Integration of Data-Level Fusion Model and Kernel Methods for Degradation Modeling and Prognostic Analysis

Changyue Song, Kaibo Liu, *Member, IEEE*, and Xi Zhang, *Member, IEEE*

***Abstract*—To prevent unexpected failures of complex engineering systems, multiple sensors have been widely used to simultaneously monitor the degradation process and make inference about the remaining useful life in real time. As each of the sensor signals often contains partial and dependent information, data-level fusion techniques have been developed that aim to construct a health index via the combination of multiple sensor signals. While the existing data-level fusion approaches have shown a promise for degradation modeling and prognostics, they are limited by only considering a linear fusion function. Such a linear assumption is usually insufficient to accurately characterize the complicated relations between multiple sensor signals and the underlying degradation process in practice, especially for complex engineering systems considered in this study. To address this issue, this study fills the literature gap by integrating kernel methods into the data-level fusion approaches to construct a health index for better characterizing the degradation process of the system. Through selecting a proper kernel function, the nonlinear relation between multiple sensor signals and the underlying degradation process can be captured. As a result, the constructed health index is expected to perform better in prognosis than existing data-level fusion methods that are based on the linear assumption. In fact, the existing data-level fusion models turn out to be only a special case of the proposed method. A case study based on the degradation signals of aircraft gas turbine engines is conducted and finally shows the developed health index by using the proposed method is insensitive for missing data and leads to an improved prognostic performance.**

***Index Terms*—Condition monitoring, data fusion, health index, kernel methods, prognostic analysis.**

Acronyms and Abbreviations

RUL Remaining useful life.
CDF Cumulative distribution function.

Manuscript received December 19, 2016; revised April 21, 2017; accepted June 10, 2017. Date of publication July 11, 2017; date of current version May 31, 2018. This work was supported in part by the National Science Foundation under Grant CMMI-1435809, in part by the Office of Naval Research under Grant N00014-17-1-2261, and in part by the NSFC under Grant 71201002, Grant 71471005, and Grant 71571003. Associate Editor: S. Li. *(Corresponding author: Xi Zhang.)*

C. Song and K. Liu are with the Department of Industrial and Systems Engineering College of Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: csong39@wisc.edu; kliu8@wisc.edu).

X. Zhang is with the Department of Industrial Engineering and Management, College of Engineering, Peking University, Beijing 100871, China (e-mail: xi.zhang@pku.edu.cn).

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/TR.2017.2715180

Nomenclature

$\mathbb{R}$	Set of real numbers.
m	Number of training units.
s	Number of selected sensors.
n_i	Time when the last measurements for unit i are collected.
$L_{i,j,t}$	Sensor measurement for unit i, sensor j at time t.
$\boldsymbol{L}_{i,\cdot,t}$	Multiple sensor measurements for unit i at time t.
$\boldsymbol{L}_{\cdot,\cdot,1}$	A matrix with each row recording the first sensor measurements for each unit.
$\boldsymbol{L}_{\cdot,\cdot,n}$	A matrix with each row recording the last sensor measurements for each unit.
$\boldsymbol{L}_{i,\cdot,\cdot}$	A matrix that records the sensor signals for unit i.
$h_{i,t}$	Health index for unit i at time t.
$y_{i,t}$	A signal measurement for unit i at time t.
$\mathcal{H}$	A reproducing kernel Hilbert space or feature space.
$\boldsymbol{\Phi}$	A mapping from the original space $\mathcal{X}$ to the feature space $\mathcal{H}$.
K	A kernel function.
$\boldsymbol{w}$	A vector of fusion coefficients to combine multiple sensor signals.
$\boldsymbol{\beta}$	A vector of coefficients to represent $\boldsymbol{w}$ by using sensor signals in space $\mathcal{H}$.
SNR^{d}	Signal-to-noise ratio for a degradation signal.
$\boldsymbol{c}_i$	Weight coefficient matrix for unit i.
$\boldsymbol{\Psi}_i$	Design matrix for unit i.
$\boldsymbol{H}_i$	Projection matrix for unit i.
$\boldsymbol{\Gamma}_i$	Random-effect parameters of the degradation model for unit i based on a specific signal.
$\boldsymbol{\mu}, \boldsymbol{\mu}_i$	Prior and the posterior mean of $\boldsymbol{\Gamma}_i$.
$\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_j$	Prior and the posterior variance of $\boldsymbol{\Gamma}_i$.
$t_i, \hat{t}_i$	Actual and estimated RUL for unit i.

I. Introduction

DEGRADATION is a natural and invertible process that widely exists in manufacturing tools, machines, and equipment. Once a certain level of degradation is reached, an engineering system will fail to operate properly. As unexpected failures of systems can lead to manufacturing downtime, delayed delivery, poor customer satisfaction, etc., it is critically important to accurately predict the remaining useful life (RUL) of the system in real time.

To achieve this goal, condition monitoring techniques have been developed to monitor and understand the degradation process of a system. During condition monitoring, the collected sensor signals that contain useful information about the underlying degradation mechanism are known as the degradation signals [1]. A typical assumption is that a failure will occur once the degradation signal reaches a certain threshold. When the analysis of the degradation signal indicates a high failure probability or a short RUL, a proper condition-based maintenance strategy can then be taken proactively to determine the optimal time for maintenance before failures [2]–[4]. As a result, how to accurately model and analyze the degradation signals is critically important to the success of the condition-based maintenance strategy.

In the literature, a number of studies have been done to model and analyze a single degradation signal [5]–[9]. For instance, Lu and Meeker proposed a general mixed-effect model to fit the degradation path of a unit [10]; Gebraeel considered a Bayesian framework to online update the RUL prediction once new observations of the degradation signal were available [11]. A review of these studies can be found in [12]. An implicit assumption in these studies is that the degradation condition of a system can be fully characterized by a single degradation signal. However, as shown in many complex systems and applications, such a simplified assumption may not be valid, thus often leading to an unreliable prognostic analysis [13], [14].

Due to the development of sensor and communication technology, multiple sensors are widely deployed to simultaneously monitor a system. While multiple degradation signals are collected, each of the signals may only contains partial information with respect to the health status of the system. To tackle this issue, data fusion approaches have been widely used that combine the information from multiple degradation signals to provide a more accurate estimation of the degradation process. Generally speaking, data fusion technologies can be classified into two categories [15], [16]: decision-level fusion [17], [18] and data-level fusion [19]–[21]. In the decision-level fusion approaches, each signal is modeled and analyzed separately, and then the prognostic results based on the analysis of each individual signal are combined to produce a final decision. For example, Sun analyzed vibration signal and oil samples separately to detect degradation in vehicles, and the two results were then combined via Bayesian inference [17]. In contrast, the data-level fusion approaches directly combine multiple sensor signals or extracted features into a one-dimensional health index, which aims to better characterize the health status of the system than each individual sensor signal. In this paper, we focus on the data-level fusion approach with two reasons: First, the one-dimensional health index facilitates data visualization and enables a continuous characterization of the health status over the entire lifecycle of the system, which provides more insights than the decision-level fusion approaches that often behave like a black box and only produce a final prediction of the RUL. Second, the constructed health index can be regarded as another sensor signal and directly used for degradation modeling and prognostics.

To construct such a health index, different data-level fusion models have been developed recently [22]–[25]. While these methods have shown a promising result, they are all limited by considering only a linear fusion function. In other words, these methods construct the health index by linear combination of the original multiple sensor signals. Unfortunately, for complex engineering system, the multiple sensor signals are usually generated by complicated mechanisms and from multiple sources, and thus may have nonlinear relations with the underlying degradation process. The linear assumption in previous studies is usually insufficient to accurately characterize these complex relations, and thus restricts the effectiveness and applicability of the data-level fusion method in practice.

To fill the current literature gap and capture the complex relations between sensor signals and the underlying degradation process, this paper proposes a generic degradation monitoring approach by integrating the data-level fusion and the kernel methods. Kernel methods have been successfully employed in various models in the literature, e.g., support vector machines [26], Gaussian processes [27], and kernel principle components analysis [28]; combining kernel methods and data fusion models for degradation modeling and prognostic analysis is, *however*, totally new and with great challenges. Generally speaking, kernel methods construct a mapping $\mathbf{\Phi}: \mathcal{X} \rightarrow \mathcal{H}$ from the original raw data space $\mathcal{X}$ to a feature space $\mathcal{H}$, where the inner product of two vectors can be measured by a kernel function K:

$$< \mathbf{\Phi}(\boldsymbol{x}_1), \mathbf{\Phi}(\boldsymbol{x}_2) >_{\mathcal{H}} = K(\boldsymbol{x}_1, \boldsymbol{x}_2). \tag{1}$$

Here $<,>$ refers to the inner product. Our main idea here is to leverage the kernel methods to map the original sensor signals into a feature space and then derive the best linear combination of the transformed signals in the feature space for health index development. While this idea is theoretically sound, as we will show later in Section III, introducing the kernel methods into the data-level fusion approach is not straightforward and needs significant new research efforts. Since different kernel functions can be used to construct different feature spaces, the proposed method is capable to explore the complex relations between multiple sensor signals and the underlying degradation process. Enabled by the kernel methods, we expect that the generality and the performance of data-level fusion approaches can be significantly enhanced.

