Remaining Useful Life Prediction of IIoT-Enabled Complex Industrial Systems with Hybrid Fusion of Multiple Information Sources

Wen, Pengfei; Li, Yong; Chen, Shaowei; Zhao, Shuai

Published in:
IEEE Internet of Things Journal

DOI (link to publication from Publisher):
10.1109/JIOT.2021.3055977

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Remaining Useful Life Prediction of IIoT-enabled Complex Industrial Systems with Hybrid Fusion of Multiple Information Sources

Pengfei Wen, *Student Member, IEEE*, Yong Li, Shaowei Chen, *Member, IEEE*, and Shuai Zhao, *Member, IEEE*

***Abstract*—Industrial Internet of Things has significantly boosted predictive maintenance for complex industrial systems, where the accurate prediction of remaining useful life with high-level confidence is challenging. By aggregating multiple informative sources of system degradation, information fusion can be applied to improve the prediction accuracy and reduce the uncertainty. It can be performed on the data-level, feature-level, and decision-level. To fully exploit the available degradation information, this paper proposes a hybrid fusion method on both the data level and decision level to predict the remaining useful life. On the data level, Genetic Programming is adopted to integrate physical sensor sources into a composite health indicator, resulting in an explicit nonlinear data-level fusion model. Subsequently, the predictions of the remaining useful life based on each physical sensor and the developed composite health indicator are synthesized in the framework of belief functions theory, as the decision-level fusion method. Moreover, the decision-level method is flexible for incorporating other statistical data-driven methods with explicit estimations of the remaining useful life. The proposed method is verified via a case study on NASA's C-MAPSS data set. Compared to the single-level fusion methods, the results confirm the superiority of the proposed method for higher accuracy and certainty of predicting the remaining useful life.**

***Index Terms*—Industrial Internet of Things, Information Fusion, Multiple Sources, Prognostics, Remaining Useful Life.**

I. Introduction

CATASTROPHIC failure of critical industrial systems will bring huge economic losses. In order to prevent potential risks that may lead to such failure, Condition-Based Maintenance (CBM) and Predictive Maintenance (PdM) have been developed to identify or predict latent problems using Condition Monitoring (CM) information [1]. Industrial Internet of Things (IIoT) provides an information-rich era since big data can be transmitted in IIoT, also boosting PdM for industrial systems [2]. Currently, the focuses of PdM in IIoT are mainly on the development of hardware and software for tracking the health state of components of monitored systems [3].

Pengfei Wen, Yong Li, and Shaowei Chen are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China. (Email: wenpengfei@mail.nwpu.edu.cn; ruikel@nwpu.edu.cn; cgong@nwpu.edu.cn)

Shuai Zhao is with the Department of Energy Technology, Aalborg University, Aalborg 9220, Denmark. (Email: szh@et.aau.dk)(*Corresponding author: Shuai Zhao*)

Accurate Remaining Useful Life (RUL) prediction is one of the central tasks of PdM, which can effectively help to reduce economic cost of maintenance. RUL estimation for complex industrial systems is still challenging due to the complicated mechanisms. The developed sensing techniques and IIoT create a premise for these issues, where multiple sensors are simultaneously employed to collect CM signals, serving as various degradation features or Health Indicators (HIs). These signals can be *in-situ* and on-chip processed in equipment plane, or be transmitted by the sensor networks to a cloud or edge computation center in IIoT.

However, for complex deteriorating systems, a single sensor is incapable of providing sufficient information of the latent degradation processes. As a result, it is desirable to fuse information from various sensors, with the expectation of compensating for the limitations of each other. Information fusion from various sources can be realized on the data level, feature level, and decision level [4]. Data produced by data-level fusion techniques usually possess similar form and structure to the raw data, so data-level fusion can be readily integrated with the existing prognostic techniques to further improve the performance. Besides, data-level fusion can facilitate data visualization and enables continuous characterization of health state [10]. Based on it, practitioners will have an insight into the latent health state of complex deteriorating systems. Liu et al. conducted a series of research works about data-level fusion for RUL prognostics of aero-engines. They built fusion models by optimizing the fusion result in terms of several desired properties, such as monotonicity and variance of failure threshold [5], fitting error using their proposed degradation model [6], a designed Signal-to-Noise Ratio (SNR), [7] and the error between the actual RUL and the predicted one in both linear [8] and the nonlinear ways [9]. Due to the limited representation capability of the linear models, it may not handle the multi-source information from complex systems produced by complicated mechanisms. Song et al. [10] further combined a proposed property in [7] and the Kernel Method (KM) to realize nonlinear fusion, in which eigenvalue decomposition is needed to be implemented for the kernel matrix produced by training data, consuming a long time and huge memory space. Furthermore, fusion models built by using KM are usually implicit due to the implicitly defined nonlinear transformation by the inner product. Li et al. [12] chose CM signals from four sensors as the Physical HIs (PHIs). Then the information was fused as the Euclidean distance between each multi-dimensional measurement and a

Table I
LITERATURE REVIEW IN TERMS OF THE LEVEL THAT FUSION IS PERFORMED

Category	Methods	Advantages	Limitations
Data-level	Optimization-based linear combination [5], [6], [7], [8], [9]	(1) Introduce less manual intervention; (2) Facilitate data visualization [10]; (3) Enable continuous characterization of health state [10];	(1) Performance is closely related to the quality of data; (2) Commonly expensive computation [11]; (3) Raw data are usually in quite different scales;
	Optimization-based KM [10]		
	Distance-based [12], [13], [14]		
	CNN [15]*, [16]		
	LSTM [15]*		
	GPR [17]*		
Feature-level	KPCA [18], [19]	(1) Degradation characteristics can be enhanced by using independent feature-level analysis techniques; (2) Massive data are reduced into a manageable amount and structure [11];	(1) Highly dependent on the quality of the raw data [5]; (2) Highly related to the used feature extraction approaches [5]; (3) Features usually have different ranges [19];
	GRU-RNN [20]*, [19]*		
	Bi-directional LSTM [21]*		
	Random forest [22]*		
	GP [23], [24], [25]		
Decision-level	Linear combination [26], [17], [27]*, [22]*	(1) Flexible to aggregate the estimation from independent prognostic techniques [5]; (2) Reduce the variance of the estimation [28]; (3) Enhance the confidence of the decision [28];	(1) Highly related to the constructed decisions [5]; (2) Require detailed prior knowledge of combining decisions [6]; (3) Cannot provide insights on the correlation between data and health state [7]
	Correlation [29]*		
	Dempster's rule [17]*		
	Bayesian network [13]		
	AdaBoost [30]		
	Stacking [30]		

* : hybrid information fusion on more than a single level

predefined Failure Threshold (FT), i.e., the predicted last value from a regression model for each CM signal at the End of Life (EoL) of each degraded unit. This category of distance-based methods fuses the information in a nonlinear way without optimizing desired properties for prediction, and the pattern of fusion is also fixed. Deep Neural Networks (DNN) have also been one of the most popular nonlinear models since they are capable to deal with the latent relations between information sources of complex systems. Yu et al. [31] used a bi-directional Recurrent Neural Network (RNN) to fuse the sensing data by minimizing the reconstruction error. However, the fusion results of DNNs lack interpretability due to their black-box feature, limiting their applications where the closed-form estimation of uncertainty is required. Zhao et al. [32] fused drain-to-source on-state resistance and the gate-to-source threshold voltage of SiC MOSFETs into a failure precursor in a nonlinear way by using Genetic Programming (GP), in which the candidate models are explicit and represented by tree structures in GP to be optimized.

The main limitations of data-level fusion include that 1) its performance is closely related to the quality of raw data; 2) the computation is expensive for health management in the case of big data [11]; 3) data produced by different sources are usually within significantly diverse scales. As a result, data-level techniques which are sensitive to the numerical scale need to be combined with normalization or standardization methods. To handle the issues of the unavailability of raw data, feature extraction techniques have been developed to discover advanced features to characterize the latent deterioration. In most cases, the extracted features also possess a similar form with raw data, facilitating the practice of the developed data-level fusion techniques on the feature level. As a result, information fusion on feature level is expected to be flexible to aggregate the information from independent feature-level analysis techniques [5]. With feature-level information fusion, massive data are reduced to be manageable [11], since another important application of feature extraction is to reduce the dimension of raw data. However, the quality of raw data still poses an inevitable impact on the performance of the fusion, and intuitively, it will be highly related to the used feature extraction approaches [5]. Moreover, the problems caused by the diverse numerical scales still exist in the extracted features. Generally, most methodologies of data-level fusion can be implemented on the feature level. Liu et al. [11] extracted 6 features from the degradation data of cutting tools and then fused them into a 1-dimensional Health Indicator (HI) by using Principal Component Analysis (PCA), which can represent tool wear conditions. KM can also be applied to transform linear fusion models produced by PCA into nonlinear ones, such that Kernel PCA (KPCA) was formed to fuse the extracted time domain, frequency domain, time-frequency domain features [19], and cumulative features [18]. Liao [23] first introduced GP into information fusion, to automatically discover advanced composite features based on the extracted ones, and another application to fuse extracted features by using GP was provided by Qin et al. [24]. While performing a feature-level fusion, Wang et al. [25] further improved GP by improving the operator of the crossover in GP, such that

crossover was only carried out on a specific class of nodes to preserve several key features.

For the sake of utilizing and aggregating information from multiple sources in a more flexible way [5], decision-level fusion has also been introduced into health monitoring. One of the advantages of decision-level fusion is its capability to reduce the variance of the estimation and enhance the confidence of the decision [28]. However, the performance of decision-level fusion is related to the constructed decisions [5]. Furthermore, it usually requires detailed prior knowledge of combining decisions [6] and cannot provide insights into the correlation between collected monitoring data and the latent health state [7]. Ensemble learning is a typical practice of decision-level information fusion. Li et al. [22] used a random-forest-based regression method to estimate RUL, in which a group of CART trees is produced. Then the estimation from each CART tree was aggregated as the global estimation. Ma et al. [30] extracted several features and on each of them, a Support-Vector-Regression (SVR) sub-model was built for prognostics. An ensemble of prediction of all the sub-models was conducted by using AdaBoost and Stacking. These ensemble-learning-based decision-level fusion methods commonly focus on the fusion of point estimators of RUL for highly accurate point estimation, in which fusion process the uncertainty management is not investigated. Baraldi et al. [17] aggregated the RUL independently predicted by using two disparate models, a similarity-based model and a Gaussian Process Regression (GPR) model, where the point estimation and the uncertainty estimation were handled separately. The point estimations were fused by taking the mean of them, while the uncertainty estimation was combined under the Dempster's rule in the framework of Belief Functions Theory (BFT), where probability needs to be reassigned as the belief since BFT is more general than probability theory [33], challenging the combination of it with many other developed probability-based methods.

