Di Wang, *Member, IEEE*, Yuhui Wang, and Xiaochen Xian, *Member, IEEE*

[[1]](#footnote-1)

Degradation Modeling and Prognostics Via A Data Fusion-Based Multitask Learning Approach

*Abstract*—Remaining useful lifetime (RUL) prediction based on degradation signals of sensors is of critical importance in prognostics and health management (PHM) of machines in manufacturing systems. The machines are often under the same operational and environmental condition and thus show dependent but heterogeneous degradation profiles. However, many existing studies ignored these characteristics among machines, leading to poor RUL prediction results especially when missing data exists. To address this issue, this paper develops a data fusion-based multi-task learning (MTL) method for degradation modeling and prognostics of machines based on multiple sensor signals. Specifically, we first adopt a quadratic polynomial model to fit the degradation processes of machines based on health index (HI) constructed by a linear fusion model. Considering the dependency and heterogeneity among machines, we innovatively integrate MTL into the degradation model. And for parameter estimation, we develop a unique iterated quadratic optimization-integrated expectation-maximum (QOIEM) algorithm that considers the properties of HI construction and simultaneously updates parameters of the fusion model and the degradation model through quadratic programming (QP) and maximum a posteriori-based expectation-maximum (MAP-EM). The QOIEM has been proved to have satisfactory convergence performance in theory in this study. Numerical experiments and a real case study on aircraft turbine engines have been implemented to compare the prognosis performance and robustness on incomplete dataset for the proposed and benchmark methods.

*Note to Practitioners*— This paper aims to develop a data fusion-based MTL approach for degradation modeling and prognostics based on multiple sensor signals. The proposed method addresses the following challenges in practice: 1) how to collectively utilize the common information with a group of machines under similar or dependent operational conditions; 2) how to handle missing values in observed signals to achieve satisfactory RUL prediction results. There are four steps for implementing this method in practice: 1) collecting multiple sensor signals of machines (e.g., aircraft engines); 2) establishing the data fusion-based MTL model that shares common information among machines for degradation modeling and prognostics; 3) conducting the QOIEM algorithm to estimate model parameters; 4) making prognostic predictions for in-service machines. The novelty of the proposed method is that it utilizes the data fusion-based MTL model to characterize the dependency and heterogeneity among machines for better degradation modeling and prognostics. The proposed method has superior prognosis results especially when missing data exists comparing with state-of-the-art benchmarks.

*Index Terms*—Data fusion-based multitask learning, quadratic optimization-integrated expectation-maximum algorithm, missing data

# INTRODUCTION

D

EGRADATION of machines is inevitable in most of manufacturing systems, such as aircraft engines [1], [2], bearing systems [3], [4] and lithium batteries [5], [6]. Prognostics and health management (PHM) of machines by considering their degradation processes is critically important to ensure the stability of the overall manufacturing systems. The remaining useful lifetime (RUL) prediction of machines is one of the most essential tasks in PHM. Based on accurately predicted RUL, maintenance or replacement of machines can be arranged reasonably.

In existing studies, many physics-based methods have been operated to obtain the prediction of RUL through analyzing the physical processes of machines in manufacturing systems [7], [8]. However, these methods are based on an in-depth understanding of the degradation mechanisms and thus are applicable only to simple systems with known degradation mechanisms. They may encounter trouble in RUL prediction of complex systems with unfamiliar degradation mechanisms. To alleviate this problem, data-driven methods predict the RUL based on collected sensor signals that reflect the degradation status of machines. Single sensor signal-based analysis selects a single type of sensor signals that can mainly reflect the degradation status of corresponding machine and then fit a degradation model based on the selected sensor signals, such as stochastic process models [9], [10] and general path models [11], [12]. However, as single sensor signal only contains partial information about the degradation status, analysis based on single sensor signal fails to characterize the degradation processes of machines in complex systems which have multiple sensors that contain comprehensive information of the degradation status, and therefore gains inaccurate estimation of RUL [1], [13]. To fully utilize such information, researchers have focused on analyzing multiple sensor signals in systems.