The rest of this paper is organized as follows. In Section II, we will first review the related works on data-level fusion for degradation modeling and prognostic analysis. The details of our proposed method is introduced in Section III and validated with a case study in Section IV. Section V provides a conclusion and discusses future work.

II. Related Works

In this section, we will review existing studies on data-level fusion for degradation modeling and prognostic analysis that are closely related to the proposed method. One of such efforts was made by Yang *et al.*, who assumed the health index to be a deterministic linear function of time and estimated the fusion function by regressing sensor signals against the health index based on training units [29]. However, this assumption oversimplifies the problem and cannot capture the randomness of

each individual unit. To address this issue, Liu *et al*. considered the unit-to-unit variability and introduced two desired properties that the health index should possess to ensure successful degradation modeling and prognosis [22]:

Property 1: Once an initial fault occurs, the trend of the health index should be monotonic.

Property 2: Given the same environmental condition and failure mode, the variance of the failure threshold in the developed health index should be minimal.

Based on these two properties, Liu *et al.* proposed a nonparametric data-level fusion model that minimized the violation of monotonicity and the variance of failure threshold in the constructed health index.

While the nonparametric data-level fusion model provides a promising approach, it regards the health index development and the prognostics as two disjoint tasks, and thus cannot guarantee the constructed health index is favorable for the selected prognostic model. To address this issue, Liu and Huang proposed a semiparametric data-level fusion approach that integrated the data fusion procedure and the degradation modeling in a unified manner [23]. Specifically, they considered another property:

Property 3: Given the selected degradation model, the model fitting errors for the developed health index should be minimal.

With this property, the semiparametric data-level fusion approach minimizes the model fitting errors and the variance of the failure threshold in the constructed health index.

Based on the semiparametric model, recently Liu *et al.* [24] showed that the range information of the health index was another important criterion for successful prognostics:

Property 4: The range information of the health index should be maximal.

By combining these properties mentioned above, Liu *et al.* [24] further proposed a metric, signal-to-noise ratio for degradation signals ($\mathrm{SNR}^{\mathrm{d}}$), to measure the quality of a degradation signal:

$$\mathrm{SNR}^{\mathrm{d}} = \frac{R^2}{\delta^2 + \upsilon} \tag{2}$$

where R was the range information of the degradation signal, δ^2 was the variance of the model fitting errors, and υ was the variance of the failure threshold. Consequently, they constructed a health index with a maximized $\mathrm{SNR}^{\mathrm{d}}$ metric.

While utilizing the four properties and the $\mathrm{SNR}^{\mathrm{d}}$ metric as reviewed above, most of the aforementioned efforts made an additional assumption that the fusion function is linear. In other words, the existing health indices are constructed via a linear combination of the original sensor signals. Currently, the existing data fusion literature for degradation modeling and prognostics still lacks a generic approach that goes beyond the linear relation assumption. In this paper, we aim to address this issue and propose a more generalized data fusion model through integration with kernel methods. More specifically, we will construct a health index (HI-kernel) by introducing kernel methods into each part of the $\mathrm{SNR}^{\mathrm{d}}$ metric and meanwhile provide an optimal fusion coefficient solver to achieve the best health status representation of the monitored system. This proposed method is expected to better characterize the complex relations between multiple sensor signals and the underlying degradation process, and thus significantly enhance the performance and the applicability of the data-level fusion approaches for degradation modeling and prognostics.

III. Methodology Development

A. Introducing Kernel Methods into Health Index Development

Recall that kernel methods construct a mapping $\boldsymbol{\Phi}: \mathcal{X} \rightarrow \mathcal{H}$ from the original raw data space $\mathcal{X}$ to a reproducing kernel Hilbert space or feature space $\mathcal{H}$. When dealing with complex relations, kernel methods have the unique advantage as it can transform data into an implicit and even infinite-dimensional feature space (e.g., $\boldsymbol{x} \rightarrow \boldsymbol{\Phi}(\boldsymbol{x})$) where the complex relations (e.g., nonlinear relation) in the original space become simple relations (e.g., linear relation) in the feature space. The beauty of the kernel method is that it does not require explicit definition of the mapping function $\boldsymbol{\Phi}(\cdot)$ when calculating $<\boldsymbol{\Phi}(\boldsymbol{x}_1), \boldsymbol{\Phi}(\boldsymbol{x}_2)>_{\mathcal{H}}$, which is the inner product of two points in the feature space, as long as the kernel function $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = <\boldsymbol{\Phi}(\boldsymbol{x}_1), \boldsymbol{\Phi}(\boldsymbol{x}_2)>_{\mathcal{H}}$ has an analytic expression that can be computed quickly. Because the explicit computation of $\boldsymbol{\Phi}(\cdot)$ is avoided, and the kernel function is often much cheaper to compute, this approach is known as the "*kernel trick*" [30].

For the rest part of the paper, we use "unit" interchangeably with "engineering system" to be consistent with the literature and to emphasize that we regard the system as an integral. Denote $L_{i,j,t}$ to be the sensor measurement for unit i, sensor j at time t, and s to be the total number of sensors. Without loss of generality, we assume that sensor measurements are collected at time $t = 1, 2, \ldots, n_i$ for unit i. The observation $\boldsymbol{L}_{i,\cdot,t} = [L_{i,1,t}, \ldots, L_{i,s,t}] \in \mathbb{R}^{1\times s}$ collected at time t for unit i contains multiple sensor measurements and is mapped to a row vector $\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t}) \in \mathcal{H}$ in the feature space. Then, the health index $h_{i,t}$ for unit i at time t can be constructed via the linear combination of $\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t})$ in the feature space:

$$h_{i,t} = \boldsymbol{\Phi}\left(\boldsymbol{L}_{i,\cdot,t}\right)\boldsymbol{w} \tag{3}$$

where $\boldsymbol{w} \in \mathcal{H}$ is a column vector and it is called the fusion coefficients for the mapped multiple sensor signals $\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t})$.

With different mappings $\boldsymbol{\Phi}$, we are able to explore complex relations among multiple sensor signals $\boldsymbol{L}_{i,\cdot,t}$ when constructing the health index $h_{i,t}$. For example, suppose $\boldsymbol{x}_1 = [x_{11}, x_{12}]$, $\boldsymbol{x}_2 = [x_{21}, x_{22}]$, and kernel function

$$K\left(\boldsymbol{x}_1, \boldsymbol{x}_2\right) = \left(\boldsymbol{x}_1\boldsymbol{x}_2^T\right)^2 = \left(x_{11}x_{21} + x_{12}x_{22}\right)^2. \tag{4}$$

Accordingly, the mapping $\boldsymbol{\Phi}$ can be expressed as

$$\boldsymbol{\Phi}\left(\boldsymbol{x}\right) = \left[x_2^2, x_1^2, \sqrt{2}x_1x_2\right] \tag{5}$$

where $\boldsymbol{x} = [x_1, x_2]$. In this way, $<\boldsymbol{\Phi}(\boldsymbol{x}_1), \boldsymbol{\Phi}(\boldsymbol{x}_2)> = [x_{12}^2, x_{11}^2, \sqrt{2}x_{11}x_{12}][x_{22}^2, x_{21}^2, \sqrt{2}x_{21}x_{22}]^T = (x_{11}x_{21} + x_{12}x_{22})^2 = K(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Thus, the feature space $\mathcal{H}$ is a reproducing kernel Hilbert space and

$$h = \boldsymbol{\Phi}\left(\boldsymbol{x}\right)\boldsymbol{w} = w_1x_2^2 + w_2x_1^2 + w_3\sqrt{2}x_1x_2 \tag{6}$$

represents a quadratic relation when constructing h. Please note that here $\boldsymbol{w} = [w_1, w_2, w_3]^T \in \mathcal{H}$. From this illustrative example, we can see that the linkage between the health index and the sensor signals is possible to be generalized via the kernel methods.

B. Degradation Modeling

According to [24], the model fitting errors in the constructed health index should be minimized. Therefore, a single-signal degradation model needs to be chosen before applying the data-level fusion method. Without loss of generality, in this study, we focus on the degradation model with the pth order polynomial form as a demonstration:

$$y_{i,t} = \sum_{k=0}^{p} \Gamma_i^{(k)} t^k + \varepsilon_{i,t} \tag{7}$$

where $y_{i,t}$ is a signal measurement for unit i at time t; $\boldsymbol{\Gamma}_i = [\Gamma_i^{(0)}, \ldots, \Gamma_i^{(p)}]^T$ is the random-effect parameter which is often assumed to follow a multivariate normal distribution, i.e., $\boldsymbol{\Gamma}_i \sim N_{p+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are signal-specific hyper-parameters (i.e., different signals have different hyper-parameters); and $\varepsilon_{i,t}$ is the random noise.