Because information fusion schemes on different levels possess their advantages and limitations, hybrid fusion on more than one level is favorable. The benefits expected from developing the structure of hybrid fusion are based on the complementary merits of the fusion techniques which can mitigate the limitations of fusion on one level by the advantages of another level, while the systematic research on the structure of hybrid fusion needs more investigations. Chen et al. [20] proposed a framework of data-feature-level fusion, in which raw data were fused into several composite features by using KPCA, and then the extracted features were fused by a Gated-Recurrent-Unit Recurrent Neural Network (GRU-RNN) and mapped to the output RUL. Al-Dulaimi et al. [15] proposed a Hybrid Deep Neural Network (HDNN) whose former layers consist of a Convolutional Neural Network (CNN) and a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN), and its latter layers form a Multi-Layer Feed Forward Neural Network (MLFNN). In the proposed HDNN, it can be regarded that information was fused on data level by using the CNN and the RNN respectively, in which process abstract features were output. Then the output features were deeply combined in the latter full-connection layers, such that a data-feature-level hybrid fusion was implemented. As a result, the information fusion techniques on different levels and their corresponding limitations and advantages are summarized in Table I.

To provide a new way to fully exploit the multi-source information in IIoT, a framework of dual-level hybrid information fusion for RUL prognostics is built. Considering the representative ability of nonlinear data-level fusion models, the preprocessed sensing data are fused into a composite HI by using GP. Compared with KM-based and DNN-based data-fusion models, GP-based fusion models consume less training time and computer memories, and the produced fusion models are explicit and computationally efficient, significantly facilitating its deployment to the monitored system in equipment plane or cloud and edge computation centers. Moreover, advantages and limitations of this GP-based fusion model have been discussed in [34]. Then the composite HI is regarded as another information source like the physical sensors, and the Probability Density Function (PDF) of RUL is provided independently based on each source, including the fused HI. The PDF acquired from each source is further aggregated in the framework of BFT, considering that in the existing BFT-based decision-level fusion methods, the fusion of point estimators lacks a quantitative interpretation under a certain theoretical framework. Besides, in this paper, the derived decision fusion model can also be interpreted based on Bayesian inference in the framework of probability theory, which also facilitates the combination of the estimated representation of uncertainty and other probability-theory-based methods. The proposed framework is validated on a case study of aircraft engines, and the contributions of this study are: 1) an easy-to-deploy nonlinear data-level fusion model is provided to improve the prognostic performance compared with any single CM sensor in IIoT environment; 2) a lightweight decision-level fusion model is applied to improve the prognostic accuracy and significantly reduce the uncertainty; 3) a dual-level framework of hybrid information fusion is developed, providing a practice of information fusion for complex multi-source information systems. Furthermore, the contributions 1) and 2) are convenient to be transferred, facilitating the practice for other practitioners.

The rest of this paper is organized as follows. In Section II, the degradation modeling, parameter estimation, and the RUL prediction methods are introduced, with which methods RUL can be predicted independently based on a single information source. Section III describes a GP-based data-level fusion method, which produces an extra information source. Section IV derives a rule of combination in the framework of BFT, which produces the formalism of decision-level fusion. Section V summarizes the proposed IIoT-enabled framework of hybrid information fusion for practical RUL prognostics. Section VI verifies the proposed dual-level hybrid fusion on the data set of aircraft engines and compares the results with the existing benchmark methods. Section VII summarizes the paper with a conclusion.

II. Estimation of Remaining Useful Life Based on Single Information Source

Degradation data can be obtained from the CM sensors over time, containing useful information about the latent degradation process. In this section, a mixed-effect degradation path model is adopted to quantify the degradation process using the collected data from a single sensor. The model is then combined with a widely-used definition of RUL, to estimate its conditional Cumulative Distribution Function (CDF). The prognostic method generally contains two steps: model development, and RUL estimation.

A. Degradation Modeling

For the step of model development, degradation path models are built for both the historical run-to-failure systems and the *in-situ* ones. The methods of estimating the corresponding model parameters are distinct, i.e., the Weighted Least Squares (WLS) method, and a reconstruction method, respectively. In this way, information included in the run-to-failure systems, as well as partially available CM data, are integrated into this framework to predict the RUL of the *in-situ* operating and monitored systems.

1) Degradation Path Model with Historical Data: For any single CM signal, Lu and Meeker [35] proposed a general degradation path model:

$$\begin{aligned} y_{i,k} &= y_i\left(t_k\right), \\ &= \psi\left(t_k; \boldsymbol{\Phi}, \boldsymbol{\theta}_i\right) + \varepsilon_{i,k}, \quad (1) \\ &= \psi_{i,k} + \varepsilon_{i,k}, \quad (2) \\ i &= 1, 2, \ldots, M, \\ k &= 1, 2, \ldots, N_i, \end{aligned}$$

where $y_{i,k}$ denotes the k^{th} CM measurement of the monitored system i at time t_k; under the same failure mode and operation condition, $\boldsymbol{\Phi}$ is the vector of fixed-effect parameters while $\boldsymbol{\theta}_i$ is the vector of non-fixed-effect parameters, characterizing homogeneity and heterogeneity between different monitored systems, respectively; $\varepsilon_{i,k}$ is the measurement error; $\psi_{i,k}$ represents the actual degradation path at t_k; M is the number of the monitored systems and N_i is the number of collected measurements of the monitored system i.

Similar to [34], here the degradation path model is applied to the logarithm scale of the original CM measurements and the monitoring time:

$$\begin{aligned} y_{i,k} &= \ln\left(s_{i,k} - \phi\right) \\ &= \theta_i^{(0)} + \theta_i^{(1)} \ln t_k + \varepsilon_{i,k}, \end{aligned} \quad (3)$$

where $s_{i,k}$ represents the original CM measurement; ϕ is the initial degradation level, which is common for all monitored systems, i.e., the fixed-effect parameter; the degradation path $\psi_{i,k} = \theta_i^{(0)} + \theta_i^{(1)} \ln t_k$; $\boldsymbol{\theta}_i = \begin{bmatrix} \theta_i^{(0)} & \theta_i^{(1)} \end{bmatrix}^{\mathrm{T}}$ represents non-fixed-effect parameters, which are specified for each system; $\varepsilon_{i,k} \sim N\left(0, \sigma^2/c_{i,k}\right)$ with the variance $\sigma^2/c_{i,k}$ and $c_{i,k}$ is the weight coefficient of the corresponding measurement error, tuning the impact imposed by the short-term data and long-term data. Here, $c_{i,k} \geq 0$ and $\sum_{k=1}^{N_i} \sqrt{c_{i,k}} = 1$. The detailed approaches to set $c_{i,k}$ can be found in [34], [5]. Besides, the non-random part of (3) is defined as the degradation trajectory:

$$\eta_{i,k} = \eta_i\left(t_k\right) = \phi + \exp\left(\theta_i^{(0)} + \theta_i^{(1)} \ln t_k\right), \quad (4)$$

and then $\eta_{i,k} = \phi + \exp\left(\psi_{i,k}\right)$.

2) Model Parameter Estimation: Eq. (3) is rewritten as the matrix form:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \quad (5)$$

where $\boldsymbol{y}_i = \begin{bmatrix} y_{i,1} & y_{i,2} & \cdots & y_{i,N_i} \end{bmatrix}^{\mathrm{T}} \in \mathrm{R}^{N_i \times 1}$; $\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i,1} & \varepsilon_{i,2} & \cdots & \varepsilon_{i,N_i} \end{bmatrix}^{\mathrm{T}} \in \mathrm{R}^{N_i \times 1}$; $\boldsymbol{X}_i$ is the design matrix:

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{N_i} \end{bmatrix} = \begin{bmatrix} 1 & \ln t_1 \\ 1 & \ln t_2 \\ \vdots & \vdots \\ 1 & \ln t_{N_i} \end{bmatrix} \in \mathrm{R}^{N_i \times 2}.$$

Then WLS is adopted here to estimated the degradation non-fixed-effect parameters of historical monitored systems, by solving the normal equation [34]:

$$\hat{\boldsymbol{\theta}}_i = \left(\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{C}_i \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{C}_i \boldsymbol{y}_i, \quad (6)$$

where $\boldsymbol{C}_i = \mathrm{diag}\left(c_{i,1}, c_{i,2}, \cdots, c_{i,N_i}\right)$.

B. Prediction of Remaining Useful Life for In-situ Monitored Systems

RUL is defined as the time left before several degradation features, such as measurements, paths, or trajectories, exceed a pre-specified Failure Threshold (FT) [36]. In this case, failures are regarded as soft failures [37]. To take full advantage of the degradation information provided by the historical run-to-failure systems, the historical database is also involved in degradation modeling for the *in-situ* monitored systems. Specifically, the degradation path and the FT of an *in-situ* system are reconstructed by a weighted sum of those of all the historical run-to-failure systems, where the weight coefficients w between the *in-situ* system and each historical run-to-failure system is calculated by minimizing the reconstruction error [38].

