Integrating Local Models and Importance Sampling for Non-parametric Degradation Modeling

Changyue Song, *Member, IEEE*, Andi Wang, Di Wang, Member, IEEE

***Abstract*—Sensors have been widely used to monitor the degradation process of various engineering systems and units. Based on the collected sensor signals, previous studies have developed different models to predict the remaining useful life (RUL) of the unit and make maintenance decisions. The majority of the existing models are parametric and usually have restrictive assumptions such as the specific functional form of the sensor signals. This study proposes a non-parametric model with few assumptions that can be potentially adopted under various practical situations. The main idea is to predict the RUL of the in-field unit based only on the historical units with the most similar degradation trends. Specifically, we first extract features to characterize the progression of each sensor signal, based on which the distance between units can be defined. Then, local models are constructed to capture the relationship between features and failure time for RUL prediction. A challenge is that the limited sensor measurements from the in-field unit may cause high uncertainty of the features and lead to unreliable predictions. To address this issue, this study integrates the importance sampling (IS) technique to incorporate the population-level information, which improves the prognostic performance. A case study with two datasets on the degradation of aircraft engines is used to validate the proposed method.**

***Index Terms*—Non-parametric degradation modeling, heterogeneous units, importance sampling, local linear regression**

I. Introduction

SENSOR-BASED prognostics has attracted much research interest in recent years due to the rapid development of the Internet of Things. In particular, sensors are installed on the operating engineering units to collect measurements that reflect the conditions of the units. By analyzing the signals recorded by these sensors, the remaining useful life (RUL) of the units can be predicted in real time to enable effective and efficient maintenance decisions. This can enhance the system reliability and achieve relatively low maintenance costs.

Degradation models play a critical role in analyzing the sensor signals, inferring the underlying degradation processes, and predicting the RUL. Recent reviews on the related literature can be found in [1]–[4]. Existing degradation models can be generally categorized into machine learning models and statistical models. Machine learning models usually regard the relationship between the sensor signals and the RUL as a black box, and the RUL prediction problem can be solved by machine learning or deep learning techniques [5]–[8]. In contrast, statistical models describe the system dynamics with a probabilistic model and thus usually enjoy better interpretability. Therefore, this study focuses on statistical models.

Statistical degradation models can be further divided into parametric models and non-parametric models. Some of the most widely used parametric models include the mixed-effect models [9]–[11] and stochastic processes such as the Wiener process [12]–[14], the gamma process [15], [16], and the inverse-Gaussian process [17], [18]. Parametric models are usually based on the assumptions on the dynamics of the degradation process. For example, mixed-effect models usually assume that the sensor signals have specific functional form and use random parameters with specific prior distributions to characterize the individual variations of units. Stochastic processes also rely on the assumption of specific parametric distributions for the signal. Parametric models can achieve good performance when the assumptions are satisfied. However, many parametric models require some prior knowledge such as the functional form of the signal, and may not be applicable in other situations. In contrast, non-parametric models make few assumptions to achieve better flexibility. For example, Kontar *et al*. proposed to use a multivariate Gaussian process to model the intercorrelation of the signals among units [19]. Enlightened by the recommender systems, Chehade and Liu proposed to represent the signal of a unit as a linear combination of other units' signals [20]. Since non-parametric models incorporate less engineering knowledge, they usually require a large sample size. In addition, a common limitation of many existing statistical degradation models is that they predict the RUL mainly through forecasting the future sensor measurements with the assumption of a pre-specified failure threshold. However, in practice, the failure threshold may not be straightforward to determine or does not even exist. For example, if there are multiple failure modes, units of different failure modes may require different failure thresholds, which will be difficult to determine when little prior information is available [21].

Recently, Song *et al*. proposed a new non-parametric degradation model to directly predict the failure time of the in-service unit based on its nearest historical units [22]. In particular, features are extracted from all units to measure the unit similarity. Then, local linear models are constructed to model the relationship between features and the failure

Changyue Song is with the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: csong14@stevens.edu).

Andi Wang is with the School of Manufacturing Systems and Networks, Arizona State University, Tempe, AZ 85281 USA (e-mail: Andi.Wang@asu.edu).

Di Wang is with the Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: d.wang@sjtu.edu.cn).

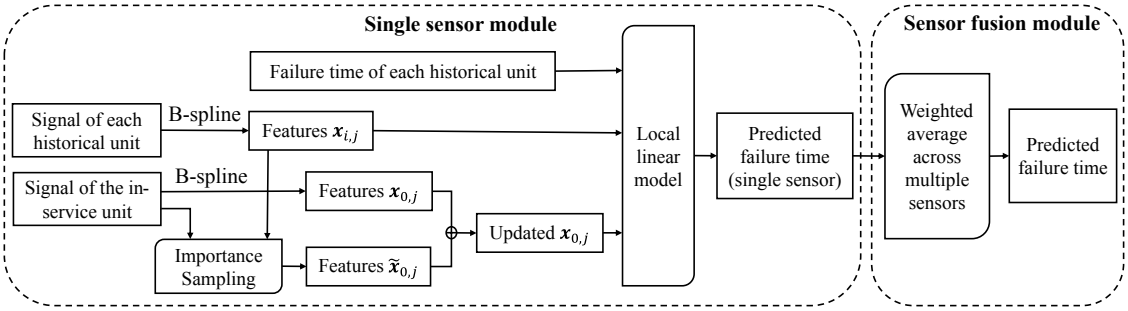


Fig. 1. Overview of the proposed method.

time. In this way, the failure time of the in-service unit can be predicted. This method enjoys the advantages of high flexibility and good interpretability, without the need of a pre-specified failure threshold. However, a critical limitation of [22] is the unsatisfactory prognostic performance for in-service units in the early degradation stage. The main issue lies in the extracted features of the in-service unit. Specifically, if the in-service unit is in the early degradation stage, the available sensor measurements are limited and there is little information about the degradation trend. Consequently, the extracted features from the in-service unit usually has high variance and thus leads to unreliable prediction.

This paper aims to fill this gap and achieve good performance even for early-stage in-service units. Enlightened by the mixed-effect models [9], [10], the key idea is to extract population-level information from the historical units and incorporate the information into the features of the in-service unit. Instead of describing the population-level information as a parametric prior distribution for the features, this study considers a non-parametric representation by regarding historical units as independent samples from the distribution. This ensures the flexibility of the model. Then, importance sampling (IS) is adopted to incorporate the collected sensor measurements from the in-service unit and estimate the posterior distribution [23], [24].

Figure 1 shows the overview of the proposed method. Each sensor is analyzed separately using the single sensor module and then the output of the modules for all sensors are combined using the sensor fusion module. In the single sensor module, features are extracted from the sensor measurements of each historical unit and the in-service unit, using monotonic B-spline regression. Since the features extracted from the in-service unit may be unreliable, importance sampling method is applied to provide another estimation and update the in-service unit features. Then, a local linear model is constructed to predict the failure time of the in-service unit based on the features and the failure time of the historical units. The sensor fusion module computes the weighted average of each sensor's prediction to obtain the final prediction.

The remaining of this paper is organized as follows. Section II briefly introduces the mixed-effect models. Section III elaborates the proposed method. Section IV presents the case study, and Section V draws the conclusion.