Many degradation models discussed in the literature can be transformed to the form in (7), such as the simple random coefficient growth model [31], [32] and the exponential functional form model [22], [33], [34]. For example, an exponential degradation model

$$z_{i,t} = \phi + \alpha_i e^{\Gamma_i^{(1)} t + \Gamma_i^{(2)} t^2 + \varepsilon_{i,t} - \sigma^2/2} \tag{8}$$

was considered in [22], where $z_{i,t}$ was a sensor measurement for unit i at time t; ϕ was the sensor-specific fixed-effect parameter; α_i, $\Gamma_i^{(1)}, \Gamma_i^{(2)}$ were the random-effect parameters; and $\varepsilon_{i,t}$ was the random noise following $N(0, \sigma^2)$. After a log-transformation, the logged signal was modeled with the polynomial form:

$$\begin{aligned} y_{i,t} &= \ln(z_{i,t} - \phi) \\ &= \Gamma_i^{(0)} + \Gamma_i^{(1)} t + \Gamma_i^{(2)} t^2 + \varepsilon_{i,t}\,. \end{aligned} \tag{9}$$

Here, $\Gamma_i^{(0)} = \ln\alpha_i - \sigma^2/2$, and the random-effect parameters were assumed to follow a multivariate normal distribution, i.e., $\boldsymbol{\Gamma}_i = [\Gamma_i^{(0)}, \Gamma_i^{(1)}, \Gamma_i^{(2)}]^T \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

C. Kernel-Based Data Fusion Model

Recall that our goal is to integrate the kernel methods and the data-level fusion models to enhance the performance of the constructed health index for degradation modeling and prognostic analysis. Specifically, to highlight our main idea, we choose the SNR-based data-level fusion model in [24] as our baseline method and aim to construct a health index with a maximized $\mathrm{SNR}^{\mathrm{d}}$ metric in (2) through integration with the kernel methods.

First, we consider the range information of the health index. In particular, the range information of the health index for unit i can be expressed as $R_i = h_{i,n_i} - h_{i,1} = (\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,n_i}) - \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,1}))\,\boldsymbol{w}$, where n_i is the time when the last sensor measurements are collected before unit i fails. In this way, the average range information for all m training units is formulated as $R = \sum_{i=1}^{m}(h_{i,n_i} - h_{i,1})/m = \mathbf{1}^T(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n}) - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1}))\boldsymbol{w}/m$, where $\mathbf{1}$ is a column vector with all entries equal to 1, and $\boldsymbol{L}_{\cdot,\cdot,1} = [\boldsymbol{L}_{1,\cdot,1}; \ldots; \boldsymbol{L}_{m,\cdot,1}] \in \mathbb{R}^{m\times s}$ and $\boldsymbol{L}_{\cdot,\cdot,n} = [\boldsymbol{L}_{1,\cdot,n_1}; \ldots; \boldsymbol{L}_{m,\cdot,n_m}] \in \mathbb{R}^{m\times s}$ are two matrices with each row recording the first and the last sensor measurements for each unit, respectively. To simplify the notation, here $\boldsymbol{\Phi}$ is regarded as a row-by-row operator, i.e., for matrix $\boldsymbol{X} = [\boldsymbol{x}_1; \ldots; \boldsymbol{x}_n] \in \mathbb{R}^{n\times s}$, $\boldsymbol{\Phi}(\boldsymbol{X}) = [\boldsymbol{\Phi}(\boldsymbol{x}_1); \ldots; \boldsymbol{\Phi}(\boldsymbol{x}_n)]$, where $\boldsymbol{x}_i$ is the ith row of $\boldsymbol{X}$; therefore, $\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1}) = [\boldsymbol{\Phi}(\boldsymbol{L}_{1,\cdot,1}); \ldots; \boldsymbol{\Phi}(\boldsymbol{L}_{m,\cdot,1})]$. As different sensors may exhibit different trend information (increasing or decreasing), we further consider the squared range information:

$$\begin{aligned} R^2 = \boldsymbol{w}^T &\left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})^T - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})^T\right) \boldsymbol{A}\,(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n}) \\ &- \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1}))\,\boldsymbol{w} \end{aligned} \tag{10}$$

where $\boldsymbol{A} = \mathbf{1}\mathbf{1}^T/m^2 \in \mathbb{R}^{m\times m}$ is a symmetric constant matrix.

Second, we consider the model fitting errors of the constructed health index. Recall that while our idea is generic, in this study, we focus on the polynomial degradation model as described in Section III-B. By using the polynomial degradation model in (7), we get $\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t})\boldsymbol{w} = h_{i,t} = [t^0, \ldots, t^p]\,\boldsymbol{\Gamma}_i + \varepsilon_{i,t}$. Then, we use the weighted least square approach to estimate the random-effect parameter: $\hat{\boldsymbol{\Gamma}}_i = (\boldsymbol{\Psi}_i^T \boldsymbol{c}_i \boldsymbol{\Psi}_i)^{-1}(\boldsymbol{\Psi}_i^T \boldsymbol{c}_i \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})\boldsymbol{w})$, where $\boldsymbol{\Psi}_i$ is the design matrix:

$$\boldsymbol{\Psi}_i \in \mathbb{R}^{n_i\times(p+1)} = \begin{bmatrix} 1^0 & \cdots & 1^p \\ \cdots & \cdots & \cdots \\ t^0 & \cdots & t^p \\ \cdots & \cdots & \cdots \\ n_i^0 & \cdots & n_i^p \end{bmatrix} \tag{11}$$

where $\boldsymbol{c}_i = \mathrm{diag}(c_{i,1}, \ldots, c_{i,n_i}) \in \mathbb{R}^{n_i\times n_i}$ is the weight coefficient matrix (we will discuss the details on the selection of $\boldsymbol{c}_i$ in Section IV-C), and $\boldsymbol{L}_{i,\cdot,\cdot} = [\boldsymbol{L}_{i,\cdot,1}; \ldots; \boldsymbol{L}_{i,\cdot,n_i}] \in \mathbb{R}^{n_i\times s}$ is the multiple sensor signals for unit i. As a result, the weighted model fitting errors for unit i are $\boldsymbol{c}_i^{1/2}\boldsymbol{e}_i = \boldsymbol{c}_i^{1/2}(\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})\boldsymbol{w} - \boldsymbol{\Psi}_i\hat{\boldsymbol{\Gamma}}_i) = (\boldsymbol{I} - \boldsymbol{H}_i)\,\boldsymbol{c}_i^{1/2}\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})\boldsymbol{w}$, where the projection matrix $\boldsymbol{H}_i = \boldsymbol{c}_i^{1/2}\boldsymbol{\Psi}_i(\boldsymbol{\Psi}_i^T\boldsymbol{c}_i\boldsymbol{\Psi}_i)^{-1}\boldsymbol{\Psi}_i^T\boldsymbol{c}_i^{1/2} \in \mathbb{R}^{n_i\times n_i}$. Therefore, the model fitting errors of the constructed health index can be estimated by

$$\begin{aligned} \delta^2 &= \sum_{i=1}^{m}\left(\left(\boldsymbol{c}_i^{1/2}\boldsymbol{e}_i\right)^T\boldsymbol{c}_i^{1/2}\boldsymbol{e}_i\right)/m \\ &= \sum_{i=1}^{m}\boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})^T\boldsymbol{B}_i\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})\,\boldsymbol{w} \end{aligned} \tag{12}$$

where $\boldsymbol{B}_i = (\boldsymbol{c}_i^{1/2}(\boldsymbol{I} - \boldsymbol{H}_i)\boldsymbol{c}_i^{1/2})/m \in \mathbb{R}^{n_i\times n_i}$ is a symmetric constant matrix.

The third component in (2) is the variance of failure threshold and the unbiased estimator is

$$\upsilon = \sum_{i=1}^{m} \left(h_{i,n_i} - \bar{h}_{\cdot,n_i}\right)^2/(m-1) = \boldsymbol{w}^T \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})^T \boldsymbol{D} \Phi(\boldsymbol{L}_{\cdot,\cdot,n})\, \boldsymbol{w} \quad (13)$$

where $\bar{h}_{\cdot,n_i} = \frac{1}{m}\sum_{i=1}^{m} h_{i,n_i}$, $\boldsymbol{D} = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/m)/(m-1) \in \mathbb{R}^{m\times m}$ is a symmetric constant matrix, and $\boldsymbol{I}$ is the identity matrix.

By combining (10), (12), and (13), we derive a new SNR^{d} metric (named as KSNR^{d}) of the health index with integration of kernel methods:

$$\text{KSNR}^{\text{d}} = \frac{\boldsymbol{w}^T\left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})^T - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})^T\right)\boldsymbol{A}\left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n}) - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})\right)\boldsymbol{w}}{\sum_{i=1}^{m}\boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})^T\boldsymbol{B}_i\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})\,\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})^T\boldsymbol{D}\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})\boldsymbol{w}}. \quad (14)$$

Thus, our goal here is to find the optimal fusion coefficients $\boldsymbol{w}$ such that the derived KSNR^{d} metric of the constructed health index can be maximized, as mentioned in Section II. However, since the fusion coefficients $\boldsymbol{w} \in \mathcal{H}$ may be an infinite vector, and the explicit expression of the kernel mapping $\boldsymbol{\Phi}$ may not even exist, it is nearly impossible to solve (14) directly. As a result, while the reconstruction of the data-level fusion model by using the kernel methods is theoretically sound, solving (14) is a challenging task and the approach in the existing work [22]–[25] cannot be used here.