1) Data-Driven Parameter Reconstruction: The degradation path model of an *in-situ* system can be constructed as:

$$y_{*,k} = \theta_*^{(0)} + \theta_*^{(1)} x_{*,k} + \varepsilon_{*,k} = \psi_{*,k} + \varepsilon_{*,k}, \quad (7)$$

where $y_{*,k}$ is the collected *in-situ* measurement at t_k^*, and $x_k^* = \ln t_k^*$. The model reconstruction is realized by reconstructing the degradation path:

$$\psi_{*,k} = \sum_{i=1}^{M} w_i \psi_{i,k}, \quad (8)$$

where $\psi_{*,k}$ is the degradation path of the *in-situ* system at t_k; w_i is the reconstruction coefficient between the *in-situ* system and the historical run-to-failure system i, and (7) can be rewritten as:

$$\boldsymbol{y}_* = \boldsymbol{X}_* \boldsymbol{\theta}_* + \boldsymbol{\varepsilon}_* = \boldsymbol{X}_* \boldsymbol{\Theta} \boldsymbol{w} + \boldsymbol{\varepsilon}_*, \quad (9)$$

where

$$\boldsymbol{\theta}_* = \boldsymbol{\Theta}\boldsymbol{w}; \tag{10}$$

$\boldsymbol{y}_* = \left[\begin{array}{cccc} y_{*,1} & y_{*,2} & \cdots & y_{*,N_*} \end{array}\right]^{\mathrm{T}} \in \mathrm{R}^{N_*\times 1}$; N_* is the number of collected CM measurements of the *in-situ* monitored system; $\boldsymbol{w} = \left[\begin{array}{cccc} w_1 & w_2 & \cdots & w_M \end{array}\right]^{\mathrm{T}} \in \mathrm{R}^{M\times 1}$; $\boldsymbol{\varepsilon}_* = \left[\begin{array}{cccc} \varepsilon_{*,1} & \varepsilon_{*,2} & \cdots & \varepsilon_{*,N_*} \end{array}\right]^{\mathrm{T}} \in \mathrm{R}^{N_*\times 1}$; and

$$\boldsymbol{\Theta} = \left[\begin{array}{cccc} \boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \cdots & \boldsymbol{\theta}_M \end{array}\right] = \left[\begin{array}{cccc} \theta_1^{(0)} & \theta_2^{(0)} & \cdots & \theta_M^{(0)} \\ \theta_1^{(1)} & \theta_2^{(1)} & \cdots & \theta_M^{(1)} \end{array}\right].$$

$\boldsymbol{w}$ can be calculated by minimizing the reconstruction error [38]. To estimate degradation parameters of historical systems, logarithmic transformation is applied to the measurements and the monitoring time in (3), and consequently trend information of the degradation paths and the specific failure time are blurred due to the logarithmic transformation. As a result, for better reconstruction, the reconstruction error is calculated as the fitting residual between the reconstructed degradation trajectory $\eta_{*,k} = \phi + \exp\left(\psi_{*,k}\right) = \phi + \exp\left(\sum_{i=1}^{M} w_i \psi_{i,k}\right)$, and the collected measurements $s_{*,k}$ of the *in-situ* system, where $k = 1, 2, \ldots, N_*$. Accordingly, $\boldsymbol{w}$ can be determined as:

$$\begin{aligned} \boldsymbol{w} &= \arg\min \sum_{k=1}^{N_*} \left(s_{*,k} - \eta_{*,k}\right)^2 \\ &= \arg\min \sum_{k=1}^{N_*} \Bigg[s_{*,k} - \phi \\ &\qquad - \exp\left(\sum_{i=1}^{M} w_i \left(\theta_i^{(0)} + \theta_i^{(1)} x_{*,k}\right)\right)\Bigg]^2, \\ &\quad s.t.\ \ w_i \geq 0, \sum_{i=1}^{M} w_i = 1. \end{aligned} \tag{11}$$

2) Estimation of Remaining Useful Life: Let D denote the FT. In the historical database, since each monitored system is run-to-failure, their corresponding FT is assumed to follow a Gaussian distribution with the same variance:

$$D_i \sim N\left(\eta_{i,N_i}, \delta^2\right),$$

where η_{i,N_i} is the value of degradation trajectory of the historical run-to-failure system i at its EoL, i.e., the last value of the trajectory; δ^2 is estimated as:

$$\hat{\delta}^2 = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{N_i - 2} \sum_{k=1}^{N_i} \left(s_{i,k} - \eta_{i,k}\right)^2. \tag{12}$$

In addition to reconstructing the degradation trajectory of the *in-situ* monitored system, the reconstruction coefficients $\{w_i | i = 1, 2, \cdots, M\}$ can also be used to reconstruct its FT [38]. Based on the run-to-failure systems in the historical database, FT D_* of the *in-situ* monitored system is assumed to follow such a Gaussian distribution:

$$D_* \sim N\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\eta}^{thd}, \delta^2 \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}\right),$$

where $\boldsymbol{\eta}^{thd} = \left[\begin{array}{cccc} \eta_{1,N_1} & \eta_{2,N_2} & \cdots & \eta_{M,N_M} \end{array}\right]^{\mathrm{T}} \in \mathrm{R}^{M\times 1}$.

Since in degradation modeling, logarithmic transformation is also applied to the CM time, i.e., $x = \ln t$, to obtain the precise time when several degradation feature(s) exceed the FT, RUL is estimated as the time left before the degradation trajectory η_* of the *in-situ* system exceed D_*, i.e.,

$$T^* = \inf\left(t : \eta_*\left(t_{N_*} + t\right) \geq D_*\right),$$

where t_{N_*} is the current CM time. Conditioning that the RUL of the *in-situ* system is greater or equal to 0 at the current monitoring time, the conditional CDF of the RUL can be derived as [34]:

$$F_{*|RUL\geq 0}\left(t\right) = \frac{\Phi\left(g\left(\eta_*\left(t_{N_*} + t\right)\right)\right) - \Phi\left(g\left(\eta_*\left(t_{N_*}\right)\right)\right)}{1 - \Phi\left(g\left(\eta_*\left(t_{N_*}\right)\right)\right)}, \tag{13}$$

where $\Phi\left(\cdot\right)$ is the CDF of standard normal distribution and

$$g\left(\cdot\right) = \frac{\cdot - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\eta}^{thd}}{\delta\sqrt{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}}. \tag{14}$$

Based on $F_{*|RUL\geq 0}\left(t\right)$, different statistics can be used as the estimators of RUL, such as mean, mode, median and other quantiles. Besides, the Confidence Interval (CI) of the estimated RUL can be obtained by truncating the conditional CDF $F_{*|RUL\geq 0}\left(t\right)$. More details of the model formulation and derivations of RUL can be found in [34].

III. Data-Level Fusion

In the integrated framework of dual-level information fusion, GP is applied to fuse information on the data level, which is one of heuristic search techniques. It usually starts from a random initial population containing unfit individuals (or called programs), fitting for a predefined optimization task, in which the objective serves as the fitness function, by applying operators analogous to natural genetic processes, including:

1) Selection: a series of individuals are selected with a certain probability from the current generation to serve as parents for the next generation, such that individuals with higher fitness are more likely to be selected.

2) Crossover: random parts of the selected parents are swapped to produce new and different offspring, becoming individuals of the next generation.

3) Mutation: random parts of several individuals are modified and they consequently change into new ones.

Typically, individuals in a new generation are statistically better than those in a previous generation, and the evolution is terminated when several individuals fit the task at a pre-specified level, or when the iteration reaches a pre-specified number. Individuals are traditionally represented by tree structures, as shown in Fig. 1. Each intermediate node corresponds to a mathematical operator and each terminal node stores an operand, making objective (fitness) functions easy to evolve and evaluate. Accordingly, in the GP-based data-level fusion model, the tree structure in Fig. 1 represents

$$\begin{aligned} h_{i,k} &= \mathcal{F}\left(s_{i,\cdot,k}\right), \\ &= \ln\left|\left(s_{i,1,k} + s_{i,2,k}\right)\cos\left(s_{i,3,k} \cdot s_{i,4,k}\right)\right|, \end{aligned}$$

where $\mathcal{F}\left(\cdot\right)$ is a row-wise fusion operator fusing each multi-dimensional CM measurement into the composite HI; $h_{i,k}$ represents the HI at time t_k of the monitored system i, and $s_{\cdot,j,\cdot}$ denotes the data from the information source j. More details of GP can be found in [32].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3055977, IEEE Internet of Things Journal

$$
\begin{aligned}
f_{*,\cap(1:J)}\left([x,y]\right) &= \left(f_{*,1} \oplus f_{*,2} \oplus \cdots \oplus f_{*,J}\right)\left([x,y]\right) \\
&= \begin{cases} \frac{1}{1-K} \int\int_{x_1,y_1} \cdots \int\int_{x_J,y_J} \prod_{j=1}^{J} f_{*,j}\left([x_j,y_j]\right) \mathrm{d}x_1 \mathrm{d}y_1 \cdots \mathrm{d}x_J \mathrm{d}x_J, & [x,y] = \bigcap_{j=1}^{J} [x_j,y_j], \\ 0, & \text{otherwise}, \end{cases} \\
&= \begin{cases} \left(\frac{1}{\sqrt{2}}\right)^J \frac{\prod_{j=1}^{J} P_{*,j|RUL\geq 0}\left(\frac{x+y}{2}\right)}{1-K}, & x = y, \\ 0, & \text{otherwise}, \end{cases}
\end{aligned} \tag{15}
$$

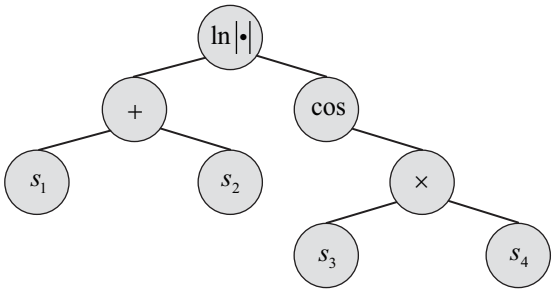

Figure 1. A data fusion function represented as a tree structure.

IV. Decision-Level Fusion

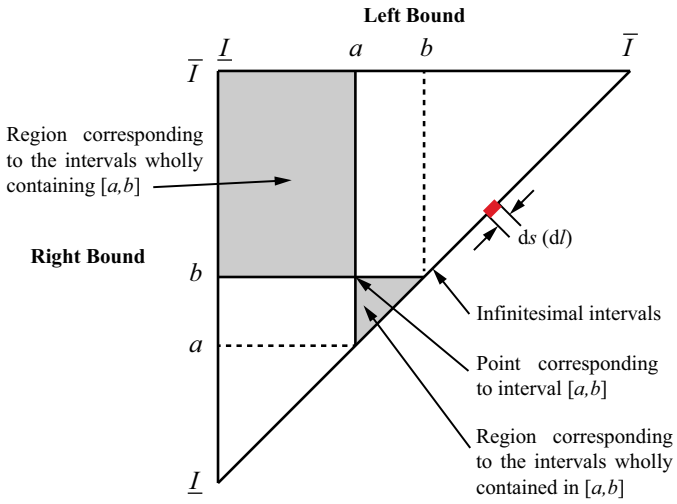

Figure 2. The continuous frame of discernment.

The combination of the predicted RUL obtained from different sources is the central idea in decision-level fusion, and different rules of combination have been proposed and applied by taking the mean [22], the median [27], the trimmed mean [27], or more flexible schemes such as the AdaBoost and Stacking [30]. A framework of combining both point estimation and uncertainty estimation is expected to significantly improve the uncertainty management.

A. Belief Functions Theory

In the framework of Belief Functions Theory (BFT), when an unknown variable is estimated based on independent sources, its subjective probabilities represent the degrees of belief in terms of the sources, and Dempster's rule can be used to combine the probabilities and provide a global estimation.