II. Review of Mixed-Effect Models

Mixed-effect models have been widely used for degradation modeling [9], [10], [25]–[27], which usually represent the sensor signal with the following functional form

$$L_{i,j}(t) = \eta_{i,j}(t) + \varepsilon_{i,j}(t) = \eta(t|\boldsymbol{\alpha}_{i,j}) + \varepsilon_{i,j}(t). \quad (1)$$

Here $L_{i,j}(t)$ is the measurement collected from unit i, sensor j, at time t, $\eta_{i,j}(t) = \eta(t|\boldsymbol{\alpha}_{i,j})$ is a pre-specified function to represent the signal path, $\boldsymbol{\alpha}_{i,j}$ is the random-effect parameter that captures the unit-to-unit variations, and $\varepsilon_{i,j}(t) \sim N(0, \sigma^2_{\varepsilon,j})$ is the measurement error.

The advantage of mixed-effect models is the ability to incorporate population-level information by assuming a prior distribution G_j for the random-effect parameter $\boldsymbol{\alpha}_{i,j}$ as

$$\boldsymbol{\alpha}_{i,j} \sim G_j, \quad (2)$$

where G_j can be estimated based on the historical units. Consequently, if sensor measurements have been collected at time epochs $\{t^1_{i,j}, t^2_{i,j}, \cdots\}$, the posterior distribution of the random-effect parameter $\boldsymbol{\alpha}_{i,j}$ can be updated as

$$p(\boldsymbol{\alpha}_{i,j}|\boldsymbol{L}_{i,j}) \propto p(\boldsymbol{L}_{i,j}|\boldsymbol{\alpha}_{i,j})p(\boldsymbol{\alpha}_{i,j}), \quad (3)$$

where $\boldsymbol{L}_{i,j} = [L_{i,j}(t^1_{i,j}), L_{i,j}(t^2_{i,j}), \cdots]^T$ is the collection of all available measurements from the sensor j of unit i, the likelihood $p(\boldsymbol{L}_{i,j}|\boldsymbol{\alpha}_{i,j})$ can be evaluated based on (1), and the prior probability distribution $p(\boldsymbol{\alpha}_{i,j})$ is given as G_j above. If the unit has a large amount of sensor measurements with clear degradation trend, the likelihood $p(\boldsymbol{L}_{i,j}|\boldsymbol{\alpha}_{i,j})$ will have a high weight to affect the posterior distribution $p(\boldsymbol{\alpha}_{i,j}|\boldsymbol{L}_{i,j})$. Therefore, the posterior will generally be determined from the observed sensor measurements. On the other hand, if the in-service unit is at an early stage with few measurements, the prior distribution $p(\boldsymbol{\alpha}_{i,j})$ will have a high weight, and the posterior distribution $p(\boldsymbol{\alpha}_{i,j}|\boldsymbol{L}_{i,j})$ will follow the general population behavior.

Existing mixed-effect models usually make explicit assumptions on the parametric forms of the signal path $\eta(t|\boldsymbol{\alpha}_{i,j})$ and the prior distribution G_j. This restricts the flexibility of existing mixed-effect models because if the parametric forms are not properly specified, the model performance cannot be guaranteed. In addition, the parametric forms may be difficult to specify, for example, when multiple failure modes exist. In this study, we borrow the idea of mixed-effect models but adopt a non-parametric representation to ensure the model flexibility.

III. METHODOLOGY

Denote the in-service unit of interest as unit 0, and denote $\mathcal{M} = \{1, 2, \cdots, m\}$ as the set of historical units that have failed. The failure time of each historical unit i is known as T_i. Assume each unit is monitored by s sensors. The objective is to predict the failure time of the in-service unit.

A. Feature Extraction

The first step is to extract features to characterize the progression of the sensor signals. In particular, we decompose each sensor signal as

$$L_{i,j}(t) = \eta_{i,j}(t) + \varepsilon_{i,j}(t), i \in \{0\} \cup \mathcal{M}. \tag{4}$$

The signal path $\eta_{i,j}(t)$ is regarded as a function of time t. Denote τ as the time when the latest measurement of the in-service unit was collected. Following [22], the zero-order, first-order, and second-order derivatives of $\eta_{i,j}(t)$ at time τ are extracted from each unit. In other words, we consider to use $[\eta_{i,j}^{(0)}(\tau), \eta_{i,j}^{(1)}(\tau), \eta_{i,j}^{(2)}(\tau)]^T$ to characterize each sensor signal, where $\eta_{i,j}^{(p)}(t)$ denotes the pth-order derivative of $\eta_{i,j}(t)$, $\eta_{i,j}^{(0)}(t) = \eta_{i,j}(t)$, and $i \in \{0\} \cup \mathcal{M}$. Here, the same time point τ is considered for the in-service unit and all historical units to provide a fair comparison.

To extracted the features, an important step is to estimate $\eta_{i,j}(t)$. Due to the great flexibility to approximate complex functions, in this study, the monotonic B-spline with infinite support is adopted for estimating $\eta_{i,j}(t)$. As discussed in [28], we can easily restrict the fitted sensor signal to be monotonic by imposing appropriate constraints, which is desired in degradation modeling because degradation process is monotonic in nature. A degree of 2 is chosen for the monotonic B-spline to avoid overfitting but it still provides enough flexibility [28], [29].

To ensure the model flexibility, each sensor signal is separately considered with different sets of basis functions. Denote $\boldsymbol{\psi}_{i,j}(t) = [\psi_{i,j}^0(t), \cdots, \psi_{i,j}^{K+2}(t)]^T$ as the collection of the basis functions for $L_{i,j}(t)$, where K is the number of knots. Note that the value of K should also be determined separately for each signal. The sensor measurement $L_{i,j}(t)$ can be decomposed as a linear combination of the basis functions

$$L_{i,j}(t) = \boldsymbol{\psi}_{i,j}(t)^T \boldsymbol{\Gamma}_{i,j} + \varepsilon_{i,j}(t), \tag{5}$$

where $\eta_{i,j}(t) = \boldsymbol{\psi}_{i,j}(t)^T \boldsymbol{\Gamma}_{i,j}$, and $\boldsymbol{\Gamma}_{i,j}$ is a $(K+3) \times 1$ vector of coefficients. Equation (5) can be written in matrix format as

$$\boldsymbol{L}_{i,j} = \boldsymbol{\Psi}_{i,j} \boldsymbol{\Gamma}_{i,j} + \boldsymbol{\varepsilon}_{i,j}, \tag{6}$$

where $\boldsymbol{L}_{i,j} = [L_{i,j}(t_{i,j}^1), L_{i,j}(t_{i,j}^2), \cdots]^T$ is the collection of all available measurements for unit i, sensor j, at time epochs $\{t_{i,j}^1, t_{i,j}^2, \cdots\}$, $\boldsymbol{\Psi}_{i,j} = \left[\boldsymbol{\psi}_{i,j}\left(t_{i,j}^1\right)^T; \boldsymbol{\psi}_{i,j}\left(t_{i,j}^2\right)^T; \cdots\right]$, and $\boldsymbol{\varepsilon}_{i,j} = [\varepsilon_{i,j}(t_{i,j}^1), \varepsilon_{i,j}(t_{i,j}^2), \cdots]^T$. Consequently, the coefficient $\boldsymbol{\Gamma}_{i,j}$ can be estimated as

$$\hat{\boldsymbol{\Gamma}}_{i,j} = \underset{\boldsymbol{\Gamma}}{\operatorname{argmin}} (\boldsymbol{L}_{i,j} - \boldsymbol{\Psi}_{i,j}\boldsymbol{\Gamma})^T (\boldsymbol{L}_{i,j} - \boldsymbol{\Psi}_{i,j}\boldsymbol{\Gamma}). \tag{7}$$