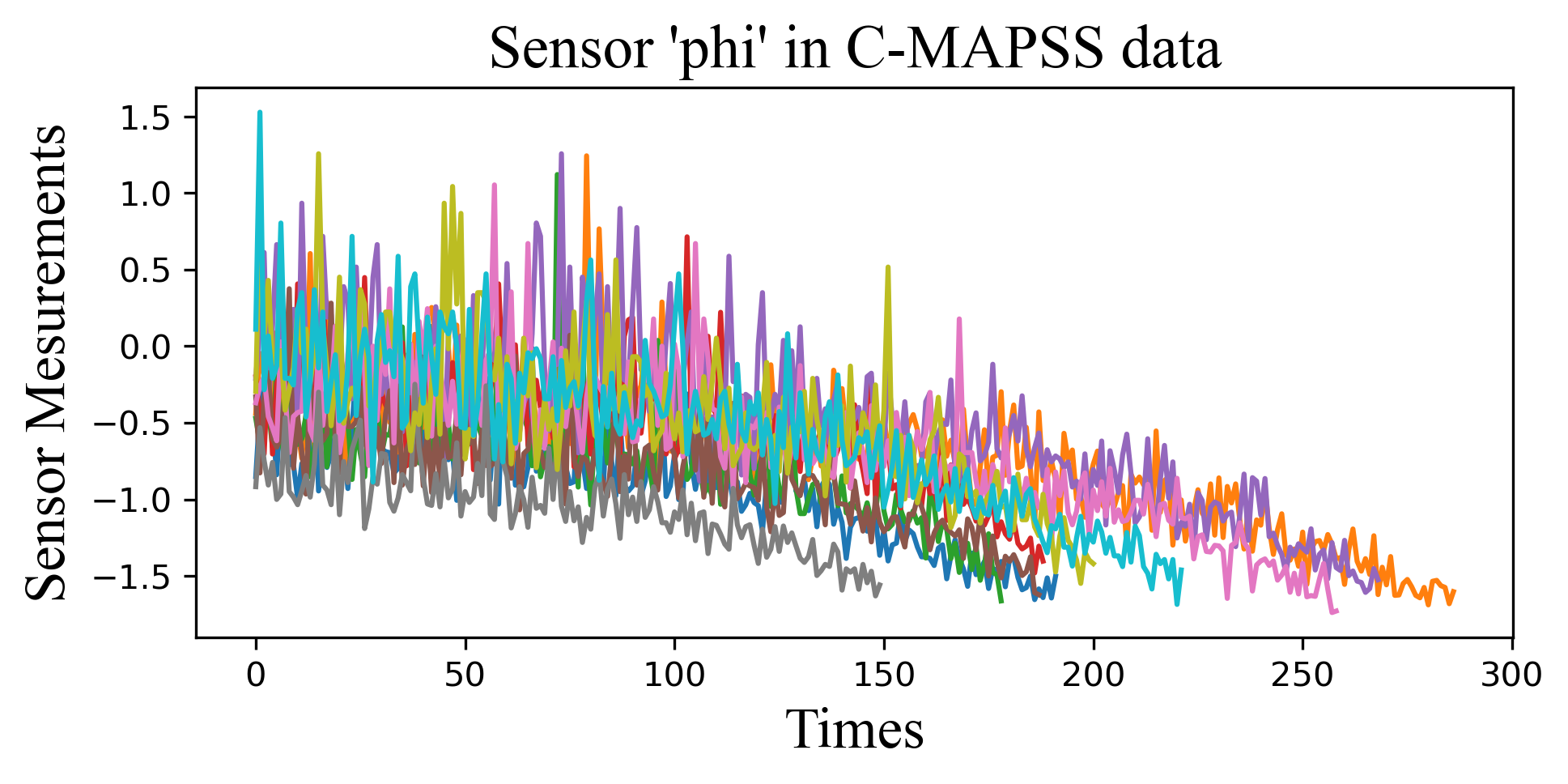


Fig. 1. Degradation signals of a group of engines in an aircraft.