In theory, the key challenge here lies in that the kernel tricks cannot be played directly. For example, we know how to calculate $\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})^T$ via kernel tricks but not $\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})^T\boldsymbol{D}\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n})$.

D. Deriving the Optimal Solution

In this section, we will propose a new method to numerically solve (14). To begin with, we propose to first prove *the following proposition*:

Proposition: The optimal fusion coefficients $\boldsymbol{w}$ can be represented as a linear combination of $\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t})$ (see Appendix for proof), i.e.,

$$\boldsymbol{w} = \sum_i \sum_t \beta_{i,t}\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t})^T = \boldsymbol{\Phi}(\boldsymbol{L})^T\boldsymbol{\beta} \quad (15)$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1; \ldots; \boldsymbol{\beta}_m] \in \mathbb{R}^{\sum_{i=1}^{m} n_i \times 1}$, $\boldsymbol{\beta}_i = [\beta_{i,1}, \ldots, \beta_{i,n_i}]^T \in \mathbb{R}^{n_i\times 1}$, $\boldsymbol{L} = [\boldsymbol{L}_{1,\cdot,\cdot}; \ldots; \boldsymbol{L}_{m,\cdot,\cdot}] \in \mathbb{R}^{\sum_{i=1}^{m} n_i \times s}$, and $\boldsymbol{L}_{i,\cdot,\cdot} = [\boldsymbol{L}_{i,\cdot,1}; \ldots; \boldsymbol{L}_{i,\cdot,n_i}] \in \mathbb{R}^{n_i\times s}$.

In this way, (14) can be transformed as

$$\text{KSN R}^{\text{d}} = \frac{\boldsymbol{\beta}^T\boldsymbol{Q}_1^T\boldsymbol{A}\boldsymbol{Q}_1\boldsymbol{\beta}}{\sum_{i=1}^{m}\boldsymbol{\beta}^T\boldsymbol{Q}_{2,i}^T\boldsymbol{B}_i\boldsymbol{Q}_{2,i}\boldsymbol{\beta} + \boldsymbol{\beta}^T\boldsymbol{Q}_3^T\boldsymbol{D}\boldsymbol{Q}_3\boldsymbol{\beta}}. \quad (16)$$

Here, $\boldsymbol{Q}_1 = \boldsymbol{K}(\boldsymbol{L}_{\cdot,\cdot,n_i}, \boldsymbol{L}) - \boldsymbol{K}(\boldsymbol{L}_{\cdot,\cdot,1}, \boldsymbol{L}) \in \mathbb{R}^{m\times\sum_{k=1}^{m} n_k}$, $\boldsymbol{Q}_{2,i} = \boldsymbol{K}(\boldsymbol{L}_{i,\cdot,\cdot}, \boldsymbol{L}) \in \mathbb{R}^{n_i\times\sum_{k=1}^{m} n_k}$, and $\boldsymbol{Q}_3 = \boldsymbol{K}(\boldsymbol{L}_{\cdot,\cdot,n_i}, \boldsymbol{L}) \in \mathbb{R}^{m\times\sum_{k=1}^{m} n_k}$, where $\boldsymbol{K}$ is a row-wise inner product operator for matrices in space $\mathcal{H}$, i.e., for matrix $\boldsymbol{X}_k = [\boldsymbol{x}_{k,1}; \ldots; \boldsymbol{x}_{k,n}] \in \mathbb{R}^{n\times s}$, $k = 1, 2$, then

$$\boldsymbol{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\Phi}(\boldsymbol{X}_1)\cdot\boldsymbol{\Phi}(\boldsymbol{X}_2)^T = \begin{bmatrix} K(\boldsymbol{x}_{1,1}, \boldsymbol{x}_{2,1}) & \cdots & K(\boldsymbol{x}_{1,1}, \boldsymbol{x}_{2,n}) \\ \vdots & \ddots & \vdots \\ K(\boldsymbol{x}_{1,n}, \boldsymbol{x}_{2,1}) & \cdots & K(\boldsymbol{x}_{1,n}, \boldsymbol{x}_{2,n}) \end{bmatrix} \in \mathbb{R}^{n\times n}. \quad (17)$$

Here K is the kernel function, $\boldsymbol{x}_{k,i}$ is the ith row of $\boldsymbol{X}_k$, and $\boldsymbol{Q}_1$, $\boldsymbol{Q}_{2,i}$, $\boldsymbol{Q}_3$ are constant matrices once the kernel function is selected.

With this new formulation to (16), the problem now becomes to find the optimal $\boldsymbol{\beta}$ to maximize the KSNR^{d} metric instead. It is straightforward to see that KSNR^{d} does not depend on the scale of $\boldsymbol{\beta}$, i.e., KSNR^{d} value will remain the same if we multiply $\boldsymbol{\beta}$ by a nonzero number. In addition, we can further prove that the optimal solution $\boldsymbol{\beta}^*$ to (16) is the eigenvector corresponding to the maximum eigenvalue of the generalized eigenvalue problem (see Appendix for details):

$$\bar{\boldsymbol{A}}\boldsymbol{\beta} = \lambda_{\max}\,\bar{\boldsymbol{F}}\boldsymbol{\beta} \quad (18)$$

where $\bar{\boldsymbol{A}} = \boldsymbol{Q}_1^T\boldsymbol{A}\boldsymbol{Q}_1$, $\bar{\boldsymbol{F}} = \sum_{i=1}^{m}\boldsymbol{Q}_{2,i}^T\boldsymbol{B}_i\boldsymbol{Q}_{2,i} + \boldsymbol{Q}_3^T\boldsymbol{D}\boldsymbol{Q}_3$, and $\lambda_{\max}$ is the maximized value of KSNR^{d} in (16). Equation (18) can be solved by a variety of existing eigen-decomposition algorithms [35], [36]. We can further show that $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{F}}$ are symmetric positive semidefinite matrices (see Appendix for details). Thus, if $\bar{\boldsymbol{F}}$ is a positive definite matrix, i.e., all eigenvalues of $\bar{\boldsymbol{F}}$ are positive, then $\boldsymbol{\beta}^*$ is the eigenvector corresponding to the maximum eigenvalue of $\bar{\boldsymbol{F}}^{-1}\bar{\boldsymbol{A}}$; otherwise, we can introduce a small positive number θ, so that $\bar{\boldsymbol{F}} + \theta\boldsymbol{I}$ becomes positive definite and $\boldsymbol{\beta}^*$ can be derived as the eigenvector corresponding to the maximum eigenvalue of $(\bar{\boldsymbol{F}} + \theta\boldsymbol{I})^{-1}\bar{\boldsymbol{A}}$ [37].

Once the optimal solution $\boldsymbol{\beta}^*$ is derived, we can then use it to calculate the health index during condition monitoring. In particular, for a partially degraded testing unit i with multiple sensor signals $\boldsymbol{L}_{i,\cdot,\cdot} = [\boldsymbol{L}_{i,\cdot,1}; \ldots; \boldsymbol{L}_{i,\cdot,n_i}] \in \mathbb{R}^{n_i\times s}$ collected up to time n_i, the health index $\boldsymbol{h}_i = [h_{i,1}; \ldots; h_{i,n_i}]$ can be calculated as

$$\boldsymbol{h}_i = \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})\,\boldsymbol{w}^* = \boldsymbol{K}(\boldsymbol{L}_{i,\cdot,\cdot}, \boldsymbol{L})\,\boldsymbol{\beta}^* \quad (19)$$

where $h_{i,t}$ is the health index at time t for the testing unit i. Once the health index is constructed, various single-signal degradation models (e.g., described in [11]) can be used to predict the RUL. In Section IV-E of the case study, we will further demonstrate the procedure of RUL prediction based on a single signal. In addition, due to the flexibility of the kernel method, when the kernel function is linear, i.e., $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1^T\boldsymbol{x}_2$, the $\boldsymbol{\Phi}$ operator maps a vector to itself, i.e., $\boldsymbol{\Phi}(\boldsymbol{x}) = \boldsymbol{x}$. As a result, the existing data fusion models can be regarded as only a special case of our proposed generic approach with a linear kernel function. In other words, the performance of our proposed method is guaranteed to be as good as existing ones [22]–[24]. On the other hand, when the true relation between the underlying degradation process and the sensor signals is nonlinear, we can always find an appropriate kernel function such that the performance of the proposed method is better.

TABLE I
DETAILED DESCRIPTION OF THE 21 SENSORS [14]

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

IV. CASE STUDY

In this section, we will implement and evaluate our proposed kernel-based data fusion model based on a simulated dataset that involves the degradation of aircraft engines. In particular, we will compare the performance of the constructed health index by using our proposed method with each original individual sensor signal and the health indices developed by aforementioned three benchmark approaches including the nonparametric [22], the semiparametric [23], and the SNR-based [24] data fusion models.

