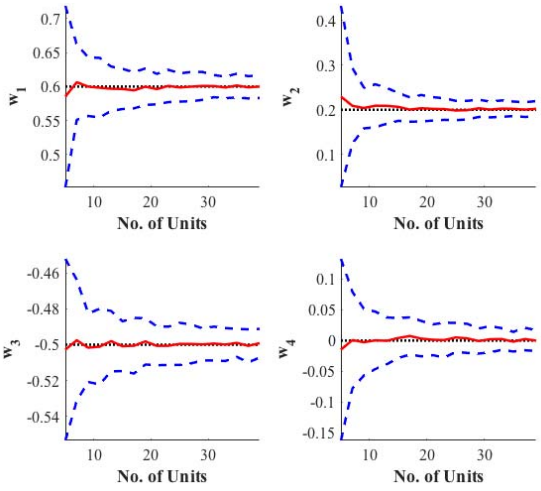


Fig. 4. Estimation results under the random failure threshold scenario. Solid line: mean estimation for each entry of $\boldsymbol{w}_0$. Dashed lines: one standard deviation of the fusion coefficients estimation. Dotted horizontal line: true value of each entry of $\boldsymbol{w}_0$.

C. Sensitivity to Random Failure Threshold

In this section, a simulation is further carried out to test the proposed method with a relaxation of the assumption that the failure threshold l is a fixed value. In real-world applications, different units indeed may fail at different levels of degradation status [21]. Thus, we generate a new data set following the same procedures as described in Section IV-A, except that the failure threshold is uniformly distributed in $[375, 425]$ rather than fixed. We apply the proposed method to the new data set while still assuming the failure threshold is a fixed value.

Following the same procedure as in Section IV-B, m units are randomly selected as historical units. Then, the fusion coefficients estimation is repeated 100 times for each m and is shown in Fig. 4 which indicates that the estimations are still very accurate. This is because, as described in Section III-F, the estimation of the fusion coefficients does not require to know the exact value of failure threshold.

D. Sensitivity to Data Sparsity

In practice, it is common that the collected sensor signals are sparse or incomplete due to limited resources for data collection or data losses during transmission. To evaluate

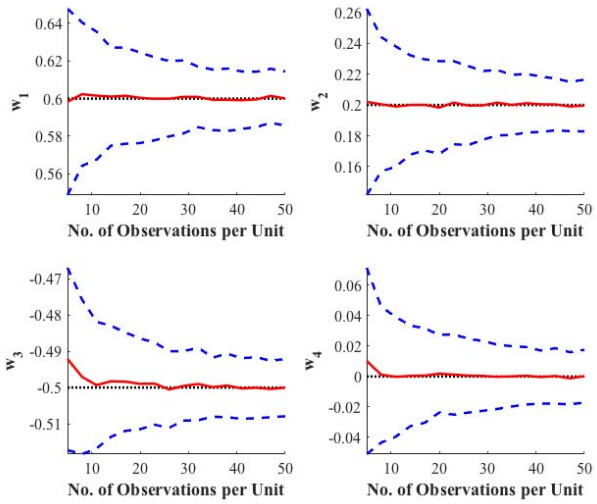


Fig. 5. Estimation results when only sparse measurements are available. Solid line: mean estimation for each entry of $\boldsymbol{w}_0$. Dashed lines: one standard deviation of the fusion coefficients estimation. Dotted horizontal line: true value of each entry of $\boldsymbol{w}_0$.

the sensitivity of the proposed method to data sparsity, we randomly choose 10 units as the training set. For each unit, we randomly sample a number of measurements to estimate the true fusion coefficients. This procedure is repeated 500 times, and the result of the fusion coefficients estimation is shown in Fig. 5. The x-axis means the number of available measurements for each historical unit. The dotted 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 estimation of the fusion coefficients. Fig. 5 shows that as the number of available sensor measurements per unit increases; the fusion coefficient estimation becomes more accurate.

V. Case Study

In this section, we employ the proposed method to predict the RUL of aircraft gas turbine engines. In addition, the results are compared with the benchmark approach: quantile regression data fusion model in [23]. The benchmark model utilizes quantile regression for HI construction and has been shown to outperform other existing data-level fusion methods (see [19]–[22]) as well as each single sensor signal in RUL prediction for the same data set.

A. Overview of the System and Data Set

The degradation-based sensor data are generated from C-MAPSS, a software widely used to simulate the degradation of turbofan aircraft engines [33]. The degradation in engine performance is due to wear and tear according to the usage pattern. To make it more realistic, each unit starts with a different degree of initial wear and manufacturing variation.

At each measurement, a total of 21 sensor signals are collected. The detailed descriptions of these 21 sensors are given in Table II. The data set consists of 100 historical units (i.e., $m = 100$) that include a total of 20 631 measurements (i.e., $\sum_{i=1}^{m} n_i = 20\,631$) and 100 in-service units that include a total of 13 096 measurements. All units have a single failure mode and operates under the same environmental condition.