According to [28], the monotonicity of $\eta_{i,j}(t)$ can be ensured by imposing proper constraints. For example, the following constraints ensure $\eta_{i,j}(t)$ to be increasing:

$$\begin{aligned} \Gamma_{i,j}^k &\le 0, k = 0, 1 \\ \Gamma_{i,j}^k - \Gamma_{i,j}^{k-1} &\ge 0, k = 3, \cdots, K, \\ \Gamma_{i,j}^k &\ge 0, k = K+1, K+2, \end{aligned} \tag{8}$$

where $\Gamma_{i,j}^k$ is the kth entry of $\boldsymbol{\Gamma}_{i,j}$.

After the estimation $\hat{\boldsymbol{\Gamma}}_{i,j}$ is obtained, the features can be calculated as $\eta_{i,j}^{(p)}(\tau) = \boldsymbol{\psi}_{i,j}^{(p)}(\tau)^T \hat{\boldsymbol{\Gamma}}_{i,j}$, where $\boldsymbol{\psi}_{i,j}^{(p)}(t)$ is the pth-order derivative of $\boldsymbol{\psi}_{i,j}(t)$ obtained as in [22], [28]. At last, the extracted features are standardized as

$$x_{i,j}^{(p)} = \frac{\eta_{i,j}^{(p)}(\tau) - \mu_j^{(p)}}{\sigma_j^{(p)}}, \tag{9}$$

where $\mu_j^{(p)}$ and $\sigma_j^{(p)}$ are the average and standard deviation of $\{\eta_{i,j}^{(p)}(\tau), \forall i \in \mathcal{M}\}$. Denote $\boldsymbol{x}_{i,j} = \left[x_{i,j}^{(0)}, x_{i,j}^{(1)}, x_{i,j}^{(2)}\right]^T$ as the standardized features.

B. Importance Sampling

One challenge lies in the uncertainty or variance of the extracted features $\boldsymbol{x}_{0,j}$ for the in-service unit. Since historical units usually have a large number of sensor measurements and clear degradation trend, the uncertainty of the extracted $\{\boldsymbol{x}_{i,j}, i \in \mathcal{M}\}$ is small. However, if the in-service unit is in the early degradation stage, the collected sensor measurements can be limited and the degradation trend may not be clear. In this case, the measurement errors $\varepsilon_{0,j}(t)$ have a significant impact on the estimated $\hat{\boldsymbol{\Gamma}}_{0,j}$, which leads to unreliable features $\boldsymbol{x}_{0,j}$. This issue has not been considered in the literature.

Enlightened by mixed-effect models, this study proposes to consider a prior distribution G_j to describe the population-level behavior. If the in-service unit is in the early degradation stage, the prior distribution will guide the estimation to follow the general population behavior. In this way, we can improve the reliability of the extracted features $\boldsymbol{x}_{0,j}$. However, G_j in mixed-effect models is usually assumed to have a parametric form, which limits its flexibility. For example, if G_j is assumed to be a single-modal distribution such as the Gaussian distribution, the mixed-effect model will not be applicable when the units are from multiple sub-populations corresponding to significantly different degradation behaviors.

To address the challenge, this study considers a non-parametric representation of G_j. We borrow the idea from particle filtering to represent the prior distribution with a number of samples, and use importance sampling (IS) to derive the posterior distribution by re-weighting the samples [23], [24]. In particular, we assume the functional form of the signal path $\eta_{i,j}(t) = \eta(t|\boldsymbol{\alpha}_{i,j})$ as in (1) and $\boldsymbol{\alpha}_{i,j} \sim G_j$. Note that we only assume the functional form to *exist* instead of specifying it, which ensures the model to be flexible.

We regard the unknown "true" features of unit i as a random variable $X_{i,j}$, and $\boldsymbol{x}_{i,j}$ directly extracted from the sensor signal in Section III-A can be regarded as an estimation of $X_{i,j}$. In addition, if $\boldsymbol{\alpha}_{i,j}$ is known, the signal path $\eta_{i,j}(t) = \eta(t|\boldsymbol{\alpha}_{i,j})$

is a deterministic function of t and then $X_{i,j}$ can be easily calculated based on the derivatives of $\eta_{i,j}(t)$. In other words, $X_{i,j}$ can be regarded as a deterministic mapping of $\boldsymbol{\alpha}_{i,j}$, i.e., $X_{i,j} = g(\boldsymbol{\alpha}_{i,j})$. Let the posterior distribution $\boldsymbol{\alpha}_{0,j}|\boldsymbol{L}_{0,j} \sim \tilde{G}_j$, our objective is to calculate the expectation of $X_{0,j}$ with respect to $\tilde{G}_j$ as

$$\begin{aligned} E_{\tilde{G}_j}[X_{0,j}] &= E_{\tilde{G}_j}[g(\boldsymbol{\alpha}_{0,j})] \\ &= \int g(\boldsymbol{\alpha}_{0,j}) p(\boldsymbol{\alpha}_{0,j}|\boldsymbol{L}_{0,j}) d\boldsymbol{\alpha}_{0,j} \\ &\propto \int g(\boldsymbol{\alpha}_{0,j}) p(\boldsymbol{L}_{0,j}|\boldsymbol{\alpha}_{0,j}) p(\boldsymbol{\alpha}_{0,j}) d\boldsymbol{\alpha}_{0,j} \\ &= E_{G_j}[g(\boldsymbol{\alpha}_{0,j}) p(\boldsymbol{L}_{0,j}|\boldsymbol{\alpha}_{0,j})]. \end{aligned} \tag{10}$$

Equation (10) implies that by drawing random samples from G_j, we can approximate $E_{\tilde{G}_j}[X_{0,j}]$ by the corresponding sample average. In particular, $\boldsymbol{\alpha}_{i,j}$ from each historical unit can be regarded as a random sample from G_j, and thus

$$\begin{aligned} E_{\tilde{G}_j}[X_{0,j}] \propto & E_{G_j}[g(\boldsymbol{\alpha}_{0,j}) p(\boldsymbol{L}_{0,j}|\boldsymbol{\alpha}_{0,j})] \\ \approx & \frac{1}{m} \sum_{i=1}^{m} g(\boldsymbol{\alpha}_{i,j}) p(\boldsymbol{L}_{0,j}|\boldsymbol{\alpha}_{0,j} = \boldsymbol{\alpha}_{i,j}) \\ = & \frac{1}{m} \sum_{i=1}^{m} X_{i,j} w_{i,j}. \end{aligned} \tag{11}$$