A. Data Description

Our dataset is provided in [14], and consists of multiple degradation signals of turbofan engines under a single failure mode and a single environmental condition. Specifically, the training dataset consists of 100 training units (i.e., $m = 100$) with a total of 20 631 observations (i.e., $\sum_{i=1}^{m} n_i = 20\,631$). The testing dataset includes 100 partially degraded testing units with 13 096 observations. There is an additional file recording the actual RUL of the 100 testing units. Measurements of 21 sensors were simultaneously and continuously collected during condition monitoring, which contained comprehensive information such as the temperature and pressure of each engine. The detailed descriptions of these sensors are given in Table I.

The dataset was generated by C-MAPSS, a widely used model for simulating the degradation process of large commercial turbofan engines [14], [38]. However, users do not have explicit access to the simulation model (i.e., the simulation model is a "black box" to users); thus, the underlying assumption here is that we have to predict the RUL for the 100 testing units solely based on the historical records of the 100 training units and the observed signals of the 100 testing units. The predicted RULs by using different methods will be compared with the actual ones to evaluate the prognostic performance of these methods.

B. Data Preprocessing

Prior to data fusion and degradation modeling, we need to first decide which sensor signals to use in the following analysis. Previous studies [22]–[24] selected $s = 11$ sensors out of the original 21 sensors based on the criterion that the degradation signals should consistently exhibit an increasing or decreasing trend in all training units. In particular, a sensor was selected if its last measurement was consistently larger (increasing trend) or smaller (decreasing trend) than the first measurement in all training units. In this study, we adopt the same criterion in order to provide a fair comparison with previous studies. The selected sensors are T24, T50, P30, Nf, Ps30, phi, NRf, BPR, htBleed, W31, and W32 and all other sensor signals are discarded. Similarly to [22]–[24], the selected sensor signals are then standardized in the following analysis.

C. Kernel-based Data Fusion Model

In this case study, we follow previous studies and choose the exponential degradation model as they provided a good model fitting to each sensor signal in the dataset [22]–[24]. Specifically, the health index is constructed by the combination of logged sensor signals using the proposed data-level fusion method. The second-order polynomial degradation model described in Section III-B is employed to model the health index.

Recall that $\boldsymbol{c}_i = \operatorname{diag}(c_{i,1}, \ldots, c_{i,n_i})$ is the weight coefficient matrix for the residuals in the fitted degradation model in (12). Considering that the prognostic result becomes more sensitive as the unit approaches to failure, here we adopt the guidelines proposed by Liu and Huang [23] and assign increasing weights, i.e., $0 < c_{i,1} \leq c_{i,2} \leq \ldots \leq c_{i,n_i}$. In addition, Liu and Huang [23] proposed two options to assign $\{c_{i,t}\}$ as either an arithmetic series or a geometric series. For the arithmetic series

$$c_{i,t} = c_{i,1} + (t - 1)\frac{2 - 2c_{i,1}n_i}{(n_i - 1)\,n_i} \tag{20}$$

and for the geometric series

$$c_{i,t} = c_{i,1} \cdot q^{t-1} \tag{21}$$

where q satisfies $c_{i,1}q^{n_i} - c_{i,1} - q + 1 = 0$. In this case study, the weight coefficients $\{c_{i,t}\}$ are set as an arithmetic series, which is also consistent with previous studies [22]–[24].

Our proposed data fusion model is generic and various kernel functions can be employed. Generally speaking, the kernel function measures the similarity between sensor measurements and should be selected according to the characteristics of the dataset. In particular, when no prior information is available, it can be selected via kernel learning methods [39] or cross validation. In this case study, we select a second-order polynomial kernel as a demonstration based on the following considerations.

1) We expect a better prognostic performance by using a polynomial kernel than previous studies [22]–[24] with a linear kernel, since the linear model is only a special case of the polynomial model.
2) Previous studies [22]–[24] have shown the effectiveness of the linear kernel, which implicates the existence of linear

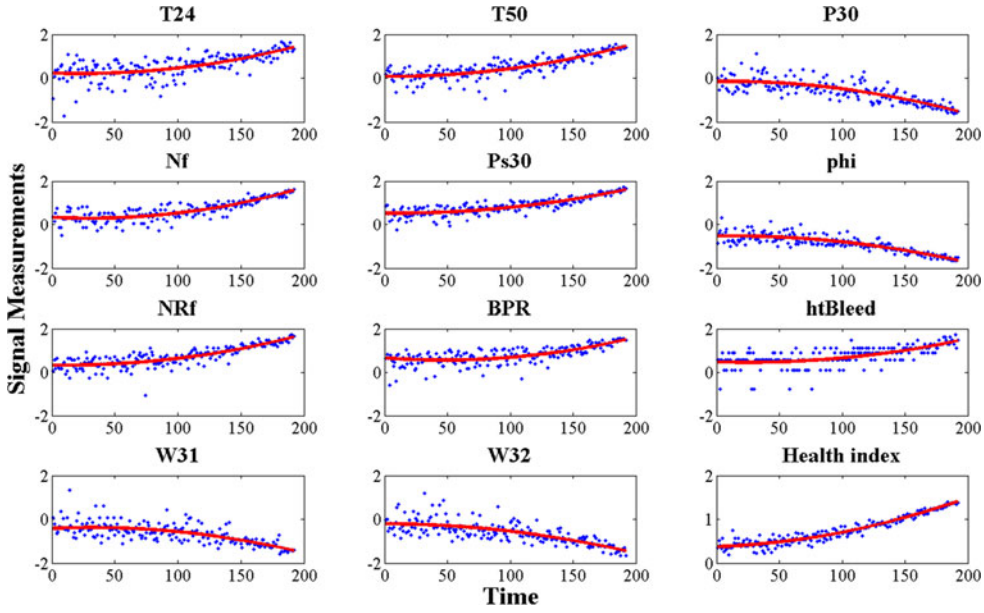

Fig. 1. An illustration plot of the model fitting results for the selected sensor signals and the developed health index using the proposed method for a training unit.

TABLE II
SNR METRIC VALUES OF ALL SELECTED SENSORS AND THE HEALTH INDICES BASED ON DIFFERENT DATA FUSION MODELS

Name	T24	T50	P30	Nf	Ps30
Value	9.912	25.985	18.580	10.147	32.835
Name	phi	NRf	BPR	htBleed	W31
Value	23.154	11.225	13.392	10.165	14.581
Name	W32	HI-non	HI-semi	HI-SNR	HI-kernel
Value	12.878	108.840	141.142	143.379	171.313

relations in the sensor signals, and these linear relations can be incorporated by a polynomial kernel.

3) Polynomial kernels are popular in various studies and applications, and the mapping function $\mathbf{\Phi}$ associated with polynomial kernels has explicit mathematical expressions and can be well explained.

By using our proposed method in Section III-D, we can derive the optimal solution in (16). In order to compare the health index constructed by the proposed method (HI-kernel) with each selected sensor signal and the health indices developed by the aforementioned three benchmark data-level fusion methods including the nonparametric method (HI-non), the semiparametric method (HI-semi), and the SNR-based method (HI-SNR), we also model each logged sensor signal and the three benchmark health indices separately by the second-order polynomial degradation model. The SNR^{d} metrics for all sensor signals and health indices are calculated and summarized in Table II. It clearly shows that our proposed new method obtains the highest SNR^{d} value, indicating the effectiveness of our approach in developing a health index with desired properties.

Fig. 1 further illustrates the degradation model fitting results for the developed health index by using our method and each selected sensor signal for a training unit. It shows that the developed health index exhibits a clearer degradation trend and a much better model fitting result.

D. Sensitivity to Missing Data

Missing data (e.g., due to sensor malfunction) is a common issue in practice, which often results in a huge challenge for data fusion and analysis. In this study, we analyze the sensitivity of our proposed method for constructing the health index when there are different degrees of missing data. Specifically, we consider the constructed health index based on all measurements of the selected sensors for the training units as the control group, which is denoted as $\{h_{i,t} : i = 1, \ldots, m; t = 1, \ldots, n_i\}$. Then, we randomly hide a fraction of sensor measurements and recalculate the health index based on the remaining measurements. In particular, for unit i, we randomly draw N_i samples from $\{1, \ldots, n_i\}$, which are denoted as $\{t_1^{(i)}, \ldots, t_{N_i}^{(i)}\}$. We assume only the sensor measurements $\boldsymbol{L}_{i,\cdot,t}$ at time $t = t_1^{(i)}, \ldots, t_{N_i}^{(i)}$ are available and all other sensor measurements are missing. The new health index constructed based on the sampled sensor measurements is denoted as $\{h'_{i,t} : i = 1, \ldots, m; t = t_1^{(i)}, \ldots, t_{N_i}^{(i)}\}$. Please note that the first and last measurements for each unit are needed in our model for calculating the range information and failure threshold, and thus we assume that they are always available in the dataset, i.e., $t_1^{(i)} = 1$, $t_{N_i}^{(i)} = n_i$.