To generalize the frame of discernment of BFT to real-number cases, it is assumed that masses can only be given to a series of intervals along the number line. In the generalized frame, the belief functions are defined on an infinite frame of discernment, and the focal elements are the closed intervals $[a,b]$ in $\mathcal{R}$ [39]. Then Basic Belief Assignment (BBA) $m\left([a,b]\right)$ is generalized as Basic Belief Density (BBD) $f\left([a,b]\right)$ for $a \leq b$ and $f\left([a,b]\right) = 0$ when $a > b$, and the sum operator in the finite and discrete frame of discernment are consequently converted into integrals. Another constraint was introduced in [40] that the frame of discernment only contains contiguous intervals, i.e., only a collection of intervals distributing end to end on the number line can constitute a union. As a result, any focal element $[a,b] \subseteq \left[\underline{I},\overline{I}\right]$ can be represented by the location of a point in a triangular area in Cartesian coordinates as shown in Fig. 2, where the x-axis represents the lower bound while the y-axis specifies the upper bound of the intervals. Accordingly, the area in the lower right side of the point (a,b) represents the set consisting of the intervals contained in $[a,b]$, while the area in the upper left side of the point (a,b) represents the set consisting of the intervals containing $[a,b]$. Besides, the real numbers, which can also be regarded as infinitesimal intervals along the number line, are represented by the points on the hypotenuse of the triangle.

B. Decision Fusion for Estimating Remaining Useful Life

For an *in-situ* operating monitored system, based on the signal from source j, CDF of its RUL can be independently derived as $F_{*,j|RUL\geq 0}\left(t\right)$ according to (13), and the Probability Density Function (PDF) can be obtained as $P_{*,j|RUL\geq 0}\left(t\right) = \mathrm{d}F_{*,j|RUL\geq 0}\left(t\right)/\mathrm{d}t$, and consequently, they can be combined in the frame of BFT.

Based on the PDF of the estimated RUL, different BBD can be induced accordingly. Here the BBD is induced as:

$$
f_*\left([x,y]\right) = \begin{cases} \frac{1}{K_1} P_{*|RUL\geq 0}\left(\frac{x+y}{2}\right), & x = y,\ x,y \geq 0, \\ 0, & \text{otherwise}, \end{cases}
$$

where K_1 is a normalization constant that ensures

$$
\frac{1}{K_1} \int_L f_*\left([x,y]\right) \mathrm{d}s = 1; \tag{16}
$$

$L : x = y,\ x,y \geq 0$ is on the hypotenuse in Fig. 2, which is equal to $L : x = l,\ y = l,\ l \geq 0$; $\mathrm{d}s$ represents an elementary

arc length along L, and thus $\mathrm{d}s = \sqrt{2}\mathrm{d}l$. It can be seen that BBD is taken as the PDF multiplied by a constant factor:

$$\begin{aligned}\frac{1}{K_1}\int\int_{x,y} f_*\left([x,y]\right)\mathrm{d}x\mathrm{d}y &= \frac{1}{K_1}\int_L P_{*|RUL\geq 0}\left(\frac{x+y}{2}\right)\mathrm{d}s \\ &= \frac{\sqrt{2}}{K_1}\int_0^{+\infty} P_{*|RUL\geq 0}\left(l\right)\mathrm{d}l \\ &= \frac{\sqrt{2}}{K_1}. \end{aligned} \tag{17}$$

As a result, $K_1 = \sqrt{2}$ due to the property of normalization of BBD. According to Dempster's rule of combination, the focal elements of the combined BBD are the intersections of those of each source, which are also a series of infinitesimal intervals along the number line, and accordingly, J BBDs acquired from J sources can be combined as (15), where $\oplus$ represents the combination operator; $(1-K)$ is also a normalization constant that ensures $\int_L f_{*,\cap(1:J)}\left([x,y]\right)\mathrm{d}s/\left(1-K\right) = 1$, and:

$$\begin{aligned} K &= 1 - \int_L \left(\frac{1}{\sqrt{2}}\right)^J \prod_{j=1}^{J} P_{*,j|RUL\geq 0}\left(\frac{x+y}{2}\right)\mathrm{d}s, \\ &= 1 - \int_0^{+\infty}\left(\frac{1}{\sqrt{2}}\right)^{J-1}\prod_{j=1}^{J} P_{*,j|RUL\geq 0}\left(l\right)\mathrm{d}l; \end{aligned} \tag{18}$$

Then the combined PDF $P_{*,\cap(1:J)|RUL\geq 0}\left(t\right)$ can be derived according to the corresponding BBD $f_{*,\cap(1:J)}\left([t,t]\right)$ multiplied by the normalization factor:

$$\begin{aligned} P_{*,\cap(1:J)|RUL\geq 0}\left(t\right) &= K_1 f_{*,\cap(1:J)}\left([x,y]\right)|_{x=y=t,t\geq 0}, \\ &= \left(\frac{1}{\sqrt{2}}\right)^{J-1}\frac{\prod_{j=1}^{J} P_{*,j|RUL\geq 0}\left(t\right)}{1-K}, \\ &= \frac{\prod_{j=1}^{J} P_{*,j|RUL\geq 0}\left(t\right)}{\int_0^{+\infty}\prod_{j=1}^{J} P_{*,j|RUL\geq 0}\left(l\right)\mathrm{d}l}. \end{aligned} \tag{19}$$

Eq. (19) is identical to that derived by using Bayesian inference in the frame of probability theory since probability theory is one of the particular forms of BFT [33]. Fused CDF of RUL can also be accordingly acquired as $F_{*,\cap(1:J)|RUL\geq 0}\left(t\right) = \int_0^t P_{*,\cap(1:J)|RUL\geq 0}\left(\tau\right)\mathrm{d}\tau$, and different statistics can be used to estimate the RUL based on the combined CDF, which is the same as that with any single information source.

V. Proposed Framework of Hybrid Information Fusion for Predictive Maintenance in an Industrial Internet of Things Environment

A typical implementation of the proposed framework embedded in IIoT is shown in Fig. 3, which can be generally divided in to three planes, i.e., cloud plane, edge plane, and equipment plane [2]. By using multiple sensors, several channels of CM information are simultaneously collected, and then the data-level fusion can be achieved on an embedded processor in equipment plane by using a preset or a trained GP-based fusion model to provide a composite HI. CM signals as well as HI are transmitted through a sensor network. With CM data and HI, 1) GP-based fusion model and the parameter reconstruction model can be trained and updated on the cloud computation center; 2) RUL can be predicted on the edge computation devices by using the constructed parameter reconstruction prognostics model, where the reconstruction coefficients are provided by the cloud center. Finally, the decision-level fusion can be implemented on the edge devices for man-in-loop decision or on the *in-situ* embedded processor for real-time adaptive health management.

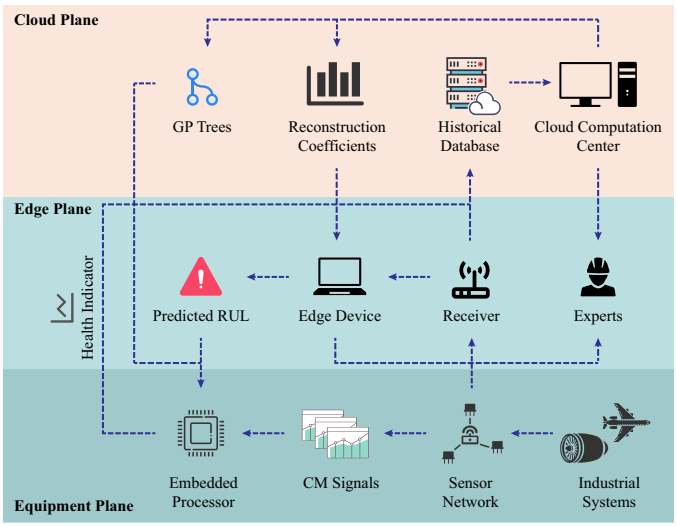

Figure 3. The proposed framework of the hybrid information fusion in an IIoT environment to predict RUL of industrial systems.

The information flow of the proposed framework is shown in Fig. 4. As can be seen, sensing data from J sensors are aggregated on data level to form a composite HI, and the estimated RULs based on all the $(J+1)$ sources (J sensors and the composite HI) are further combined on decision level.

VI. Case Study

The proposed integrated framework of the dual-level information fusion method for RUL prediction is verified on the C-MAPSS data set, which is acquired from a commercial simulation software simulating the degradation process of aircraft engines. The prognostic results are then compared with those provided by using different information fusion schemes to verify its superiority of enhancing the prediction certainty. The integrated framework is also compared with several published works, to confirm its performance under the premise of providing explicit information fusion models.

A. Data Set Introduction

C-MAPSS data set contains four subsets, and the subset involving a single failure threshold and a single operation condition (FD001) is adopted here. This subset consists of a training set and a test set, where the former contains 100 run-to-failure training units and a total of 20,631 CM measurements, forming the historical database, and the latter contains 100 test units and 13,096 measurements. It should be mentioned that only partial measurements of the test units were acquired, based on which the users can estimate their RUL by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3055977, IEEE Internet of Things Journal

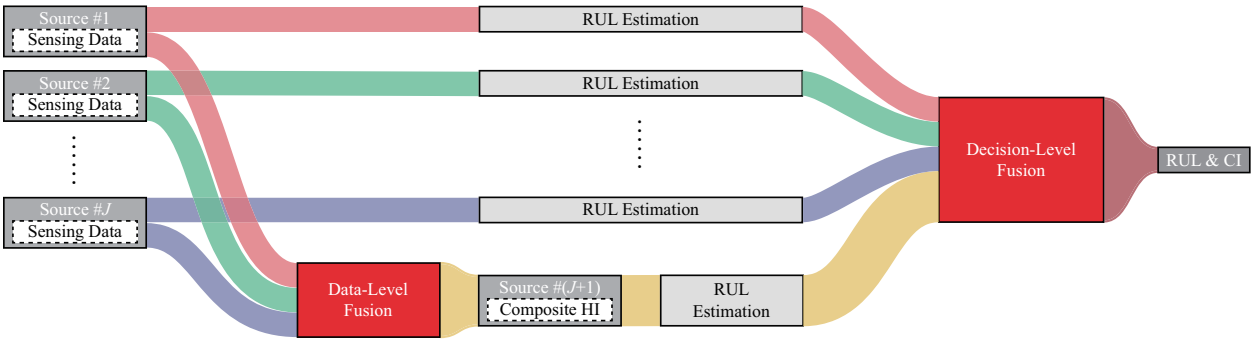

Figure 4. Information flow in the proposed framework of dual-level information fusion for RUL prediction.

using the knowledge learned from the training set. There are 21 monitoring sensors, i.e., information sources, from which 21 columns of CM signals are acquired simultaneously at every monitoring and details of all sensors can be accessed by referring to [5]. Besides, the actual RUL of each test unit is provided for users to evaluate their prediction.