TABLE II
C-MAPSS Outputs to Measure System Response

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	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	rpm
Nc	Physical core speed rpm	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

The sensor signals for each historical unit are collected until failure, whereas the sensor signals for each in-service unit are truncated at some random point prior to its failure. The failure time of all historical units and the actual RUL of all in-service units are also recorded.

B. Data Preprocessing

We first rule out 10 sensors to avoid the construction of constant HI as discussed in Section III-D. Specifically, if the sensor does not exhibit consistent increasing or decreasing trend in all historical units or if its variance is less than 10^{-4}, it is excluded. As a result, 11 candidate sensors are selected out of 21 sensors; including T24, T50, P30, Nf, Ps30, phi, NRf, BPR, htBleed, W31, and W32. To achieve a fair comparison, for these sensors, we then apply a log transformation and standardize all logged sensor signals in the same way as in [23].

C. Results and Comparison

The quadratic degradation model (i.e., $\boldsymbol{\psi}(t) = [1, t, t^2]$) is applied since it provides a good fitting based on the existing studies [19], [21], [23]. At first, we conduct sensor selection based on the historical units. When implementing the adaptive lasso, we consider three choices for γ: 0.5, 1, and 2. The fivefold cross validation is employed to find the optimal λ for a given γ. As a result, $\gamma = 0.5$ and $\lambda = 0.015$ are chosen as an optimal pair and all 11 sensors are selected as informative sensors. Our sensor selection result turns out to agree with previous studies [20]–[23], which manually selected the 11 sensors. The estimates of the fusion coefficient for each sensor are presented in Table III. Note that since the failure threshold l does not affect the RUL prediction, here we arbitrarily set $l = 2$ in the fusion coefficients estimation.

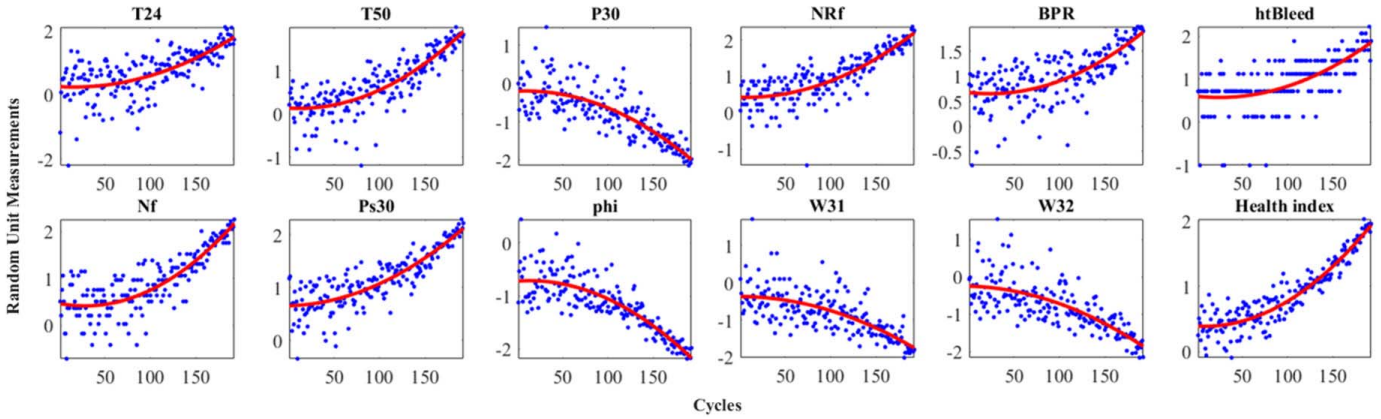


Fig. 6. Degradation signals plot and model fittings for 11 selected sensors and the constructed HI of a randomly selected in-service unit.

TABLE III
ESTIMATED FUSION COEFFICIENTS $\hat{\boldsymbol{w}}$ FOR EACH SENSOR

Sensor	Value	Sensor	Value
T24	0.184	NRf	-0.058
T50	0.289	BPR	0.146
P30	-0.074	htBleed	0.131
Nf	-0.071	W31	-0.115
Ps30	0.129	W32	-0.245
phi	-0.140		

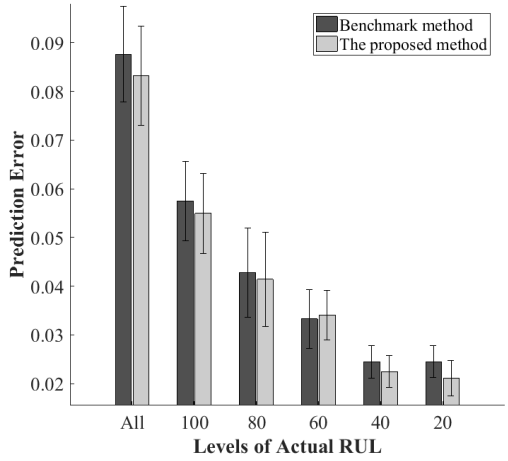


Fig. 7. Comparison results of the RUL prediction errors for the in-service units by using the benchmark method and the proposed method.