For multiple sensor signal-based analysis, data fusion is an effective way because it can fuse comprehensive degradation information from multiple aspects of machines for better RUL prediction. Existing data fusion methods can be divided into two categories, i.e., decision-level fusion and data-level fusion. Decision-level fusion focuses on fusion methods of different prediction results. The different prediction results can be provided by considering multiple algorithms [14], multiple failure modes [13] and multiple environmental conditions [15]. The limitation of decision-level fusion is that it regards each sensor signal as a single degradation process and analyzing different sensors separately, owing to which the results are sensitive to the sensor signals that cannot well characterize the degradation status. Data-level fusion focuses on combining information from multiple sensor signals to obtain a more informative fused signal. Such data-level fusion methods in the literature include functional principal components analysis [16], hidden Markov models [17], [18], neuro-fuzzy systems [19]–[21] and neural networks [22], [23].

In particular, health index (HI)-based methods in data-level fusion have received widespread attention in recent years. HI is designed to characterize the underlying degradation processes of machines, and it can also be interpreted as a visualization of machines’ health statuses. Liu et al. [24] designed HI as a weighted sum of multiple sensor signals. Specifically, they proposed several theoretical assumptions to ensure the monotonicity, stability etc. of the constructed HI and then applied these assumptions as constraints in a quadratic programming model to obtain optimal fusion weight coefficient for each sensor. Therefore, the constructed HI has better monotonicity and fuses complete degradation information compared to any single sensor signal, and thus yields a better degradation model by fitting the constructed HI. As further researches, Song et al. [25] and Li et al. [26] proposed new fusion methods to construct HI that has nonlinear relations with sensor signals in complicated systems. These studies enriched the theories and construction methods of HI, and obtained fine prognosis results.

However, in real production processes, a group of machines in a manufacturing system are often under similar or dependent operational and environmental conditions (e.g., from the same engines in the same aircrafts), leading to dependent but possibly heterogeneous degradation profiles of each individual machine. For example, Fig. 1 illustrate degradation signals of a group of engines in an aircraft, where the data are originated from the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) [27]. Since these engines are under the same operational and environmental condition in the aircraft, their degradation processes present similar but not identical profiles, which indicates the dependency and heterogeneity of the engines. Such characteristics may contain key information for HI construction, degradation modeling and RUL prediction, but have been ignored in the above studies. Another challenge lies in that sensor signals for prognostic analysis may not be completely and appropriately gathered in practice for unexpected reasons, e.g., sensor aging, data reading errors, and wireless communication failures, leading to the existence of missing data. Currently, there is a lack of an effective method for degradation modeling and prognostics by considering dependency and heterogeneity among machines especially when missing data exist in sensor signals, though it remains an essential task to be solved.

There have been attempts in the literature that partially addressed the issues with missing data and considered similarities among different data sources. Models such as state-space model [28], naive bayes model [29] and kernel-based data-level fusion model [25] have been applied to time series data with missing values. Meanwhile, interpolation and extrapolation methods have been adopted to fill up incomplete datasets, such as linear interpolation, Kalman filtering [30], neural network [31], autoregressive (AR) predictor [32] and spline interpolation [33]. These methods can be regarded as single-task learning (STL), by which information is learned from only a single data source and thus fail to reflect domain-specific knowledge in related tasks, leading to poor model performance when a large number of data are missing. Instead, multi-task learning (MTL) is designed to share common information between related individual tasks [34]. Yu et al. [35] adopted a gaussian process from multiple tasks to characterize common features among the tasks, and then applied the maximum a posteriori based expectation-maximum (MAP-EM) algorithm to estimate model parameters. Such capacity in capturing dependency and heterogeneity from related tasks allows MTL to achieve a better performance than STL for handling incomplete datasets. For instance, Wan et al. [36] proposed a Bayesian MTL model to reconstruct the incomplete datasets for structural health monitoring, and validated that the model had better performance than traditional Bayesian STL model. Moreover, MTL has been applied to areas that are related to degradation process, e.g., Alzheimer’s disease (AD) neuroimaging [37], solar panel PV [38], aircraft engines [39] etc. However, these methods either ignore data fusion from multiple sensor signals to construct HI or divide degradation modeling and data fusion into two independent parts, leading to poor RUL prediction results.