Here m is the total number of historical units. Since historical units usually have a large amount of measurements with clear degradation trend, the extracted features $\boldsymbol{x}_{i,j}$ can be regarded as a good estimation of $X_{i,j}$ with negligible variance, and thus we can simply replace $X_{i,j}$ with $\boldsymbol{x}_{i,j}$ in (11). The likelihood $p(\boldsymbol{L}_{0,j}|\boldsymbol{\alpha}_{0,j} = \boldsymbol{\alpha}_{i,j})$ can be regarded as the weight of unit i and be calculated as

$$\begin{aligned} w_{i,j} &= p(\boldsymbol{L}_{0,j}|\boldsymbol{\alpha}_{0,j} = \boldsymbol{\alpha}_{i,j}) \\ &= p(\boldsymbol{L}_{0,j}|\eta_{0,j}(t) = \eta_{i,j}(t)) \\ &\approx p(\boldsymbol{L}_{0,j}|\eta_{0,j}(t) = \hat{\eta}_{i,j}(t)) \\ &= N(\boldsymbol{L}_{0,j}|\boldsymbol{\Psi}_{i,j}^0 \hat{\boldsymbol{\Gamma}}_{i,j}, \hat{\sigma}_{\varepsilon,j}^2 \boldsymbol{I}) \\ &\propto \exp\Big\{ -\frac{1}{2\hat{\sigma}_{\varepsilon,j}^2} \sum_n \Big[L_{0,j}(t_{0,j}^n) - \boldsymbol{\psi}_{i,j}(t_{0,j}^n)^T \hat{\boldsymbol{\Gamma}}_{i,j} \Big]^2 \Big\}, \end{aligned} \tag{12}$$

where $\boldsymbol{I}$ is the identify matrix, $N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the likelihood of a normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{x}$, and $\boldsymbol{\Psi}_{i,j}^0 = [\boldsymbol{\psi}_{i,j}(t_{0,j}^0)^T; \boldsymbol{\psi}_{i,j}(t_{0,j}^1)^T; \cdots]$ is obtained by evaluating the functional basis $\boldsymbol{\psi}_{i,j}(t)$ for unit i at time points $\{t_{0,j}^n, n = 1, 2, \cdots\}$ when sensor measurements are collected from the in-service unit. Note that in (12), we simply replace $\eta_{i,j}(t)$ with $\hat{\eta}_{i,j}(t) = \boldsymbol{\psi}_{i,j}(t)^T \hat{\boldsymbol{\Gamma}}_{i,j}$ because for historical units, the fitted signal path $\hat{\eta}_{i,j}(t)$ should be a good estimation of $\eta_{i,j}(t)$ with negligible variance. If the historical unit i has a similar degradation trend as the in-service unit, the likelihood $N(L_{0,j}(t)|\boldsymbol{\psi}_{i,j}(t)^T \boldsymbol{\Gamma}_{i,j}, \sigma_{\varepsilon,j}^2)$ will be large, leading to higher weight of unit i. The estimated variance of the sensor noise $\hat{\sigma}_{\varepsilon,j}^2$ can be obtained by

$$\hat{\sigma}_{\varepsilon,j}^2 = \frac{(\boldsymbol{L}_{0,j} - \boldsymbol{\Psi}_{0,j} \hat{\boldsymbol{\Gamma}}_{0,j})^T (\boldsymbol{L}_{0,j} - \boldsymbol{\Psi}_{0,j} \hat{\boldsymbol{\Gamma}}_{0,j})}{N_{0,j} - K - 3}, \tag{13}$$

where $N_{i,j}$ is the total number of sensor measurements collected from sensor j for unit i. Therefore, the posterior mean of the in-service feature $X_{0,j}$ can be calculated as

$$\tilde{\boldsymbol{x}}_{0,j} = E_{\tilde{G}_j}[X_{0,j}] \propto \sum_{i=1}^{m} w_{i,j} \boldsymbol{x}_{i,j}. \tag{14}$$

As a result, two sets of features are obtained for the in-service unit, including the features $\boldsymbol{x}_{0,j}$ directly extracted from the sensor signal as in Section III-A and the features $\tilde{\boldsymbol{x}}_{0,j}$ calculated by IS as described above. Theoretically, if the number of historical units is large enough, $\tilde{\boldsymbol{x}}_{0,j}$ is a good estimation of the features. However, in practice, the number of historical units is usually limited, and thus $\tilde{\boldsymbol{x}}_{0,j}$ can be biased. Therefore, we propose to combine $\boldsymbol{x}_{0,j}$ and $\tilde{\boldsymbol{x}}_{0,j}$, and update $\boldsymbol{x}_{0,j}$ as:

$$\boldsymbol{x}_{0,j} \leftarrow \lambda \tilde{\boldsymbol{x}}_{0,j} + (1 - \lambda) \boldsymbol{x}_{0,j}. \tag{15}$$

Here $\lambda \in [0, 1]$ is a tuning parameter that represents the weight of $\tilde{\boldsymbol{x}}_{0,j}$. If the in-service unit is in the early degradation stage and $\boldsymbol{x}_{0,j}$ contains high uncertainty, λ can be set close to 1 to rely more on $\tilde{\boldsymbol{x}}_{0,j}$. On the other side, if the in-service unit has shown clear degradation trend and the variance of $\boldsymbol{x}_{0,j}$ is low, λ can be set close to 0 to mitigate the bias of $\tilde{\boldsymbol{x}}_{0,j}$. The selection of λ is discussed in Section III-E.

The algorithm of importance sampling to update the in-service unit features is summarized as follows.

Algorithm 1 Update $\boldsymbol{x}_{0,j}$ by Importance Sampling

Input: extracted features $\{\boldsymbol{x}_{i,j}, i \in \{0\} \cup \mathcal{M}\}$, monotonic B-spline basis $\{\boldsymbol{\psi}_{i,j}(t), i \in \mathcal{M}\}$, in-service unit sensor measurements $\{L_{0,j}(t), t = t_{0,j}^1, t_{0,j}^2, \cdots\}$
Output: updated in-service features $\boldsymbol{x}_{0,j}$
Algorithm:
1. For each historical unit $i \in \mathcal{M}$:
   Calculate the weight $w_{i,j}$ according to (12)
2. Normalize $w_{i,j}$ by
$$w_{i,j} \leftarrow \frac{w_{i,j}}{\sum_i w_{i,j}}$$
3. Calculate $\tilde{\boldsymbol{x}}_{0,j}$ by (14)
4. Update $\boldsymbol{x}_{0,j}$ by (15)

C. Local Linear Regression

After feature extraction, local linear regression (LLR) is adopted to establish the relationship between features $\boldsymbol{x}$ and failure time T. The LLR approximate the nonlinear relationship between $\boldsymbol{x}$ and T with a local linear relationship on a small region around $\boldsymbol{x}_{0,j}$. In particular, we consider the following linear model

$$T_0 = [1, \boldsymbol{x}_{0,j}^T] \boldsymbol{\beta}_{0,j} + \rho_j, \tag{16}$$

where T_0 is the failure time of the in-service unit, and $\boldsymbol{\beta}_{0,j}$ and ρ_j are the model coefficients and error, respectively, based on sensor j. The coefficients $\boldsymbol{\beta}_{0,j}$ can be estimated based on the nearest historical units. Specifically, a weight is assigned

to each historical unit based on its distance from the in-service unit

$$k_{i,j} = \mathcal{K}_j(c_{i,j}) = \mathcal{K}_j(\|\boldsymbol{x}_{i,j} - \boldsymbol{x}_{0,j}\|^2). \tag{17}$$