B. Data Preprocessing

Data preprocessing mainly contains the selection of useful information sources and standardization. Since GP-based data-level fusion is a heuristic optimization method, the prediction result is expected to be greatly enhanced and the computational burden can be significantly reduced after discarding the irrelevant information sources. With the empirical engineering criteria that i) signals show clear increasing or decreasing trend, and ii) signals from the identical source for different training units possess the similar trend, and according to [5], 11 sensors including T24, T50, P30, Nf, Ps30, phi, NRf, BPR, htBleed, W31, and W32 are identified as the information sources to conduct data-level fusion.

Since the selected sensors measure diverse features of the engines, e.g., temperature, pressure, and fan speed, etc, the numerical scales of the provided information are consequently different. The data from each sensor are standardized for both training and test units by using z-score standardization, and the used mean and standard deviation are calculated as:

$$Mean\left(s\right)=\frac{\sum_{i=1}^{M}\sum_{k=1}^{N_i}s_{i,k}}{\sum_{i=1}^{M}N_i}, \tag{20}$$

$$Std\left(s\right)=\sqrt{\frac{\sum_{i=1}^{M}\sum_{k=1}^{N_i}\left(s_{i,k}-Mean\left(s\right)\right)^2}{\sum_{i=1}^{M}N_i-1}}, \tag{21}$$

where $M=100$, denoting the number of training units.

C. Data-Level Fusion and Prognostics Based on a Single Information Source

For signals of different units generated from a single source, a hybrid metric to improve the prediction accuracy was proposed as [34]:

$$\begin{aligned}Hyb\left(s_{1:M,1:N_i}\right)=&\,r\cdot Con\left(s_{1:M,1:N_i}\right)\\&+\left(1-r\right)\cdot Ran\left(s_{1:M,1:N_i}\right),\end{aligned} \tag{22}$$

which consists of two desired properties for improving the prognostics, i.e., 1) the consistency among training units under the same failure mode and operation condition, denoted by $Con\left(\cdot\right)$; 2) range information, denoted by $Ran\left(\cdot\right)$. The detailed definition of the two properties can be found in [34]. r is the tuning parameter to balance the two properties. By maximizing the hybrid metric, a composite HI was constructed via a nonlinear data-level fusion model produced by using GP [34]:

$$\mathrm{HI}=\cos\left[\sin\left(\sin\mathrm{W32}\right)\right]+\mathrm{T50}+\mathrm{Ps30}, \tag{23}$$

where W32 represents the low-pressure turbine (LPT) coolant bleed; T50 represents the total temperature at the LPT outlet; Ps30 represents the static pressure at the high-pressure compressor (HPC) outlet. The data-level fused HI can also be regarded as an extra information source similar to the 11 selected sensors. Based on the selected information source and the composite HI, the prognostics can be conducted, and the results have been presented in [34].

D. Relevant Information Source Identification

Since several sources may provide less-informative contributions for prognostics-orientated decision-level fusion, discarding the outcomes from several sources can enhance the accuracy and robustness of the decision level fusion [27].

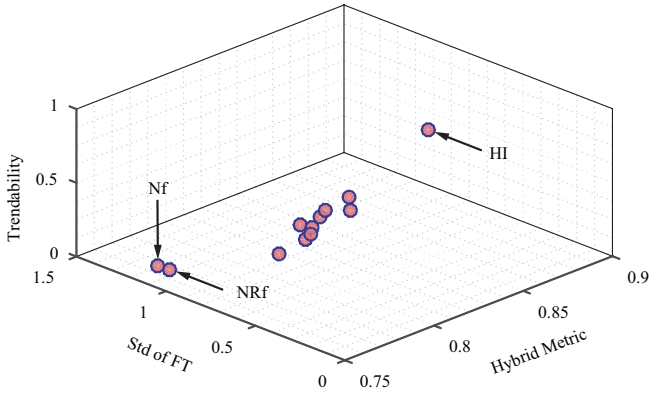

Figure 5. Graphic representation of the performance of the 11 selected sources and HI measured by the three engineering metrics.

Here three metrics proposed based on the engineering knowledge about RUL prognostics are applied to measure the

Table II
PERFORMANCE OF 11 SELECTED SOURCES AND HI MEASURED BY THREE ENGINEERING METRICS

	T24	T50	P30	Nf	Ps30	phi	NRf	BPR	htBleed	W31	W32	HI
Hybrid metric	0.7704	0.7969	0.7885	0.7584	0.7971	0.7891	0.7646	0.7786	0.7951	0.7871	0.7857	0.8588
Standard deviation	0.5724	0.4344	0.5231	1.1275	0.4448	0.5007	1.1240	0.4782	0.6710	0.5531	0.6031	0.6179
Trendability	0.3563	0.5963	0.5419	0.0699	0.6778	0.5958	0.0141	0.4941	0.2887	0.4650	0.4642	0.7601

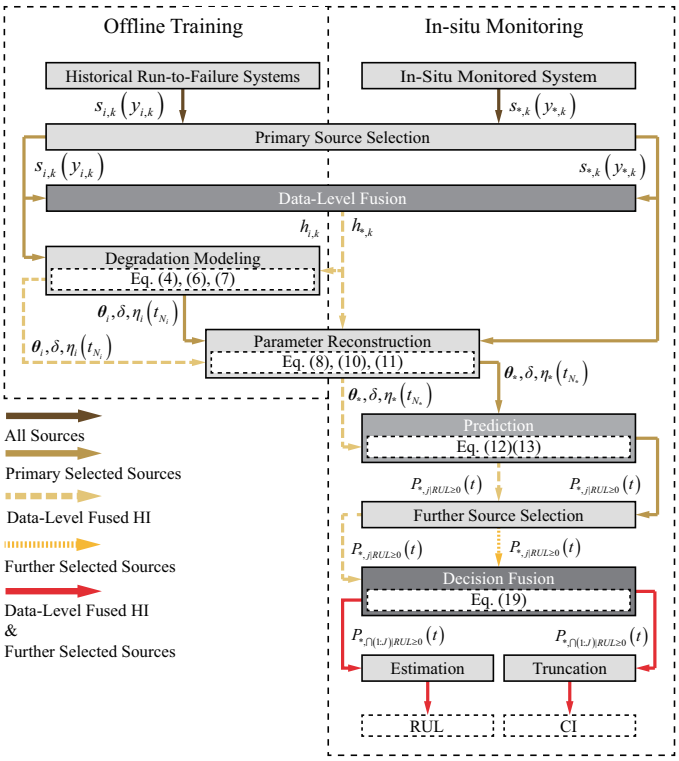

Figure 6. Framework of dual-level information fusion for RUL prediction.

performance of the 11 primarily selected sources as well as the composite HI, including 1) *hybrid metric* (22); 2) standard deviation of the last measurements of training units [6]; and 3) *trendability* [41]. The smaller the standard deviation of metric 2), the better the prediction results are expected to. The *hybrid metric* and the *trendability* quantify the performance of the sources into the interval $[0, 1]$, and the higher the values in terms of them, the better the information sources are for prognostics [41]. Performance of the sources measured by the three metrics is detailed in Table II and graphically presented in Fig. 5. In Table II, it can be seen that, compared with the primarily selected sensors, the composite HI has been improved by 7.74%∼13.24% in terms of the *hybrid metric*:

$$\frac{0.8588 - 0.7971}{0.7971} \times 100\% = 7.74\%,$$
$$\frac{0.8588 - 0.7584}{0.7584} \times 100\% = 13.24\%,$$

and by 12.15%∼5291.97% in terms of the *trendability*:

$$\frac{0.7601 - 0.6778}{0.6778} \times 100\% = 12.15\%,$$
$$\frac{0.7601 - 0.0141}{0.0141} \times 100\% = 5291.97\%.$$

Metric 2) is closely related to the range of the data, such that the performance of HI is not largely improved in terms of that metric since its range information has also been significantly improved [34]. In Fig. 5, it can be seen that Nf and NRf perform significantly inferior to other information sources. As a result, the decision-level fusion is conducted on the other 10 sources (T24, T50, P30, Ps30, phi, BPR, htBleed, W31, W32, and HI). The processes of the verification in this case study are summarized in Fig. 6.

E. Results

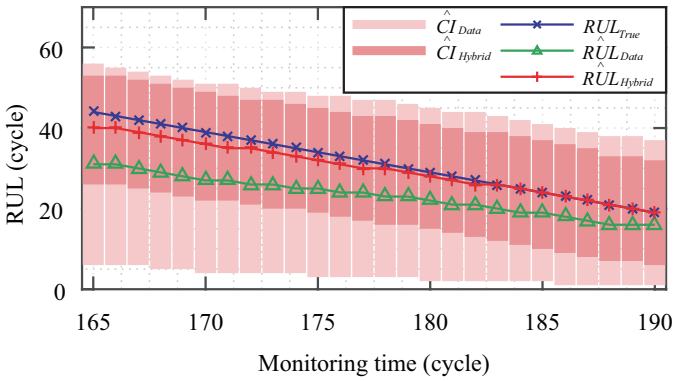

Figure 7. RUL prediction based on data-level fusion and the proposed hybrid fusion for test unit #35.

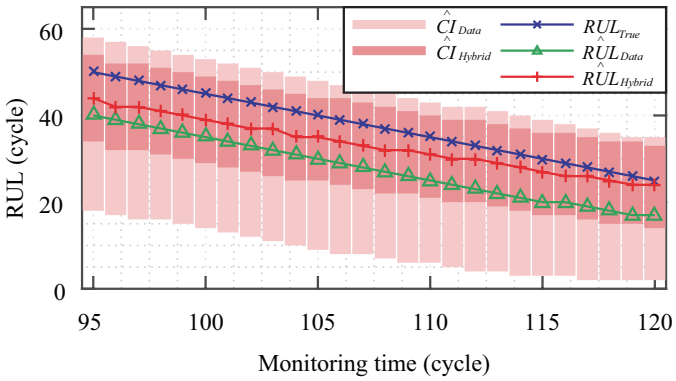

Figure 8. RUL prediction based on data-level fusion and the proposed hybrid fusion for test unit #36.