The sensitivity of our proposed method is measured by the relative difference between the health index based on all measurements of the selected sensors and the health index based on the dataset with missing data:

$$R_d = \frac{\sum_{i=1}^{m} \sum_{j=1}^{N_i} \left| h'_{i,t_j^{(i)}} - h_{i,t_j^{(i)}} \right|}{\sum_{i=1}^{m} \sum_{j=1}^{N_i} \left| h_{i,t_j^{(i)}} \right|} . \tag{22}$$

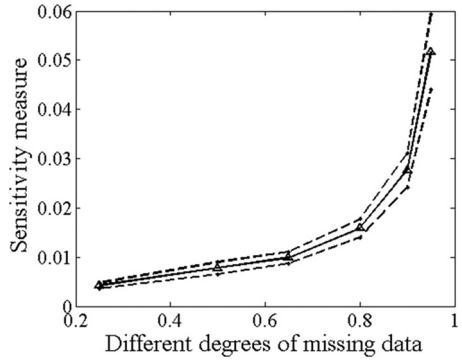

Fig. 2. Sensitivity measures for different degrees of missing data.

The procedure described above is repeated for 25 times to obtain the average and standard deviation of the sensitivity measure.

Fig. 2 illustrates the sensitivity measure of the proposed method for different degrees of missing data. For example, the label "0.2" refers that 20% of the sensor measurements are missing. The solid line represents the average sensitivity measure and the dashed lines are one standard deviation apart from the average values. A clear increasing trend is observed in the sensitivity measure, which is consistent with our intuition that more differences are expected in the constructed health index as more missing data are introduced. When 95% of the data are missing (i.e., only around ten observations are available for each training unit), the sensitivity measure still maintains only around 5%, indicating that the constructed health index by using our proposed data fusion model is insensitive to missing data.

E. RUL Prediction

In this section, we will further evaluate the performance of the constructed health index when it is used for RUL prediction. One standard approach to prognostics based on a single sensor signal is the one in [11], which aims to update the RUL prediction via a Bayesian framework in real time when new measurements are available. Here, we adopt this method and the procedure is explained below for the convenience of the readers.

As mentioned before, in this case study, each sensor signal and the health index are separately fitted by the second-order polynomial degradation model $y_{i,t} = \Gamma_i^{(0)} + \Gamma_i^{(1)} t + \Gamma_i^{(2)} t^2 + \varepsilon_{i,t}$. For a partially degraded testing unit i, the distribution of $\mathbf{\Gamma}_i$ can be updated by incorporating the *in situ* sensor measurements collected from this unit. Specifically, let $\boldsymbol{y}_{i,\cdot} = [y_{i,1}, \ldots, y_{i,n_i}]^T \in R^{n_i \times 1}$ be the collected signal for the partially degraded unit i up to the current measurement epoch n_i. It can be shown that the posterior distribution of $\mathbf{\Gamma}_i$ also follows a multivariate normal distribution [23], [24]:

$$\mathbf{\Gamma}_i | \boldsymbol{y}_{i,\cdot} \sim N_3 \left(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i \right) \tag{23}$$

where

$$\boldsymbol{\mu}_i = \left(\frac{\mathbf{\Psi}_i^T \mathbf{\Psi}_i}{\sigma^2} + \mathbf{\Sigma}^{-1} \right)^{-1} \left(\frac{\mathbf{\Psi}_i^T \boldsymbol{y}_{i,\cdot}}{\sigma^2} + \mathbf{\Sigma}^{-1} \boldsymbol{\mu} \right)$$

$$\mathbf{\Sigma}_i = \left(\frac{\mathbf{\Psi}_i^T \mathbf{\Psi}_i}{\sigma^2} + \mathbf{\Sigma}^{-1} \right)^{-1}, \mathbf{\Psi}_i \in R^{n_i \times 3} = \begin{bmatrix} 1^0 & \cdots & 1^2 \\ \cdots & \cdots & \cdots \\ t^0 & \cdots & t^2 \\ \cdots & \cdots & \cdots \\ n_i^0 & \cdots & n_i^2 \end{bmatrix}.$$

Based on the updated distribution $\mathbf{\Gamma}_i | \boldsymbol{y}_{i,\cdot}$, the RUL of the testing unit i can be estimated by the time when the degradation signal crosses the failure threshold. For a specific signal, considering that the failure threshold l may not be the same for different units, we thus model l as a random variable with a normal distribution, $l \sim N(\mu^d, v^d)$, where μ^d and v^d vary in different signals and can be estimated from the training units based on the last measurements before failures. Since $y_{i,n_i+t} = \Gamma_i^{(0)} + \Gamma_i^{(1)}(n_i + t) + \Gamma_i^{(2)}(n_i + t)^2 + \varepsilon_{i,n_i+t}$, we can see that y_{i,n_i+t} follows a normal distribution with mean $\mu_{i,n_i+t} = [1, n_i + t, (n_i + t)^2]\, \boldsymbol{\mu}_i$ and variance $\sigma^2_{i,j,n_i+t} = [1, n_i + t, (n_i + t)^2]\, \mathbf{\Sigma}_i [1, n_i + t, (n_i + t)^2]^T + \sigma^2$. Therefore, the conditional cumulative distribution function (CDF) of the RUL T_i for unit i given the *in situ* signal $\boldsymbol{y}_{i,\cdot}$ is

$$\begin{aligned} F_{T_i | y_{i,\cdot}}(t) &= P\left(T_i \le t | \boldsymbol{y}_{i,\cdot}\right) = P\left(y_{i,n_i+t} \ge l | \boldsymbol{y}_{i,\cdot}\right) \\ &= \varphi\left(\frac{\mu_{i,n_i+t} - \mu^d}{\sqrt{\sigma^2_{i,n_i+t} + v^d}} \right) = \varphi\left(g(t)\right). \end{aligned} \tag{24}$$

Here, $\varphi(\cdot)$ represents the CDF of the standard normal distribution. Since $T_i \ge 0$, the truncated CDF for T_i is

$$\begin{aligned} P\left(T_i \le t | \boldsymbol{y}_{i,\cdot}, T_i \ge 0\right) &= \frac{P\left(0 \le T_i \le t | \boldsymbol{y}_{i,\cdot}\right)}{P\left(T_i \ge 0 | \boldsymbol{y}_{i,\cdot}\right)} \\ &= \frac{\varphi\left(g(t)\right) - \varphi\left(g(0)\right)}{1 - \varphi\left(g(0)\right)}. \end{aligned} \tag{25}$$

Since the truncated CDF is skewed, the median value is used as the point estimator for the RUL, i.e., we find $\hat{t}_i$ so that $P\left(T_i \le \hat{t}_i | \boldsymbol{y}_{i,\cdot}, T_i \ge 0\right) = 0.5$ as the predicted RUL for unit i based on signal $\boldsymbol{y}_{i,\cdot}$.

By comparing the predicted RUL $\hat{t}_i$ and the actual RUL t_i for the testing unit i, the prognostic performance of each sensor signal and the developed health index can be evaluated. Define the prediction error err_i as the relative difference between $\hat{t}_i$ and t_i:

$$err_i = \frac{\left|\left(n_i + \hat{t}_i\right) - \left(n_i + t_i\right)\right|}{n_i + t_i} = \frac{\left|\hat{t}_i - t_i\right|}{n_i + t_i} \tag{26}$$

where n_i is the current measurement epoch of unit i. The average prediction errors by using the best single sensor signal, HI-non, HI-semi, HI-SNR, and HI-kernel at different levels of actual RUL are illustrated and compared in Fig. 3. For example, the label "100" means that only the testing units with actual RUL less than or equal to 100 are considered, and the label "All"

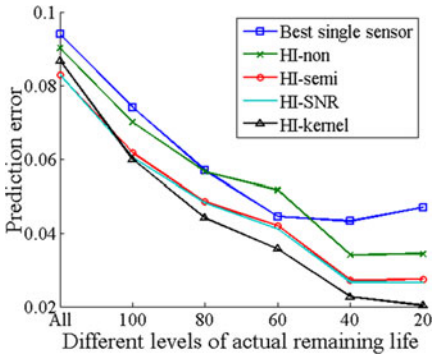

Fig. 3. Comparison of the RUL prediction errors among single sensor signal, nonparametric health index, semiparametric health index, SNR-based health index, and kernel-based health index, at different levels of actual remaining life.

means that all 100 testing units are considered. A clear decreasing trend is observed in this figure. One possible reason is that as the unit approaches to failure, we have more measurements to estimate the degradation condition of the unit and we only need to predict the RUL over a shorter period of time, thus leading to a more accurate prediction of RUL. It can also be observed that when the actual RUL is large, the average prediction error of HI-kernel is slightly higher than HI-SNR and HI-semi. However, since high uncertainty is involved in prognostics when the actual RUL is large due to the lack of sensor measurements, it is still unclear whether HI-SNR and HI-semi outperform HI-kernel in this case. On the contrary, when the actual RUL is small (less than 100), the average prediction error of HI-kernel is much smaller than all other signals with little uncertainty. In this case, the HI-kernel clearly outperforms all other signals. Actually, as RUL decreases and the unit is more prone to fail, it is more critical to predict the RUL accurately in order to determine the optimal time for maintenance. From this point of view, the HI-kernel is also more favorable than other signals. In conclusion, this result further validates the effectiveness of our proposed method.