Here $c_{i,j} = \|\boldsymbol{x}_{i,j} - \boldsymbol{x}_{0,j}\|^2$ denotes the Euclidean distance between the in-service unit and the historical unit i measured with sensor j, and $\mathcal{K}_j$ is a smoothing kernel function. A historical unit is assigned with a higher weight if it is close to the in-service unit, and it is assigned a small or even zero weight if it is far from the in-service unit. In this way, the coefficients $\boldsymbol{\beta}_{0,j}$ can be estimated by minimizing

$$\begin{aligned} h_j(\boldsymbol{\beta}) &= \sum_{i\in\mathcal{M}} k_{i,j}(T_i - [1, \boldsymbol{x}_{i,j}^T]\boldsymbol{\beta})^2 \\ &= (\boldsymbol{T} - \boldsymbol{X}_j\boldsymbol{\beta})^T \boldsymbol{K}_j(\boldsymbol{T} - \boldsymbol{X}_j\boldsymbol{\beta}), \end{aligned} \tag{18}$$

where $\boldsymbol{X}_j = [1, \boldsymbol{x}_{i,j}^T]_{i\in\mathcal{M}}$ is the m-by-4 design matrix, $\boldsymbol{T} = [T_1, T_2, \cdots]^T$ is the m-by-1 response vector, and $\boldsymbol{K}_j = \text{diag}(k_{1,j}, k_{2,j}, \cdots)$ is the diagonal weight matrix. The estimation is given by

$$\hat{\boldsymbol{\beta}}_{0,j} = \underset{\boldsymbol{\beta}}{\text{argmin}}\, h_j(\boldsymbol{\beta}) = (\boldsymbol{X}_j^T\boldsymbol{K}_j\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^T\boldsymbol{K}_j\boldsymbol{T} \tag{19}$$

Therefore, the failure time of the in-service unit can be predicted as

$$\hat{T}_{0,j} = [1, \boldsymbol{x}_{0,j}^T]\hat{\boldsymbol{\beta}}_{0,j} \tag{20}$$

based on sensor j.

As can be seen, by using LLR, no restrictive assumption is imposed on the relation between features $\boldsymbol{x}$ and failure time T and thus ensures the flexibility. For example, if sub-populations exist where units in different sub-populations exhibit significantly different behaviors, historical units from the same sub-population of the in-service unit will be assigned with higher weights and historical units from different sub-populations will be ignored. This enables the LLR to automatically determine the sub-population of the in-service unit and make accurate predictions.

D. Sensor Fusion

According to the procedures above, each sensor gives a different prediction $\hat{T}_{0,j}$ of the failure time of the in-service unit. These predictions can be combined into a more accurate prediction. To achieve this, the first step is to assess the fidelity of each prediction. Following [22], we consider the prediction variance of $\hat{T}_{0,j}$

$$v_j = \text{Var}(T_0 - \hat{T}_{0,j}) = [1, \boldsymbol{x}_{0,j}^T]\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_{0,j}}\begin{bmatrix}1\\ \boldsymbol{x}_{0,j}\end{bmatrix} + \hat{\sigma}_{\rho,j}^2, \tag{21}$$

where the covariance matrix of $\hat{\boldsymbol{\beta}}_{0,j}$ is

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_{0,j}} = \hat{\sigma}_{\rho,j}^2(\boldsymbol{X}_j^T\boldsymbol{K}_j\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^T\boldsymbol{K}_j^2\boldsymbol{X}_j(\boldsymbol{X}_j^T\boldsymbol{K}_j\boldsymbol{X}_j)^{-1}, \tag{22}$$

and

$$\hat{\sigma}_{\rho,j}^2 = \frac{h_j(\hat{\boldsymbol{\beta}}_{0,j})}{\text{tr}(\boldsymbol{K}_j) - \text{tr}[(\boldsymbol{X}_j^T\boldsymbol{K}_j\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j^T\boldsymbol{K}_j^2\boldsymbol{X}_j]}. \tag{23}$$

See Chapter 2 and Chapter 9 of [30] for the detailed derivation. A higher prediction variance v_j means that the predicted failure time $\hat{T}_{0,j}$ by sensor j has higher uncertainty and thus should be assigned with a smaller weight. The predictions of the failure time from individual sensors are combined as

$$\hat{T}_0 = \frac{\sum_j u_j\hat{T}_{0,j}}{\sum_j u_j}, \tag{24}$$

where

$$u_j = \begin{cases}\left(\frac{1}{v_j}\right)^\gamma, & v_j \le v^{(q)} \\ 0, & v_j > v^{(q)}\end{cases} \tag{25}$$

is the weight of sensor j, $v^{(q)}$ is the qth smallest value of $\{v_j, j = 1, \cdots, s\}$, and $\gamma > 0$ is a positive tuning parameter. Sensors with a smaller prediction variance will have higher weights, and only the q sensors with the smallest prediction variance are used. In this way, we only combine the most reliable sensors.

E. Selection of λ

The parameter λ should be determined based on the degradation stage of the in-service unit. In this study, we consider the signal range to estimate the degradation stage. In particular, the signal range of historical unit $i \in \mathcal{M}$, sensor j is calculated as the difference in the fitted signal between the failure time T_i and 0 as $\hat{\eta}_{i,j}(T_i) - \hat{\eta}_{i,j}(0)$, and for the in-service unit, it is defined as $\hat{\eta}_{0,j}(\tau) - \hat{\eta}_{0,j}(0)$, where $\hat{\eta}_{i,j}(t) = \boldsymbol{\psi}_{i,j}(t)^T\hat{\boldsymbol{\Gamma}}_{i,j}, i \in \{0\}\cup\mathcal{M}$. Then, we define the relative signal range as

$$r_{i,j} = \frac{\hat{\eta}_{0,j}(\tau) - \hat{\eta}_{0,j}(0)}{\hat{\eta}_{i,j}(T_i) - \hat{\eta}_{i,j}(0)}. \tag{26}$$

If $r_{i,j} < 0$, it means the in-service unit and the historical unit i have different sensor trend, i.e., one is increasing and one is decreasing, and thus they are probably from different sub-populations. Therefore, for each sensor j, historical units with negative $r_{i,j}$s are ignored.