Given the significance level $\alpha = 0.05$, the widths of CIs can be used to evaluate the uncertainty of estimation [6]. Two test units, #35 and #36, are randomly selected to illustrate the *in-situ* monitoring and prognostics, as shown in Fig. 7 and Fig. 8, where RUL_{True} represents the actual RUL; $\hat{RUL}_{Data}$ and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3055977, IEEE Internet of Things Journal

$\hat{RUL}_{Hybrid}$ represent the estimated RUL respectively by using GP-based data-level fusion and the proposed hybrid dual-level fusion; $\hat{CI}_{Data}$ and $\hat{CI}_{Hybrid}$ represent the corresponding 95% CI respectively. It can be seen that compared with the data-level fusion, the prognostic accuracy (i.e., the prediction error) and uncertainty (i.e., the widths of the 95% CIs) are all improved by using the the proposed data-decision-level hybrid information fusion. Also, since more CM data can be collected along with the monitored systems, and parameter reconstruction can be performed in the cloud plane at each monitoring time, the accuracy of degradation modeling is increasing. As a result, the predicted RUL of these systems is becoming more accurate to the ground truth. The comparison at the five monitoring time for test unit #35 is further detailed in Table III. For example, at $t = 180$ cycles, the actual RUL is 29 cycles, and that estimated by using data-level fusion and hybrid fusion are 22 cycles and 28 cycles, respectively. The 95% CIs are estimated as $[2, 45]$ cycles and $[15, 41]$ cycles respectively, showing a significant reduction by 39.5% ($[(41-15)-(45-2)]/(45-2)\times100\%=-39.5\%$) of uncertainty based on hybrid fusion.

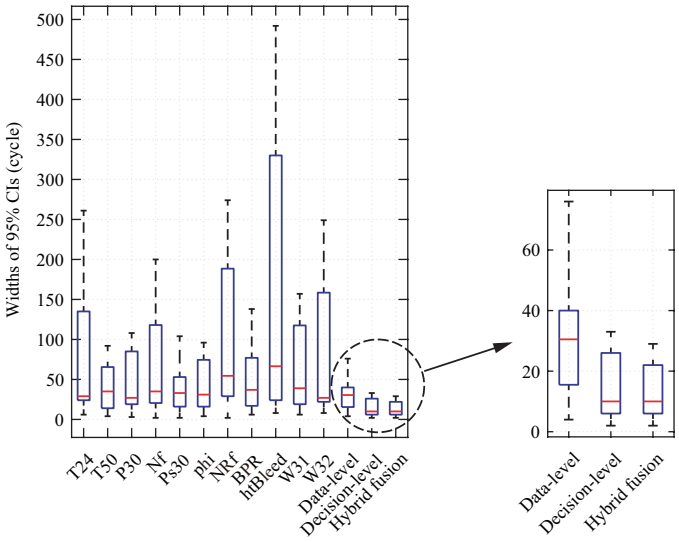

Figure 9. Widths of CIs of all the test units based on different combinations of information sources.

1) Comparisons of Remaining Useful Life Prediction Results: 95% CI widths of all the test units estimated based on each single sensor, data-level fused HI, decision-level fusion, and data-decision-level fusion are shown in Fig. 9. The decision level fusion is conducted on the PDF of RUL estimated by using the 9 further selected sensors (detailed in Subsection VI-D). When HI is not involved in the fusion, the hybrid fusion will be a decision-level fusion. Compared with widths of CIs estimated by using each single sensor, those estimated by using fusion techniques are reduced. The mean CI width is further detailed in Table IV. As can be seen in Fig. 9 and Table IV, after removing the irrelevant information sources, the uncertainty shows a decreasing trend as the number of the actually involved information sources increases. Compared with the narrowest mean 95% CI width of RUL estimated based on the best single source (sensor Ps30), that estimated based on the fusion of 3 sources (according to eq. (23)) is reduced by 19.6% ($(30.0-37.3)/37.3\times100\%=-19.6\%$); compared with the result based on the fusion of 3 sources, that estimated based on the fusion of 9 sources is reduced by 48.7% ($(15.4-30.0)/30.0\times100\%=-48.7\%$); since the fused HI can be regarded as an extra information source, compared with the result based on the fusion of 9 sources, that estimated based on the fusion of 10 sources is further reduced by 7.1% ($(14.3-15.4)/15.4\times100\%=-7.1\%$), showing a significant improvement by using information fusion in reducing the prognostic uncertainty.

Mean Absolute Percentage Error (MAPE) is adopted to evaluate the proposed hybrid fusion method for RUL prognostics, which is defined as [5]:

$$MAPE = \frac{100\%}{P}\sum_{i=1}^{P}\frac{\left|\hat{T}_i - T_i^r\right|}{t_{N_i} + T_i^r} \tag{24}$$

where T_i^r is the actual RUL of the test unit i; $\hat{T}_i$ is the estimated RUL of the test unit i; $(t_{N_i} + T_i^r)$ denotes the whole life of the test unit i; and P represents the total number of test units.

To investigate the performance of the proposed method for *in-situ* operating systems, MAPE and the mean width of 95% CI by using data-level fusion, decision-level fusion and the proposed hybrid fusion schemes are investigated in terms of the actual RUL level of the test units, as shown in Fig. 10 and Fig. 11 respectively, where label *all* under

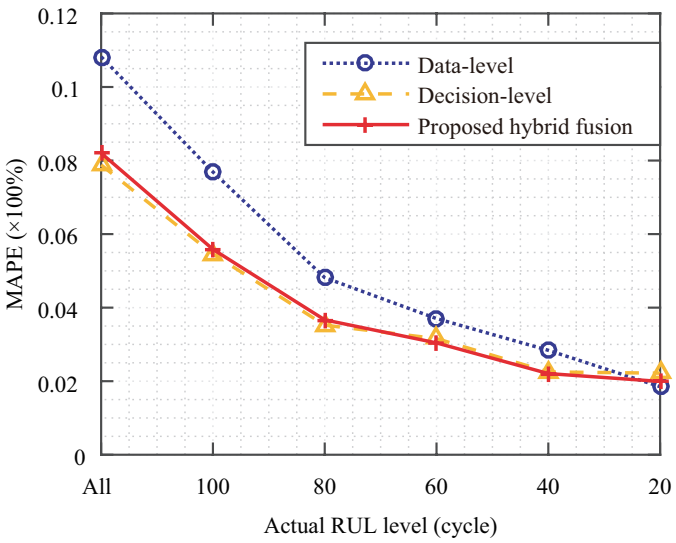

Figure 10. MAPE by using the three different fusion schemes.

the x-axes represents the corresponding metrics for all the 100 test units, and label *80* represents those metrics only for the test units whose actual RUL is not more than 80 cycles. Compared with data-level fusion, prognostics by using decision-level and the hybrid fusion achieves significant lower MAPE and narrower CI width, indicating the effectiveness of incorporating more informative sources (3 sources for data-level fusion, 9 sources for decision-level fusion, and 10 sources for the hybrid fusion) for improving the prognostic accuracy

Table III
MEAN WIDTH OF 95% CI OF THE TEST UNIT #35 ESTIMATED BY USING DIFFERENT SENSORS AND FUSION SCHEMES (UNIT: CYCLE)

Monitoring time	True RUL	RUL Estimation		95% CI Estimation	
		$\hat{RUL}_{Data}$	$\boldsymbol{\hat{RUL}_{Hybrid}}$	$\hat{CI}_{Data}$	$\boldsymbol{\hat{CI}_{Hybrid}}$
t =165	44	31	**40**	[6, 56]	**[26, 53]**
t =170	39	27	**36**	[4, 51]	**[22, 49]**
t =175	34	25	**32**	[3, 48]	**[19, 45]**
t =180	29	22	**28**	[2, 45]	**[15, 41]**
t =185	24	19	**24**	[2, 41]	**[10, 37]**
t =190	19	16	**19**	[1, 37]	**[6, 32]**

Table IV
MEAN WIDTH OF 95% CI OF TEST UNITS ESTIMATED USING DIFFERENT SENSORS AND FUSION SCHEMES (UNIT: CYCLE)

Source	T24	T50	P30	Nf	Ps30	phi	NRf
Mean CI width	80.5	38.4	49.7	68.1	37.3	43.7	97.3
Source	BPR	htBleed	W31	W32	Data-level fusion	Decision-level fusion	Hybrid fusion
Mean CI width	46.2	165.9	65.2	87.1	30.0	15.4	**14.3**

and reducing the uncertainty. Compared with decision-level fusion, the widths of CIs can be further narrowed by introducing the composite HI while a comparable prognostic accuracy is kept. Also, as can be seen, when these systems run closer to their EoL, the predicted RUL is becoming more accurate referring to the ground truth, and the prognostic uncertainty is reduced statistically in terms of the mean width of CI, which consequently improves the prognostic precision. Moreover, the proposed integration method of degradation modeling and parameter reconstruction to predict RUL can be categorized as the degradation-based method [17]. RUL prediction can also be regarded as an estimation task, and for degradation-based methods, prognostic accuracy is related to both the model assumptions and the applied methods. As a result, compared with accuracy, the prognostic precision is improved by using multi-source information fusion.

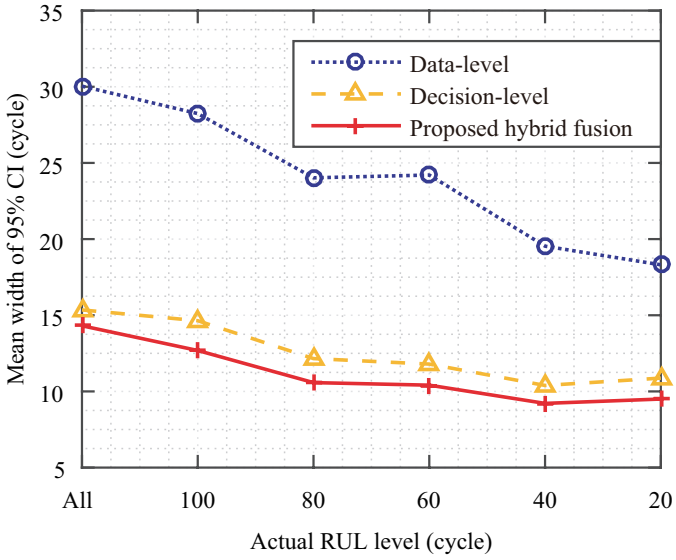

Figure 11. Mean width of the 95% CI by using the three different fusion schemes.