The HI of each unit is then constructed using the estimated fusion coefficients.

Fig. 6 compares each individual sensor and the constructed HI of a randomly selected in-service unit. From Fig. 6, we can see that the constructed HI provides a much better model fitting result than original single sensors.

Based on the constructed HI, we then predict the RUL for each in-service unit. To provide a fair comparison, we also adopt the assumption in [23] that the random-effect parameter $\boldsymbol{\Gamma}_i$ follows a multivariate normal distribution. In assessing the prediction error, we use the following error criteria:

$$\text{error} = \frac{|\text{Estimated RUL} - \text{Actual RUL}|}{\text{Actual Failure Time}}. \tag{21}$$

Since the sensor measurements for different in-service units are truncated at different time points, we compare the prediction error at different levels of actual RUL, as shown in Fig. 7. For example, the level "80" on the x-axis represents the prediction error of the in-service units whose actual RUL is equal to or less than 80. The prognostic results from the proposed method and the benchmark method are represented in Fig. 7. The bars refer to the average of the prediction errors and the error bars show one standard deviation. We can see that the proposed method yields lower overall RUL prediction errors than the benchmark method. In addition, the advantage of the proposed method seems to be more significant when the in-service units approach to the end of life, which is very important for practical applications.

VI. DISCUSSION AND CONCLUSION

Utilization of multiple sensors in condition monitoring has received much attention in recent years. In particular, different sensor signals may have different degrees of relevance to the degradation process. Hence, the key challenges in multisensor degradation modeling are twofold. One is how to select the informative sensors. The other is how to effectively combine the information from the selected sensor signals.

In this paper, we propose a generic HI-based data fusion method that constructs an HI by automatically selecting and combining the multiple sensor signals to better understand the degradation process. Unlike existing HI-based approaches, we propose a latent linear model for HI construction and a systematic sensor selection procedure, which resolve these above-mentioned two challenges in a unified manner. The proposed method has the following contributions. First, the estimated fusion coefficients converge to the true value. In fact, by solving the latent linear model, our method obtains the BLUE. To the best of our knowledge, this is the only method that has this nice property when dealing with multisensor degradation signals. Second, the proposed method requires much less computational time since the closed-form solution of the fusion coefficients is available. Third, compared to the previous HI-based methods, the proposed method is more generic with greatly relaxed assumptions. Specifically, a variety of degradation models can be employed to represent the

degradation process, and the random-effect parameter can have any distribution. Fourth, the proposed method can accurately predict RUL without requiring to know the exact value of the failure threshold. Finally, variable selection methods for linear regression models can be directly adopted to the latent linear model for a systematic sensor selection.

The effectiveness and the sensitivity of the proposed method under different scenarios were investigated through simulation studies and the case study. The simulation results showed that the proposed method estimated the fusion coefficients accurately even when the failure threshold was not fixed or only sparse measurements were available. For the case study, the degradation data set of aircraft engines was used to evaluate the proposed method, which showed that our method had better RUL prediction performance compared to the existing benchmark method.

There are several important topics for future research. First, an extension of the proposed data fusion method to the cases with multiple failure modes and multiple operation conditions will be of interest in our future research. Second, more systematic and effective approaches are needed to tackle the constant HI construction issue as described in Section III-D. Finally, to highlight our main idea, a linear fusion function is adopted in this paper when constructing the HI. It would be interesting to study how to construct the nonlinear mappings between the HI and each individual sensor signal.

REFERENCES

[1] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.

[2] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, May 1993.

[3] N. Gebraeel and N. Gebraeel, "Sensory-updated residual life distributions for components with exponential degradation patterns," *IEEE Trans. Autom. Sci. Eng.*, vol. 3, no. 4, pp. 382–393, Oct. 2006.

[4] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, and M. G. Pecht, "Remaining useful life estimation based on a nonlinear diffusion degradation process," *IEEE Trans. Rel.*, vol. 61, no. 1, pp. 50–67, Mar. 2012.

[5] Z.-S. Ye, N. Chen, and Y. Shen, "A new class of Wiener process models for degradation analysis," *Rel. Eng. Syst. Saf.*, vol. 139, pp. 58–67, Jul. 2015.

[6] X. Wang, N. Balakrishnan, and B. Guo, "Residual life estimation based on a generalized Wiener degradation process," *Rel. Eng. Syst. Saf.*, vol. 124, pp. 13–23, Apr. 2014.