For any $r_{i,j} > 0$, ideally, if the in-service unit is in the late degradation stage, $r_{i,j}$ should be close to 1, and otherwise $r_{i,j}$ should be close to 0. However, there will be variations among different historical units because different historical units usually have different signal ranges. If the variations of $r_{i,j}$ corresponding to a specific sensor j is large, it means that the corresponding sensor may not be reliable in determining the degradation stage of the in-service unit. Therefore, in this study, only the sensor with the smallest variance in $r_{i,j}$ is considered, and the index of this sensor is denoted as $p = \text{argmin}_j\{R_j\}$, where R_j is the variance of $\{r_{i,j}, \forall i \in \mathcal{M} : r_{i,j} > 0\}$. Recall that λ is the weight of $\tilde{\boldsymbol{x}}_{0,j}$, and $\tilde{\boldsymbol{x}}_{0,j}$ is calculated based on IS. A large value of $r_{i,p}$ means that the in-service unit is in the late degradation stage and thus λ should be small. Therefore, we set the value of λ by

$$\lambda = \min\{\max\{1 - b\cdot r_p^{\max}, 0\}, 1\}, \tag{27}$$

where $r_p^{\max} = \max\{r_{i,p}, \forall i \in \mathcal{M}\}$ is the maximum value of $\{r_{i,p}\}$, and $b > 0$ is a pre-specified constant. The value $1 - b\cdot r_p^{\max}$ is truncated to be within $[0, 1]$. Since $\tilde{\boldsymbol{x}}_{0,j}$ can be biased, we consider $r_p^{\max}$ and $b > 1$ to prefer the use of $\boldsymbol{x}_{0,j}$ by setting λ to a smaller number. For example, with $b = 2$, if the relative signal range $r_p^{\max} = 0.5$, indicating the in-service

unit to be in the middle degradation stage, $\lambda = 0$ will be used to fully rely on $\boldsymbol{x}_{0,j}$. In other words, even if the in-service unit is in the middle degradation stage and has not reached the maximum signal range, the degradation trend has become clear and the extracted $\boldsymbol{x}_{0,j}$ is reliable enough, and thus we can ignore $\tilde{\boldsymbol{x}}_{0,j}$ to avoid the bias.

IV. Case Study

The proposed method is implemented in two datasets on the degradation of aircraft engines.

A. Data Description

The data used in the case study were generated by C-MAPSS, which is a widely used model to simulate the degradation process of large commercial turbofan engines [31]. It includes four datasets, while in this study, dataset 1 and dataset 3 are considered. Both datasets were generated under a single operational condition. Dataset 1 has only one failure mode but the dataset 3 has two failure modes. Each dataset contains 100 historical units and 100 in-service units, and each unit is monitored by 21 sensors measuring comprehensive information such as temperature and pressure. Table I provides a detailed description of the sensors. Sensor measurements are available until the time of failure for historical units, but are truncated before failure for in-service units. The actual RUL of the in-service units is recorded in a separate file. The objective is to predict the RUL of the in-service units based on the available sensor measurements and assess the prediction accuracy through the comparison with the actual RUL.

TABLE I
Detailed Description of the 21 Sensors

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

B. Method Implementation

At first, sensors that only show constant measurements are excluded, and the remaining 14 sensors are considered for both datasets, including T24, T30, T50, P30, Nf, Nc, Ps30, phi, NRf, NRc, BPR, htBleed, W31 and W32. No further preprocessing is required.

Throughout the case study, the monotonic B-spline with infinite support of degree 2 is used for fitting the sensor signals and extracting the features. The B-spline knots are equally spaced and the number of knots K is separately determined for each unit and each sensor by Bayesian information criterion (BIC). The k-nearest-neighbor smoothing kernel function is considered for constructing the LLR, which is defined as

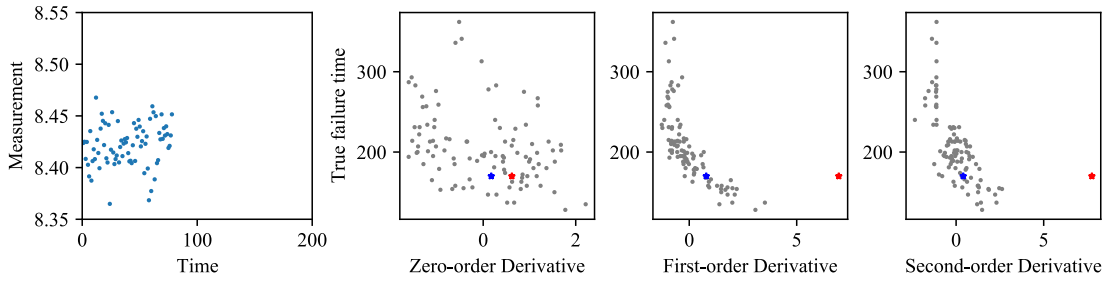
$$\mathcal{K}(c_i) = \begin{cases} 1, & c_i \leq c^{(n)}, \\ 0, & c_i > c^{(n)}. \end{cases} \tag{28}$$

Here $c^{(n)}$ is the nth smallest value of $\{c_i, i \in \mathcal{M}\}$, and the subscript j that represents the index of sensor is omitted for simplicity. This study chooses $n = 15$ in the k-nearest-neighbor kernel function for demonstration. In practice, other kernel functions can also be used, and cross validation can be adopted to choose the best kernel function. The details can be found in [22].

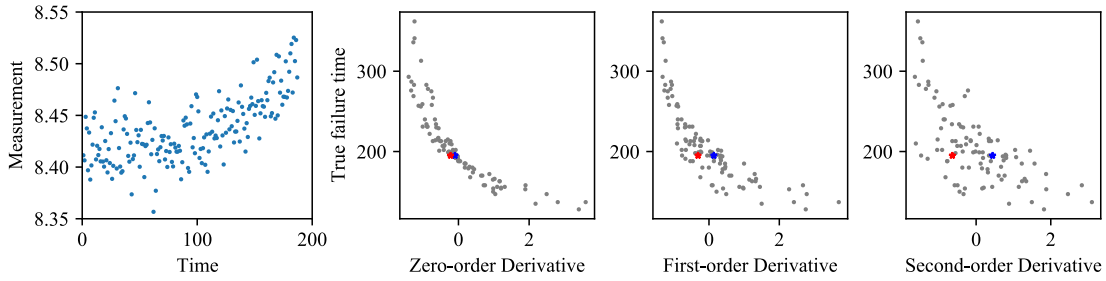
C. Results with a Single Sensor

To begin with, the method is used to predict the RUL based on the measurements from a single sensor. As a demonstration, we consider the sensor BPR and ignore all the other sensors. The parameter b is chosen to be 2 to determine λ as in (27).