2) Prognostic Accuracy: The proposed method is compared with the following 6 representative benchmark methods in terms of MAPE:

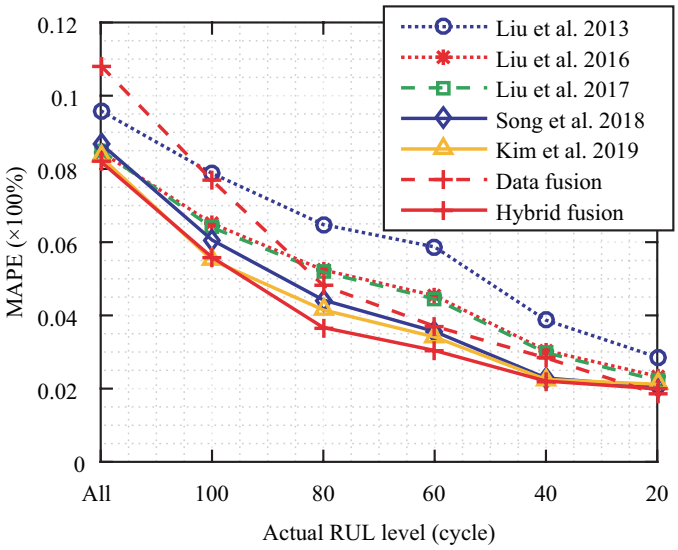

Figure 12. Comparison with the existing methods in terms of MAPE.

(1) Linear data-level fusion with optimizing two desired properties designed according to domain knowledge [5] (legend *Liu et al. 2013*);

(2) Linear data-level fusion with optimizing two desired properties designed integrating both domain knowledge and degradation modeling [6] (legend *Liu et al. 2016*);

(3) Linear data-level fusion with optimizing a designed signal-to-noise ratio (SNR) integrating both domain knowledge and degradation modeling [7] (legend *Liu et al. 2017*);

(4) Nonlinear data-level fusion based on KM with optimizing the designed signal-to-noise ratio (SNR) [10] (legend *Song et al. 2018*);

(5) Linear data-level fusion with directly optimizing the prognostic error [9] (legend *Kim et al. 2019*);

Table V
COMPARISON WITH THE EXISTING METHODS IN TERMS OF MAPE (UNIT: %)

Level of actual RUL	Number of test units	**Proposed hybrid fusion**	GP-based data-level fusion [34]	Liu et al. 2013 [5] ($RE \leq$0.05)	Liu et al. 2016 [6] ($RE \leq$0.05)	Liu et al. 2017 [7] ($RE \leq$0.01)	Song et al. 2018 [10] ($RE \leq$0.02)	Kim et al. 2019 [9] ($RE \leq$0.02)
All	100	**8.20**	10.82	9.59	8.47	8.51	8.69	8.33
$\leqslant$100	67	5.59	7.69	7.88	6.51	6.40	6.06	**5.51**
$\leqslant$80	45	**3.67**	4.82	6.48	5.24	5.18	4.42	4.15
$\leqslant$60	39	**3.04**	3.70	5.87	4.54	4.45	3.56	3.41
$\leqslant$40	28	**2.21**	2.84	3.87	3.04	2.98	2.29	2.25
$\leqslant$20	16	1.99	**1.85**	2.86	2.34	2.23	2.05	2.12

(6) Nonlinear data-level fusion based on GP with optimizing two desired properties designed according to domain knowledge [34] (legend *Data fusion*).

The results are shown in Fig. 12 and detailed in Table V, in terms of the actual RUL level of all the test units. In Table V, note that the results of Liu et al., Song et al., and Kim et al. are obtained from the figures in [5], [6], [7], [10], [9] by using *Getdata Graph Digitizer*, which is professional software that can help to read data from figures, and thus a rough upper limit of the absolute value of Reading Error (RE) is provided as well. As can be seen in Fig. 12 and Table V, the proposed hybrid fusion method outperforms linear data-level fusion [5], [6], [7], KM-based nonlinear data-level fusion [10], and GP-based nonlinear data-level fusion [34] in terms of the MAPE, or in terms of the uncertainty at a close MAPE. For instance, for the 16 test units with actual RUL $\leq$ 20 cycles, the mean width of 95% CIs is 9.5 cycles by using the hybrid fusion method, and 18.3 cycles by using the GP-based data-level fusion, showing that uncertainty can be significantly reduced by using the proposed method when the performance of prediction is closed. For data-level fusion, generally, when the optimization objectives belong to the same category, i.e., the degradation-based properties or the direct prognostic error, nonlinear models can provide better prognostics than linear ones due to their more powerful representation capability. When the fusion models belong to the same category, such as linear combination, fusing information by directly optimizing the prognostic error can improve the performance. However, fusing information by optimizing degradation-based properties can provide a better insight into the entire degradation pattern that the monitored systems will follow, consequently providing more informative and transparent outcomes. This can be useful because it allows checking the prediction consistency considering experts' domain knowledge and the information on-line acquired during the degradation [17]. To be more specific, according to the results in Table V, fusing information by optimizing the prognostic error [9] has a significant advantage of the prediction accuracy, despite fusion models with more powerful representation capability can be provided by GP [34] and KM [10] with optimizing degradation-based properties. By fusing information from more useful sources, the accuracy of degradation-based methods can be improved to a close level to those optimizing prognostic error, and the interpretability about the degradation process can be retained as well. The proposed RUL prediction method with data-level information fusion outperforms the other involved methods when the monitored systems run close to their EoL, but performs inferior in the early stage when they are a long time away from the EoL. One possible reason is that the monitored systems usually have not shown any obvious degradation trend in the early stage of their operation, and the acquired CM information is little, posing negative impacts on parameter reconstruction. As a result, if more information sources are introduced for fusion, it is expected that a small amount of information from multiple sources can compensate for each other, providing relatively global better prognostic results as shown in Table V.

Root Mean Squared Error (RMSE) is also adopted to evaluate prognostic performance, which is defined as:

$$RMSE = \sqrt{\frac{1}{P}\sum_{i=1}^{P}\left(\hat{T}_i - T_i^r\right)^2}, \tag{25}$$

and the proposed method is compared with several benchmark methods according to RMSE as well, i.e.:

(1) Wavelet feature extraction combined with 'haar' features (WFE-haar, $RMSE = 42.98$) [42];

(2) Multi-Layer Perceptron (MLP, $RMSE = 37.56$) [43];

(3) Support Vector Machine (SVM, $RMSE = 33.53$) [44];

(4) Linear Regression (LR, $RMSE = 31.21$) [45];

(5) Nonlinear data-level fusion based on GP with optimizing two desired properties designed according to domain knowledge ($RMSE = 30.95$) [34].

The proposed hybrid-fusion-based approaches can provide a prediction in terms of that $RMSE = 24.31$, outperforming the above approaches. Compared with several machine learning approaches such as SVM [44] and MLP [46], the proposed method can provide both closed-form estimated RUL and CI, which is favorable for decision-making. Also, the proposed method outperforms these two representative data-driven machine learning methods regarding the prognostic accuracy. One possible reason can be that a mixed-effect degradation model is constructed in the proposed method to characterize the latent degradation process, in which the degradation trend is modeled as a power-functional form, and specific FTs are estimated on run-to-failure units. This implementation is a key factor for improving the performance of RUL prediction, since it is based on important domain knowledge which is commonly not utilized in many pure data-driven methods, such as neural networks and KM for RUL prognostics on the basis of regression. On the other hand, from the perspective of information fusion, this domain knowledge can be regarded

as another information from sources like experts, historical cases, etc., although it is not explicitly presented in structured data. This information can serve as constraints when predicting the RUL, while it needs to be discovered from input data for those pure data-driven methods. Moreover, most of the listed approaches lack of analyses in terms of the uncertainty of the estimated RUL, which are challenging for both uncertainty management and decision making in practice.

3) Time Costs: All the approaches are performed on a regular laptop (Intel Core i7-8550U, 1.8 GHz, 16 GB RAM) with MATLAB 2015b. Time costs for parameter estimation, reconstruction, and GP-based data-level fusion model training have been investigated on the identical laptop in detail [34]. The average time cost of GP to produce the data-level fusion models is not more than 225 s. Parameter estimation and reconstruction for the degradation model consume less than 2 ms and 0.5 s respectively, indicating that the cloud computation center needs not to be extremely powerful. Here the time costs of data-level fusion and decision-level fusion are investigated. Data-level fusion model represented by (23) and decision-level fusion model represented by (19) was performed on all of the 100 testing units. The mean time cost is 10.12 μs with a 8.25 μs standard deviation for data-level fusion, and 65.90 μs with a 10.24 μs standard deviation for decision-level fusion, showing the computational-efficiency and considerable potentials to be deployed into IIoT.

F. Discussion

One of the challenges of applying the proposed data-decision-level hybrid information fusion method for PdM in IIoT is the heuristic feature of GP. Due to the fact that GP is essentially a heuristic search technique that explores an optimal or at a least suitable solution from available alternatives, it may converge to a local optimum. Multiple runs of GP can mitigate this problem to some extent. Also, selecting several sub-optimal individuals in addition to the best one in the final generation as candidate solutions is another effective approach to address this challenge. Another challenge is that the produced models are not easy to be interpreted with their physical implications [23]. Since elementary function operators can be introduced into GP in addition to the four fundamental operators of arithmetic, it can be a potential research field to discover physically interpretable information fusion models by using GP. Besides, the mechanism that GP can adaptively select a subset of the input information sources has not been fully explored and utilized, benefiting an integration of source selection and information fusion. This can also reduce manual intervention to the automatic IIoT applications. In addition to fusing information, GP has also been used for data-driven RUL prediction [47], where it can further simplify the relations between the CM data and predicted RUL as an explicit and end-to-end formula, greatly facilitating the embedded real-time monitoring in IIoT.

Since a probability distribution is much more informative than a BBD [17], more than one BBD can be induced from a PDF, and it may be challenging to select a suitable BBD to be induced. Different BBD induced from the same PDF can be ordered according to the strength of the beliefs they represent [39], and further, they can be ordered according to different specific ordering criteria, such as *pl-ordering*, *q-ordering*, and *s-ordering*. It can be practical to induce a BBD according to *q-ordering* and the principle of minimal commitment, i.e., the *q*-Least-Committed (*q*-LC) BBD, which means selecting the least committed belief function in a set of equally justified belief functions [39].