V. Discussion and Conclusion

The rapid development of sensor technology has enabled an unprecedented opportunity for condition monitoring of complex engineering systems to avoid unexpected failures. Meanwhile, new challenges arise on how to combine the information from multiple degradation signals to better estimate the degradation status and predict the RUL of engineering systems. To tackle this challenging question, data-level fusion models have been developed, which aim to construct a composite health index by combining multiple sensor signals to better characterize the underlying degradation process of a unit. While this approach has shown a promising result, the existing data-level fusion models are limited by the linear fusion assumption when combining multiple sensor signals. Unfortunately, such a linear assumption may not be valid in practice, thus greatly restricting the performance of the developed health index and limiting the applicability of the data fusion models for degradation modeling and prognostics in many applications.

The main contribution of the paper is to fill the current literature gap by establishing a generic kernel-based data fusion model, which relaxes the existing linear assumption for better characterizing the complex relations between multiple sensor signals and the underlying degradation process. In particular, various fusion functions can be explored via the kernel methods, where the kernel function can be selected according to domain knowledge or data-driven approaches such as cross validation to maximize the RUL prediction accuracy of the constructed health index. The previous studies have been shown to be only a special case of our proposed model when the fusion function is linear, which further justifies the superiority of the proposed method. Considering that the kernel-based data fusion model cannot be directly applied, we further propose an effective approach that transforms the original optimization formulation into a generalized eigen-decomposition problem. A case study based on the degradation dataset of aircraft gas turbine engines is further conducted to numerically evaluate the prognostic performance of the constructed health index. The result shows that the kernel-based health index by using our method can better characterize the degradation process, leading to a more accurate prognosis than other benchmark approaches. A sensitivity study has also been conducted, which demonstrates that our proposed method is insensitive to different degrees of missing data.

There are several topics worth further study. First, existing data-level fusion methods make the assumption that the failures are caused by a single failure mode under a single environmental condition. Extensions of the data fusion methods to multiple failure modes and multiple environmental conditions should be studied in the future. Second, this study considers the pth-order polynomial form for the degradation model. More studies for the data fusion methods are desired to be investigated when the degradation model is associated with a more general form. Third, some of the collected sensor signals may be unrelated to the degradation process and act as noises when constructing the health index. These signals should be screened out to avoid overfitting and reduce the volume of data. Thus, how to automatically select the most useful sensor signals during data fusion is another interesting topic worth further investigation.

Appendix

Proof of Proposition

It is straightforward to see that the scale of $\boldsymbol{w}$ does not affect the value of KSNR^{d} in (14). As a result, we impose another constraint $\boldsymbol{w}^T\boldsymbol{w} = 1$ to derive a unique solution of $\boldsymbol{w}$ as a unit vector. Then (14) can be reformulated as

$$\begin{aligned} \max\ & \boldsymbol{w}^T \left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})^T \right) \\ & \times \boldsymbol{A} \left(\boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,n_i}\right) - \boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,1}\right) \right) \boldsymbol{w} \\ s.t.\ & \sum_{i=1}^{m} \boldsymbol{w}^T \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})^T \boldsymbol{B}_i \boldsymbol{\Phi}\left(\boldsymbol{L}_{i,\cdot,\cdot}\right) \boldsymbol{w} + \boldsymbol{w}^T \\ & \times \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T \boldsymbol{D} \boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,n_i}\right) \boldsymbol{w} = C \\ & \boldsymbol{w}^T\boldsymbol{w} = 1 \end{aligned} \tag{27}$$

where C is a constant number. The Lagrangian expression can be written as

$$\begin{aligned} L_p = {} & \boldsymbol{w}^T \left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})^T \right) \\ & \times \boldsymbol{A} \left(\boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,n_i}\right) - \boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,1}\right) \right) \boldsymbol{w} \\ & - \alpha_1 \sum_{i=1}^{m} \boldsymbol{w}^T \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})^T \boldsymbol{B}_i \boldsymbol{\Phi}\left(\boldsymbol{L}_{i,\cdot,\cdot}\right) \boldsymbol{w} - \alpha_1 \boldsymbol{w}^T \\ & \times \Phi(\boldsymbol{L}_{\cdot,\cdot,n_i})^T \boldsymbol{D} \boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,n_i}\right) \boldsymbol{w} \\ & + \alpha_1 C - \alpha_2 \left(\boldsymbol{w}^T \boldsymbol{w} - 1 \right) \end{aligned} \tag{28}$$

where α_1, α_2 are the Lagrangian multipliers. Let $\partial L_p / \partial \boldsymbol{w} = 0$, and then we get

$$\begin{aligned} \boldsymbol{w} = {} & \frac{1}{\alpha_2} \left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})^T \right) \\ & \times \boldsymbol{A} \left(\boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,n_i}\right) - \boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,1}\right) \right) \boldsymbol{w} \\ & - \alpha_1 \alpha_2 \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T \boldsymbol{D} \boldsymbol{\Phi}\left(\boldsymbol{L}_{\cdot,\cdot,n_i}\right) \\ & \times \boldsymbol{w} - \frac{\alpha_1}{\alpha_2} \sum_{i=1}^{m} \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})^T \boldsymbol{B}_i \boldsymbol{\Phi}\left(\boldsymbol{L}_{i,\cdot,\cdot}\right) \boldsymbol{w} \\ = {} & \left(\boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1})^T \right) \left[\frac{1}{\alpha_2} \boldsymbol{A} \left(\boldsymbol{h}_{\cdot,n_i} - \boldsymbol{h}_{\cdot,1} \right) \right] \\ & - \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i})^T \left(\frac{\alpha_1}{\alpha_2} \boldsymbol{D} \boldsymbol{h}_{\cdot,n_i} \right) \\ & - \sum_{i=1}^{m} \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot})^T \left(\frac{\alpha_1}{\alpha_2} \boldsymbol{B}_i \boldsymbol{h}_{i,\cdot} \right). \end{aligned} \tag{29}$$

Please note that $\boldsymbol{h}_{\cdot,n_i} = \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,n_i}) \boldsymbol{w} \in \mathbb{R}^{m \times 1}$, $\boldsymbol{h}_{\cdot,1} = \boldsymbol{\Phi}(\boldsymbol{L}_{\cdot,\cdot,1}) \boldsymbol{w} \in \mathbb{R}^{m \times 1}$, and $\boldsymbol{h}_{i,\cdot} = \boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,\cdot}) \boldsymbol{w} \in \mathbb{R}^{n_i \times 1}$ are column vectors. Thus, this finishes the proof that there exists a $\boldsymbol{\beta}$ such that the optimal value of $\boldsymbol{w}$ in (14) can be represented as a linear combination of $\boldsymbol{\Phi}(\boldsymbol{L}_{i,\cdot,t})^T$, i.e., $\boldsymbol{w} = \boldsymbol{\Phi}(\boldsymbol{L})^T \boldsymbol{\beta}$.

Proof of (18)

Let $\bar{\boldsymbol{A}} = \boldsymbol{Q}_1^T \boldsymbol{A} \boldsymbol{Q}_1$ and $\bar{\boldsymbol{F}} = \sum_{i=1}^{m} \boldsymbol{Q}_{2,i}^T \boldsymbol{B}_i \boldsymbol{Q}_{2,i} + \boldsymbol{Q}_3^T \boldsymbol{D} \boldsymbol{Q}_3$. Then, (16) can be simplified to

$$\mathrm{KSNR}^{\mathrm{d}} = \frac{\boldsymbol{\beta}^T \bar{\boldsymbol{A}} \boldsymbol{\beta}}{\boldsymbol{\beta}^T \bar{\boldsymbol{F}} \boldsymbol{\beta}} . \tag{30}$$

Thus, maximizing $\mathrm{KSNR}^{\mathrm{d}}$ is equivalent to the following optimization problem:

$$\begin{aligned} & \max \boldsymbol{\beta}^T \bar{\boldsymbol{A}} \beta \\ & s.t.\ \boldsymbol{\beta}^T \bar{\boldsymbol{F}} \beta = C . \end{aligned} \tag{31}$$

The Lagrangian expression can be written as

$$L_p = \boldsymbol{\beta}^T \bar{\boldsymbol{A}} \boldsymbol{\beta} + \lambda \left(\boldsymbol{\beta}^T \bar{\boldsymbol{F}} \boldsymbol{\beta} - C \right) \tag{32}$$

where λ is the Lagrangian multiplier. Let $\partial L_p / \partial \boldsymbol{\beta} = 0$, and then we get $\bar{\boldsymbol{A}} \boldsymbol{\beta} = \lambda \bar{\boldsymbol{F}} \boldsymbol{\beta}$. Therefore

$$\mathrm{KSNR}^{\mathrm{d}} = \frac{\boldsymbol{\beta}^T \bar{\boldsymbol{A}} \boldsymbol{\beta}}{\boldsymbol{\beta}^T \bar{\boldsymbol{F}} \boldsymbol{\beta}} = \lambda. \tag{33}$$

This means that the maximized value of $\mathrm{KSNR}^{\mathrm{d}}$ equals to the maximum eigenvalue of the generalized eigenvalue problem $\bar{\boldsymbol{A}} \boldsymbol{\beta} = \lambda_{\max} \bar{\boldsymbol{F}} \boldsymbol{\beta}$ and the optimal value $\boldsymbol{\beta}^*$ is the corresponding eigenvector.