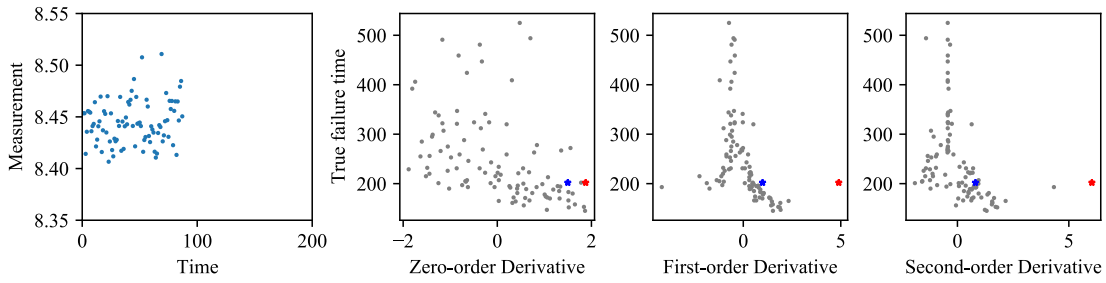
To illustrate the proposed method, we visualize the BPR signal of four selected in-service units in Fig. 2. For each unit, the plot on the left shows the collected sensor measurements. In the next three plots at the right-hand side, the three extracted features are respectively represented as the x-axis, and the true failure time are represented as the y-axis. In these three plots, the grey points represent the historical units, the red point represents the features $\boldsymbol{x}_{0,j}$ of the in-service unit directly extracted from the sensor signal, and the blue point represents the features $\tilde{\boldsymbol{x}}_{0,j}$ of the in-service unit obtained using IS. Figure 2a shows the in-service unit #48 from dataset 1. From the historical units, we can observe that the true failure time does not show strong correlation with the zero-order derivative feature, but shows clear nonlinear relationship with the first-order and second-order derivative features. Since the unit is still in the early degradation stage, the features $\boldsymbol{x}_{0,j}$ directly extracted from sensor signals are unreliable because they are far from all the historical units, especially for the first-order and second-order derivative features. Instead, the features $\tilde{\boldsymbol{x}}_{0,j}$ based on IS are within the same region of the historical units and thus are more reliable. Figure 2b shows the in-service unit #68 from dataset 1. The in-service unit is in the end degradation stage and shows clear trend. The first-order derivative feature now shows strong relationship with the true failure time. Despite some discrepancy possibly caused by the bias of IS, the features $\boldsymbol{x}_{0,j}$ by fitting the sensor signal and $\tilde{\boldsymbol{x}}_{0,j}$ by IS are quite close. Figures 2c and 2d shows in-service units #42 and #64 from dataset 3, respectively. The features clearly show two sub-populations within the units, corresponding to two different failure modes. The proposed method can automatically classify the sub-population of the in-service unit. In addition, similar conclusions can be reached



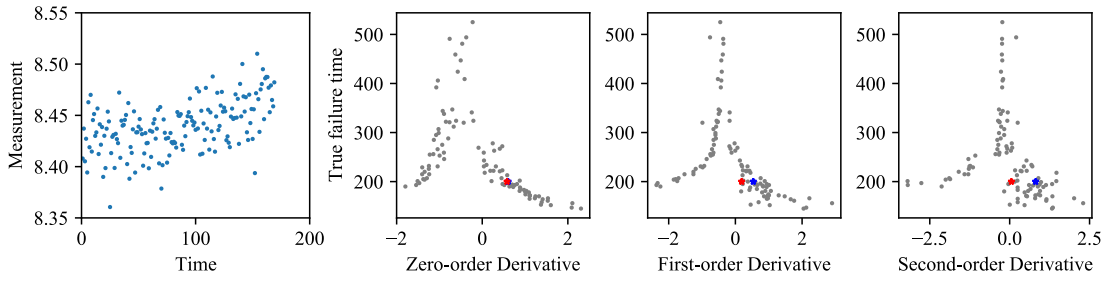
(a) In-service unit #48 from dataset 1



(b) In-service unit #68 from dataset 1



(c) In-service unit #42 from dataset 3



(d) In-service unit #64 from dataset 3

Fig. 2. Illustration of the collected sensor measurements and the true failure time versus the three extracted features. The grey points are historical units. The red points are the in-service unit by directly fitting the sensor signal, and the blue points are the in-service unit based on IS.

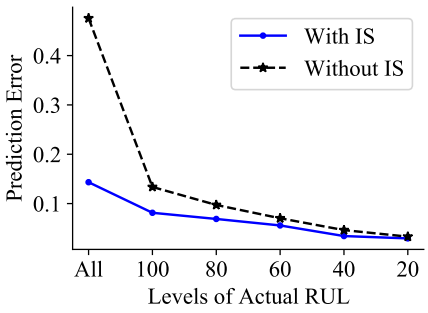
as in Fig. 2a and Fig. 2b regarding the features $\boldsymbol{x}_{0,j}$ and $\tilde{\boldsymbol{x}}_{0,j}$ for the in-service unit. Therefore, by integrating with IS, the proposed method can enhance the reliability of the extracted features for in-service units, especially for those in the early degradation stage.

The proposed method is further used to predict the failure time of the in-service units. The prediction error is evaluated as

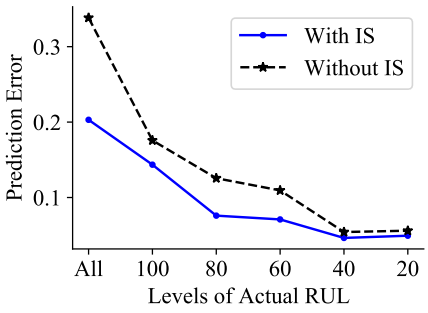
$$\text{Error} = \frac{|\hat{T}_0 - T_0|}{T_0}, \tag{29}$$

where $\hat{T}_0$ is the predicted failure time of the in-service unit and T_0 is the true failure time. The results are shown in Fig. 3. The x-axis represents the actual RUL, for example, “80” means only in-service units with actually RUL less or equal

to 80 are considered, and the y-axis is the average failure time prediction error. As can be seen, with IS integrated, the prognostic performance is greatly improved, especially for early-stage in-service units with large RUL.



(a) Dataset 1



(b) Dataset 3

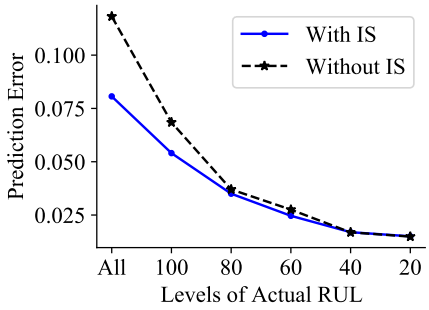
Fig. 3. Failure time prediction error of the in-service units based only on BPR for (a) dataset 1 and (b) dataset 3.

D. Results with Multiple Sensors

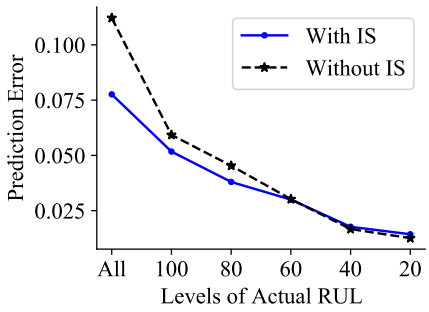
The proposed method is then implemented with the 14 selected sensors in both datasets. The sensor fusion step uses $q = 5$ and $\gamma = 1$ as in (25) for demonstration. In practice, cross validation can be adopted for fine tuning of these parameters [22]. The average prediction errors are shown in Fig. 4. By comparing with the single sensor result in Fig 3, significant improvement can be observed by fusing multiple sensors. It is worth noting that [22] has shown the superior performance of the LLR model without IS over many existing studies. Therefore, the LLR model without IS is used as the baseline for comparison. Again, it can be observed that by integrating IS, the prediction accuracy has been significantly improved, especially for early-stage in-service units with large RULs.

E. Sensitivity to b

A parameter involved in the proposed method is b as in (27). To explore the effect of b, we try different values of b and calculate the failure time prediction errors for the in-service units of dataset 1 based on the 14 sensors, which are shown in Table II. Each row of the table represents a different value of



(a) Dataset 1



(b) Dataset 3

Fig. 4. Failure time prediction error of the in-service units based on 14 sensors for (a) dataset 1 and (b) dataset 3.

b and each column represents the corresponding group of in-service units considered. For example, the "80" column means the group of the in-service units with actual RUL less or equal to 80. As can be seen, a smaller b assigns higher weight to the features $\tilde{\boldsymbol{x}}_{0,j}$ obtained by IS, and may lead to relatively smaller prediction error when the actual RUL is large, but larger prediction error when the actual RUL is small. Overall, the prediction error is not sensitive to the choice of b. Similar conclusions can be drawn for dataset 3 and thus are omitted.