VII. Conclusion

In this paper, a data-decision-level hybrid information fusion method for PdM in IIoT is proposed to estimate RUL of complex industrial systems. On data level, sensing data are fused into a composite HI by using GP, in which a nonlinear explicit fusion model is constructed. The composite HI is added as an extra information source, from which as well as other sources, the PDF of RUL is derived independently and respectively. After discarding the irrelevant sources considering the three widely-used criteria to evaluate the quality of the sources, the PDF of RUL deduced by using each source is further combined in the framework of BFT. The experimental results show that compared with sole data-level and decision-level fusion methods, the prognostic uncertainty is reduced by 52.3% ((14.3−30.0)/30.0×100%) and 7.1% ((14.3−15.4)/15.4×100%), respectively, by using the proposed dual-level information fusion in terms of the mean width of CI. Also, there is an improvement of the prediction accuracy by 24.2% ((8.20%−10.82%)/10.82%×100%) in terms of MAPE, compared to the constructed GP-based data-level fusion. With this approach, the selection of the relevant sources to construct the composite HI can be mitigated since it can be conducted automatically. The proposed decision-level fusion method also provides a convenient way to aggregate estimations acquired by using multiple sub-models. Integrated with the proposed approach, IIoT is expected to reach the objective that support decision making to optimally act on physical systems.

References

[1] S. Zhao, F. Blaabjerg, and H. Wang, "An overview of artificial intelligence applications for power electronics," *IEEE Trans. Power Electron.*, vol. 36, no. 4, pp. 4633–4658, Apr. 2021.

[2] L. Ren, Y. Liu, X. Wang, J. LÃŒ, and M. J. Deen, "Cloud-edge based lightweight temporal convolutional networks for remaining useful life prediction in IIoT," *IEEE Internet Things J.*, 2020, to be published.

[3] M. Compare, P. Baraldi, and E. Zio, "Challenges to IoT-enabled predictive maintenance for industry 4.0," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4585–4597, May 2020.

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

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

[6] 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.

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

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

[9] M. Kim, C. Song, and K. Liu, "A generic health index approach for multisensor degradation modeling and sensor selection," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 3, pp. 1426–1437, Jul. 2019.

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

[11] Y. Liu, X. Hu, and W. Zhang, "Remaining useful life prediction based on health index similarity," *Rel. Eng. Syst. Safety*, vol. 185, pp. 502–510, 2019.

[12] Y. Li, J. Shi, G. Wang, and M. Zhang, "An ensemble model for engineered systems prognostics combining health index synthesis approach and particle filtering," *Quality Rel. Eng. Int.*, vol. 33, no. 8, pp. 2711–2725, 2017.

[13] Y. Z. Rosunally, S. Stoyanov, C. Bailey, P. Mason, S. Campbell, G. Monger, and I. Bell, "Fusion approach for prognostics framework of heritage structure," *IEEE Trans. Rel.*, vol. 60, no. 1, pp. 3–13, Mar. 2011.

[14] J. Chen, D. Zhou, C. Lyu, and C. Lu, "A novel health indicator for pemfc state of health estimation and remaining useful life prediction," *Int. J. Hydrogen Energy*, vol. 42, no. 31, pp. 20 230–20 238, 2017.

[15] A. Al-Dulaimi, S. Zabihi, A. Asif, and A. Mohammadi, "A multimodal and hybrid deep neural network model for remaining useful life estimation," *Comput. Ind.*, vol. 108, pp. 186–196, 2019.

[16] Y. Song, S. Gao, Y. Li, L. Jia, Q. Li, and F. Pang, "Distributed attention-based temporal convolutional network for remaining useful life prediction," *IEEE Internet Things J.*, 2020, to be published.

[17] P. Baraldi, F. Mangili, and E. Zio, "A belief function theory based approach to combining different representation of uncertainty in prognostics," *Inform. Sci.*, vol. 303, pp. 134–149, 2015.

[18] L. Duan, F. Zhao, J. Wang, N. Wang, and J. Zhang, "An integrated cumulative transformation and feature fusion approach for bearing degradation prognostics," *Shock Vibration*, 2018.

[19] X. Q. Li, H. K. Jiang, X. Xiong, and H. D. Shao, "Rolling bearing health prognosis using a modified health index based hierarchical gated recurrent unit network," *Mechanism Mach. Theory*, vol. 133, pp. 229–249, 2019.

[20] J. L. Chen, H. J. Jing, Y. H. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Rel. Eng. Syst. Safety*, vol. 185, pp. 372–382, 2019.

[21] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, 2017.

[22] Z. Li and Q. He, "Prediction of railcar remaining useful life by multiple data source fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2226–2235, Aug. 2015.

[23] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2464–2472, May 2014.

[24] A. S. Qin, Q. H. Zhang, Q. Hu, G. X. Sun, J. He, and S. Q. Lin, "Remaining useful life prediction for rotating machinery based on optimal degradation indicator," *Shock Vibration*, 2017.

[25] H. Wang, G. Dong, and J. Chen, "Application of improved genetic programming for feature extraction in the evaluation of bearing performance degradation," *IEEE Access*, vol. 8, pp. 167 721–167 730, 2020.

[26] H. Li, H.-Z. Huang, Y.-F. Li, J. Zhou, and J. Mi, "Physics of failure-based reliability prediction of turbine blades using multi-source information fusion," *Appl. Soft Comput.*, vol. 72, pp. 624–635, 2018.

[27] P. Baraldi, E. Zio, G. Gola, D. Roverso, and M. Hoffmann, "Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems," *Ann. Nucl. Energy*, vol. 38, no. 2-3, pp. 212–220, 2011.

[28] R. Polikar, "Bootstrap-inspired techniques in computation intelligence," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 59–72, Jul. 2007.

[29] C. Duan, C. Deng, and N. Li, "Reliability assessment for CNC equipment based on degradation data," *Int. J. Advanced Manuf. Technol.*, vol. 100, no. 1-4, pp. 421–434, 2019.

[30] C. Ma, X. Zhai, Z. Wang, M. Tian, Q. Yu, L. Liu, H. Liu, H. Wang, and X. Yang, "State of health prediction for lithium-ion batteries using multiple-view feature fusion and support vector regression ensemble," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 9, pp. 2269–2282, 2019.

[31] W. Yu, I. Y. Kim, and C. Mechefske, "Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme," *Mech. Syst. Signal Process.*, vol. 129, pp. 764–780, 2019.

[32] S. Zhao, S. Chen, F. Yang, E. Ugur, B. Akin, and H. Wang, "A composite failure precursor for condition monitoring and remaining useful life prediction of discrete power devices," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 688–698, Jan. 2021.

[33] A. Appriou, *Uncertainty Theories and Multisensor Data Fusion*. John Wiley & Sons, 2014.

[34] P. Wen, S. Zhao, S. Chen, and Y. Li, "A generalized remaining useful life prediction method for complex systems based on composite health indicator," *Rel. Eng. Syst. Safety*, vol. 205, 2021.

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

[36] S. Zhao, V. Makis, S. Chen, and Y. Li, "Evaluation of reliability function and mean residual life for degrading systems subject to condition monitoring and random failure," *IEEE Trans. Rel.*, vol. 67, no. 1, pp. 13–25, Mar. 2018.

[37] S. Zhao, V. Makis, S. Chen, and Y. Li, "Health assessment method for electronic components subject to condition monitoring and hard failure," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 138–150, Jan. 2019.

[38] A. Chehade, S. Bonk, and K. Liu, "Sensory-based failure threshold estimation for remaining useful life prediction," *IEEE Trans. Rel.*, vol. 66, no. 3, pp. 939–949, Sep. 2017.

[39] P. Smets, "Belief functions on real numbers," *Int. J. Approximate Reasoning*, vol. 40, no. 3, pp. 181–223, 2005.

[40] T. Strat, "Continuous belief function for evidential reasoning," in *4th Nat. Conf. Artificial Intell.*, Austin, TX, USA, Aug. 1984, pp. 308–313.

[41] P. Baraldi, G. Bonfanti, and E. Zio, "Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics," *Mech. Syst. Signal Process.*, vol. 102, pp. 382–400, 2018.

[42] P. Lim, C. K. Goh, and K. C. Tan, "A novel time series-histogram of features (TS-HoF) method for prognostic applications," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 3, pp. 204–213, Jun. 2018.

[43] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.

[44] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017.

[45] S. K. Singh, S. Kumar, and J. Dwivedi, "A novel soft computing method for engine RUL prediction," *Multimedia Tools Appl.*, vol. 78, no. 4, pp. 4065–4087, 2019.

[46] G. S. Babu, P. Zhao, and X. L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *2016 Int. Conf. Database Syst. Advanced Appl.*, Dallas, TX, USA, Mar. 2016, pp. 214–228.

[47] C. Chen, T. H. Xu, G. Wang, and B. Li, "Railway turnout system RUL prediction based on feature fusion and genetic programming," *Measurement*, vol. 151, 2020.

Pengfei Wen (S'20) received the B.S. degree in Communication Engineering in the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, in 2017, where he is currently pursuing the Ph.D. degree in Information and Communication Engineering from 2019. His current research interests include Information Fusion, Reliability Engineering and Industrial Internet of Things.

Yong Li received the B.S. degree in Avionics Engineering, M.S. and Ph.D degrees in Circuits and Systems from Northwestern Polytechnical University, Xi'an, China, in 1983, 1988 and 2005, respectively. He joined School of Electronic Information, Northwestern Polytechnical University in 1993 and was promoted to professor in 2002. His research interests include digital signal processing and radar signal processing.

Shaowei Chen (M'15) is currently an associate professor in the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, where he is also the dean of the Department of telcommunication engineering and the Director of Perception and IoT Information Processing Laboratory. He is the principal investigator of several projects supported by the Aeronautical Science Foundation of China, Beijing, China. His research expertise is in the area of the fault diagnosis, sensors, condition monitoring, and prognosis of electronic systems. Prof. Chen is selected to receive several Provincial and Ministerial Science and Technology Awards. He is a senior member of the Chinese Institute of Electronics.

Shuai Zhao (S'14-M'18) received the B.E. (hons.), M.E., and Ph.D. degrees in information and communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2011, 2014, and 2018, respectively. He is currently a Postdoctoral Researcher with the Center of Reliable Power Electronics, Department of Energy Technology, Aalborg University, Aalborg, Denmark. From 2014 to 2016, he was a Visiting Ph.D. Student with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada, with the scholarship from China Scholarship Council. In August 2018, he was a Visiting Scholar with the Power Electronics and Drives Laboratory, Department of Electrical and Computer Science, The University of Texas at Dallas, Richardson, TX, USA. His research interests include system informatics, intelligent condition monitoring, diagnostics and prognostics, and tailored AI tools for power electronic systems.