Proof That $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{F}}$ in (18) are Positive Semidefinite

At first, we can see that $\boldsymbol{A}$ and $\boldsymbol{D}$ are positive semidefinite, because $\boldsymbol{A} = \boldsymbol{1}\boldsymbol{1}^T / m^2 \in \mathbb{R}^{m \times m}$ and $\boldsymbol{D} = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/m)/(m-1) \in \mathbb{R}^{m \times m}$. In addition, $\boldsymbol{B}_i$ is also symmetric and positive semidefinite, because $\boldsymbol{B}_i = \boldsymbol{c}_i^{1/2} (\boldsymbol{I} - \boldsymbol{H}_i) \boldsymbol{c}_i^{1/2} / m = ((\boldsymbol{I} - \boldsymbol{H}_i) \boldsymbol{c}_i^{1/2})^T (\boldsymbol{I} - \boldsymbol{H}_i) \boldsymbol{c}_i^{1/2} / m$, where $\boldsymbol{H}_i$ is the projection matrix. Since $\boldsymbol{A}$ is positive semidefinite, according to the Cholesky decomposition, it can be written as $\boldsymbol{A} = \boldsymbol{A}_0^T \boldsymbol{A}_0$, where $\boldsymbol{A}_0$ is a lower triangular matrix. As a result, $\bar{\boldsymbol{A}} = \boldsymbol{Q}_1^T \boldsymbol{A} \boldsymbol{Q}_1 = \boldsymbol{Q}_1^T \boldsymbol{A}_0^T \boldsymbol{A}_0 \boldsymbol{Q}_1 = (\boldsymbol{A}_0 \boldsymbol{Q}_1)^T \boldsymbol{A}_0 \boldsymbol{Q}_1$ is also symmetric and positive semidefinite. By similar procedures, $\bar{\boldsymbol{F}}$ can be shown to be symmetric and positive semidefinite as well.

REFERENCES

[1] W. Nelson, *Accelerated Testing Statistical Models, Test Plans, and Data Analysis*. New York, NY, USA: Wiley, 1990.

[2] R. K. Mobley, *An Introduction to Predictive Maintenance*. Burlington, MA, USA: Elsevier, 2002.

[3] X. Liu, J. Li, K. N. Al-Khalifa, A. S. Hamouda, D. W. Coit, and E. A. Elsayed, "Condition-based maintenance for continuously monitored degrading systems with multiple failure modes," *IIE Trans.*, vol. 45, no. 4, pp. 422–435, Apr. 2013.

[4] R. Moghaddass and M. J. Zuo, "A parameter estimation method for a condition-monitored device under multi-state deterioration," *Rel. Eng. Syst. Safety*, vol. 106, pp. 94–103, Oct. 2012.

[5] W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*. New York, NY, USA: Wiley, 1998.

[6] M. J. Zuo, R. Jiang, and R. C. M. Yam, "Approaches for reliability modeling of continuous-state devices," *IEEE Trans. Rel.*, vol. 48, no. 1, pp. 9–18, Mar. 1999.

[7] X. Liu, K. N. Al-Khalifa, E. A. Elsayed, D. W. Coit, and A. S. Hamouda, "Criticality measures for components with multi-dimensional degradation," *IIE Trans.*, vol. 46, no. 10, pp. 987–998, Oct. 2014.

[8] J. Guo, Z. Li, and M. Pecht, "A Bayesian approach for Li-Ion battery capacity fade modeling and cycles to failure prognostics," *J. Power Sources*, vol. 281, no. 1, pp. 173–184, May 2015.

[9] Z. Li, Y. Deng, and C. Mastrangelo, "Model selection for degradation-based Bayesian reliability analysis," *J. Manuf. Syst.*, vol. 37, no. 1, pp. 72–82, Oct. 2015.

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

[11] 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.

[12] 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, Oct. 2006.

[13] T. Brotherton, P. Grabill, D. Wroblewski, R. Friend, B. Sotomayer, and J. Berry, "A testbed for data fusion for engine diagnostics and prognostics," in *Proc. IEEE Aerosp. Conf.*, 2002, vol. 6, pp. 3029–3042.

[14] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. 1st Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, Oct. 2008, pp. 1–9.

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

[16] A. J. Volponi, T. Brotherton, R. Luppold, and D. L. Simon, "Development of an information fusion system for engine diagnostics and health management," Cleveland, OH, USA: NASA, Feb. 2004.

[17] Q. Sun, "Sensor fusion for vehicle health monitoring and degradation detection," in *Proc. 5th Int. Conf. Inf. Fusion*, Annapolis, MD, USA, 2002, pp. 1422–1427.

[18] B. Saha, K. Goebel, and J. Christophersen, "Comparison of prognostic algorithms for estimating remaining useful life of batteries," *Trans. Inst. Meas. Control*, vol. 31, pp. 293–308, 2009.
[19] K. Goebel and P. Bonissone, "Prognostic information fusion for constant load systems," in *Proc. 7th Annu. Conf. Inf. Fusion*, 2005, vol. 2, pp. 1247–1255.
[20] D. Simon and D. L. Simon, "Aircraft turbofan engine health estimation using constrained Kalman filtering," *ASME J. Eng. Gas Turbines Power*, vol. 127, no. 2, pp. 323–328, Apr. 2005.
[21] K. Salahshoor, M. Mosallaei, and M. Bayat, "Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended Kalman filter algorithm," *Measurement*, vol. 41, no. 10, pp. 1059–1076, Mar. 2008.
[22] K. Liu, N. 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.
[23] K. Liu and S. Huang, "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.
[24] 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.*, to be published, doi: 10.1109/TASE.2015.2446752.
[25] H. Yan, K. Liu, X. Zhang, and J. Shi, "Multiple sensor data fusion for degradation modeling and prognostics under multiple operational conditions," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1416–1426, Sep. 2016.
[26] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[27] M. Seeger, "Gaussian processes for machine learning," *Int. J. Neural. Syst.*, vol. 14, no. 2, pp. 69–106, Apr. 2004.
[28] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[29] 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.
[30] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2003.
[31] W. B. Wang, "A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance," *Int. J. Prod. Res.*, vol. 38, no. 6, pp. 1425–1436, Nov. 2000.
[32] J. C. Lu, J. Park, and Q. Yang, "Statistical inference of a time-to-failure distribution derived from linear degradation data," *Technometrics*, vol. 39, no. 4, pp. 391–400, Nov. 1997.
[33] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Life distributions from component degradation signals: A neural net approach," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 694–700, Jun. 2004.
[34] N. Gebraeel, "Prognostics-based identification of the top-k units in a fleet," *IEEE Trans. Autom. Sci. Eng.*, vol. 7, no. 1, pp. 37–48, Jan. 2010.
[35] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*. Philadelphia, PA, USA: SIAM, 2011.
[36] G. H. Golub and Q. Ye, "An inverse free preconditioned Krylov subspace method for symmetric generalized eigenvalue problems," *SIAM J. Sci. Comput.*, vol. 24, no. 1, pp. 312–334, Jul. 2002.
[37] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Rev.*, vol. 40, no. 3, pp. 636–666, 1998.
[38] S. Sarkar, X. Jin, and A. Ray, "Data-driven fault detection in aircraft engines with noisy sensor measurements," *ASME J. Eng. Gas Turbines Power*, vol. 133, no. 8, Aug. 2011, Art. no. 081602.
[39] C. A. Micchelli and M. Pontil, "Learning the kernel function via regularization," *J. Mach. Learn. Res.*, vol. 6, pp. 1099–1125, Jul. 2005.

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, China, in 2015. Currently, he is working toward the Ph.D. degree at the Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA.

His research interests are focused on process modeling and data fusion.

Kaibo Liu (M'14) received the B.S. degree in industrial engineering and engineering management 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.

Currently, he is an Assistant Professor at the Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA. His research interests are focused on data fusion for process modeling, monitoring, diagnosis, and prognostics.

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

Xi Zhang (M'14) received the B.S. degrees in mechanical engineering and automation from Shanghai Jiaotong University, Shanghai, China, in 2006, and the Ph.D. degree in industrial engineering from the University of South Florida, Tampa, FL, USA, in 2010.

Currently, he is an Assistant Professor at the Department of Industrial Engineering and Management, Peking University, Beijing, China. His research interests are focused on data-driven modeling and analysis for complex dynamic systems including physiological-data-driven analysis for healthcare delivery.

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