[7] Z.-S. Ye and N. Chen, "The inverse Gaussian process as a degradation model," *Technometrics*, vol. 56, no. 3, pp. 302–311, Jul. 2014.

[8] J. Lawless and M. Crowder, "Covariates and random effects in a gamma process model with application to degradation and failure," *Lifetime Data Anal.*, vol. 10, no. 3, pp. 213–227, Sep. 2004.

[9] V. Bagdonavičius and M. S. Nikulin, "Estimation in degradation models with explanatory variables," *Lifetime Data Anal.*, vol. 7, no. 1, pp. 85–103, 2001.

[10] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[11] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[12] C. Hu, B. D. Youn, P. Wang, and J. T. Yoon, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," *Rel. Eng. Syst. Saf.*, vol. 103, pp. 120–135, Jul. 2012.

[13] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.

[14] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *J. Intell. Manuf.*, vol. 23, no. 2, pp. 227–237, Apr. 2012.

[15] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821–832, Dec. 2013.

[16] Z. Xu, Y. Ji, and D. Zhou, "Real-time reliability prediction for a dynamic system based on the hidden degradation process identification," *IEEE Trans. Rel.*, vol. 57, no. 2, pp. 230–242, Jun. 2008.

[17] J. Yu, "State-of-health monitoring and prediction of lithium-ion battery using probabilistic indication and state-space model," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 11, pp. 2937–2949, Nov. 2015.

[18] X. Fang, K. Paynabar, and N. Gebraeel, "Multistream sensor fusion-based prognostics model for systems with single failure modes," *Rel. Eng. Syst. Saf.*, vol. 159, pp. 322–331, Mar. 2017.

[19] C. Song, K. Liu, and X. Zhang, "Integration of data-level fusion model and kernel methods for degradation modeling and prognostic analysis," *IEEE Trans. Rel.*, vol. 67, no. 2, pp. 640–650, Jun. 2018.

[20] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 652–664, Jul. 2013.

[21] L. Kaibo and H. Shuai, "Integration of data fusion methodology and degradation modeling process to improve prognostics," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 344–354, Jan. 2016.

[22] K. Liu, A. Chehade, and C. Song, "Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1504–1514, Jul. 2017.

[23] C. Song and K. Liu, "Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach," *IISE Trans.*, vol. 50, no. 10, pp. 853–867, Oct. 2018.

[24] F. Yang, M. S. Habibullah, T. Zhang, Z. Xu, P. Lim, and S. Nadarajan, "Health index-based prognostics for remaining useful life predictions in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2633–2644, Apr. 2016.

[25] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Trans.*, vol. 45, no. 9, pp. 939–952, Sep. 2013.

[26] Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht, "Degradation data analysis using Wiener processes with measurement errors," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 772–780, Dec. 2013.

[27] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Appl. Stochastic Models Bus. Ind.*, vol. 31, no. 1, pp. 16–32, 2015.

[28] G. Wang, J. Jiao, and S. Yin, "A kernel direct decomposition-based monitoring approach for nonlinear quality-related fault detection," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1565–1574, Apr. 2017.

[29] G. Wang and J. Jiao, "A kernel least squares based approach for nonlinear quality-related fault detection," *IEEE Trans. Ind. Electron.*, vol. 64, no. 4, pp. 3195–3204, Apr. 2017.

[30] R. C. Hill, W. E. Griffiths, and G. C. Lim, *Principles of Econometrics*, 4th ed. Hoboken, NJ, USA: Wiley, 2011.

[31] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.

[32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.

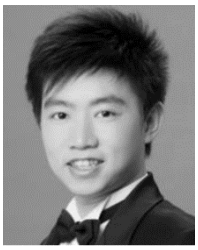
[33] Y. Liu, *User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)*, 2nd ed. Cleveland, OH, USA: NASA Glenn Research Center, 2012.

Minhee Kim received the B.S. degree in industrial and management engineering from the Pohang University of Science and Technology, Pohang, South Korea, in 2017. She is currently pursuing the Ph.D. degree with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA.

Her research interests include degradation process modeling with applications in healthcare and manufacturing.

Changyue Song received the B.S. degree in industrial engineering from Tsinghua University, Beijing, China, in 2012, and the M.S. degree in management science and engineering from Peking University, Beijing, in 2015. He is currently pursuing the Ph.D. degree with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA.

His research interests include statistical modeling for system improvement and data analytics.



Kaibo Liu (M'14) received the B.S. degree in industrial engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2009, and the M.S. degree in statistics and the Ph.D. degree in industrial engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011 and 2013, respectively.

He is currently an Assistant Professor with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA. His research interests include data fusion for process modeling, monitoring, diagnosis, and prognostics and decision making.

Dr. Liu is a member of ASQ, INFORMS, and IIE.