TABLE II
FAILURE TIME PREDICTION ERROR FOR IN-SERVICE UNITS OF DATASET 1 (%)

RUL	All	100	80	60	40	20
$b = 1$	8.2	5.2	3.7	2.8	2.1	1.8
$b = 1.5$	7.8	5.1	3.2	2.4	1.7	1.5
$b = 2$	8.1	5.4	3.5	2.5	1.7	1.5
$b = 3$	8.9	5.7	3.8	2.6	1.7	1.5

V. CONCLUSIONS AND FUTURE WORK

This study proposes a non-parametric degradation modeling approach. The features are extracted from each sensor signal and LLR is adopted to establish the nonlinear relationship between the extracted features and the failure time of the in-service unit. The research gap is that when the in-service unit is in the early degradation stage, the extracted features usually

have high levels of uncertainty and thus are not reliable. To improve the performance of LLR, this study proposes to integrate IS with LLR to incorporate the population-level information into the extracted features by treating each historical unit as a sample from the population and assigning a weight to each historical unit based on its similarity with the in-service unit on the signal path. With two datasets on the degradation of aircraft engines, the proposed method is shown to significantly improve the prognostic performance, especially for early-stage in-service units.

There are several topics worth further investigating. First, an interesting theoretical topic is to quantify the bias of the importance sampling with limited historical units, which may help to determine the value of λ. Second, this study only focuses on one operational condition, and more investigations are desired for non-parametric degradation modeling under multiple or dynamic operational conditions.

REFERENCES

[1] J. Guo, Z. Li, and M. Li, "A review on prognostics methods for engineering systems," *IEEE Trans. Rel.*, vol. 69, no. 3, pp. 1110–1129, 2019.

[2] Y. Hong, M. Zhang, and W. Q. Meeker, "Big data and reliability applications: The complexity dimension," *J. Qual. Technol.*, vol. 50, no. 2, pp. 135–149, 2018.

[3] 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.

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

[5] M. Kim and K. Liu, "A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics," *IISE Trans.*, vol. 53, no. 3, pp. 326–340, 2020.

[6] 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, 2012.

[7] 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. Reliab.*, vol. 62, no. 4, pp. 821–832, 2013.

[8] D. Wang, K. Liu, and X. Zhang, "A generic indirect deep learning approach for multisensor degradation modeling," *IEEE Trans. Autom. Sci. Eng.*, pp. 1–17, 2021, in press.

[9] 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, 1993.

[10] 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, 2006.

[11] Y. Hong, Y. Duan, W. Q. Meeker, D. L. Stanley, and X. Gu, "Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data," *Technometrics*, vol. 57, no. 2, pp. 180–193, 2015.

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

[13] Q. Zhai and Z.-S. Ye, "RUL prediction of deteriorating products using an adaptive Wiener process model," *IEEE Trans. Ind. Inform.*, vol. 13, no. 6, pp. 2911–2921, 2017.

[14] L. Hao, K. Liu, N. Gebraeel, and J. Shi, "Controlling the residual life distribution of parallel unit systems through workload adjustment," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1042–1052, 2015.

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

[16] 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, 2004.

[17] X. Wang and D. Xu, "An inverse Gaussian process model for degradation data," *Technometrics*, vol. 52, no. 2, pp. 188–197, 2010.

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

[19] R. Kontar, S. Zhou, C. Sankavaram, X. Du, and Y. Zhang, "Nonparametric modeling and prognosis of condition monitoring signals using multivariate gaussian convolution processes," *Technometrics*, vol. 60, no. 4, pp. 484–496, 2018.

[20] A. Chehade and K. Liu, "Structural degradation modeling framework for sparse data sets with an application on Alzheimer's disease," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 1, pp. 192–205, 2018.

[21] A. Chehade, C. Song, K. Liu, A. Saxena, and X. Zhang, "A data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes," *J. Qual. Technol.*, vol. 50, no. 2, pp. 150–165, 2018.

[22] C. Song, Z. Zheng, and K. Liu, "Building local models for flexible degradation modeling and prognostics," *IEEE Trans. Autom. Sci. Eng*, pp. 1–13, 2021, in press.

[23] S. T. Tokdar and R. E. Kass, "Importance sampling: a review," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 1, pp. 54–60, 2010.

[24] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.

[25] C. Song, K. Liu, and X. Zhang, "A generic framework for multisensor degradation modeling based on supervised classification and failure surface," *IISE Trans.*, vol. 51, no. 11, pp. 1288–1302, 2019.

[26] M. Kim, C. Song, and K. Liu, "Individualized degradation modeling and prognostics in a heterogeneous group via incorporating intrinsic covariate information," *IEEE Trans. Autom. Sci. Eng.*, 2021.

[27] Q. Zhou, J. Son, S. Zhou, X. Mao, and M. Salman, "Remaining useful life prediction of individual units subject to hard failure," *IIE Trans.*, vol. 46, no. 10, pp. 1017–1030, 2014.

[28] S. Jahani, R. Kontar, S. Zhou, and D. Veeramani, "Remaining useful life prediction based on degradation signals using monotonic b-splines with infinite support," *IISE Trans.*, vol. 52, no. 5, pp. 537–554, 2020.

[29] J. O. Ramsay, "Monotone regression splines in action," *Stat. Sci.*, pp. 425–441, 1988.

[30] C. Loader, *Local regression and likelihood*. Springer Science & Business Media, 2006.

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

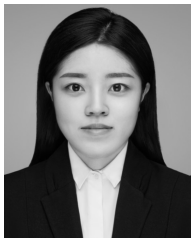
Changyue Song (Member, IEEE) received the B.S. degree in industrial engineering from Tsinghua University, Beijing, China, in 2012, the M.S. degree in management science and engineering from Peking University, Beijing, in 2015, and the M.S. degree in statistics and the Ph.D. degree in industrial engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2018 and 2020, respectively.

Currently, he is an Assistant Professor with the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, USA. His research interests include data analytics for system improvement, process modeling and prognosis, and statistical learning.

Andi Wang received the B.S. degree in statistics from Peking University, Beijing, China, in 2012, the Ph.D. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2016, and the Ph.D. degree from the School of Industrial and Systems Engineering, Georgia Tech, Atlanta, GA, USA.

His research interests include advanced statistical modeling, large-scale optimization, and machine learning for system performance improvement.

Dr. Wang is also a member of the Institute of Industrial and Systems Engineers (IISE) and the Institute for Operations Research and the Management Sciences (INFORMS).



Di Wang received the B.S. degree in industrial engineering from Nankai University, Tianjin, China, in 2015, and the Ph.D. degree in management science and engineering from Peking University, Beijing, China, in 2020.

She is currently an Assistant Professor with the Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. Her research interests include statistically modeling of spatiotemporal data and artificial intelligence of complex engineering systems.

Dr. Wang is also a member of INFORMS and IISE.

a member of IEEE, INFORMS, and IISE