To fill the research gap, this paper develops a data fusion-based multi-task learning method for degradation modeling and prognostics of machines based on multiple sensor signals. Specifically, we adopt a quadratic polynomial model to characterize the degradation processes of machines based on the HI constructed by a linear fusion model. Considering the dependency and heterogeneity among machines under the same operational and environmental condition, we innovatively integrate the MTL into the degradation model and assume the model parameters of all machines follow a common multivariate Gaussian distribution. For parameter estimation, we develop a unique quadratic optimization-integrated expectation-maximum (QOIEM) algorithm. In the QOIEM, quadratic programming, which considers the properties for HI construction including minimum fitting errors, minimum variance at the failure time and monotonicity, is integrated into the MAP-EM algorithm to update the parameters in the fusion model and degradation simultaneously. Finally, given the constructed HI and updated model parameters, the RUL of machines can be predicted correspondingly.

The proposed method addresses the challenges of degradation modeling and prognostics in the following three aspects: *First*, by fully considering the dependency and heterogeneity among machines under the same operational and environmental condition, this paper develops a data-level fusion based multi-task learning method for degradation modeling, which proposes a quadratic polynomial model to characterize the degradation status of machines and assume the model parameters of all machines follow a common multivariate Gaussian distribution. *Second*, a unique QOIEM algorithm is developed for parameter estimation, which can achieve accurate parameter estimation in the fusion model and the degradation model simultaneously and have satisfactory convergence performance in theory. *Third*, the proposed method is capable to make prognostic predictions of machines effectively when missing data exist in sensor signals, which has been validated through numerical experiments and real case study on C-MAPSS data [27]. Moreover, the proposed method has superior robustness for accurate RUL prediction under different ratios of missing data compared with state-of-the-art benchmarks.

The remainder of this paper is organized as follows. Section II introduces the proposed data fusion-based multitask learning method. Section III and IV present numerical experiments and a case study of aircraft gas turbine engines to evaluate the performance of the proposed method and compare it with existing benchmarks. Section V provides the conclusion and future work of this paper.

# Methodology

## Model formulation

We regard a group of units as our research objects, which are able to represent objects (e.g., machines) in degradation processes in real production cases. Following common settings in existing works [9], [11], we define the failure time of unit as the time when the underlying degradation status exceeds a predefined failure threshold :

, (1)

where the degradation status is a function of time . We consider units operate under similar or dependent operational and environmental condition.

Following [24], we present the definition of HI as follows. We denote as a vector of sensor signals collected from sensors of unit at time . We assume the degradation status can be recovered from a fusion function of multiple sensor signals with the contamination of white noises, i.e.,

(2)

The fusion function is a function of sensor signals . denotes the noise of unit at time , where . In particular, we define HI as a combination of multiple sensor signals by the fusion function:

, (3)

where denotes the HI of unit at time . To summarize, the relation between the HI and the underlying degradation status can be expressed as follows:

(4)

## Data fusion-based multitask learning for HI construction and degradation modeling

We consider a group of units under dependent operational and environmental conditions and therefore their degradation statuses have dependent profiles. Without loss of generality, we consider units with similar degradation statuses. For each unit () that has observed time points, we adopt the quadratic polynomial degradation model to describe its degradation status based on empirical studies and trend analysis of sensor signals [24]:

, (5)

where is the vector of quadratic basis function at time , with , and  is the vector of model parameters.

Since HI represents the health status of each unit, we design HI by setting the weight coefficient to linearly superimpose following [24], and thus can be represented as

,(6)

where is the HI at time . The extension to nonlinear fusion functions can be implemented by setting kernel functions or developing new models, which will be the focus of future study and omitted here. According to (4), (5) and (6), the relationship among the HI, multiple sensor signals and degradation status of unit can be expressed by a matrix form when there are observed time points as follows:

, (7)

where denotes the vector of , denotes the matrix of , denotes the matrix of , and denotes the vector of .

To characterize the dependency and heterogeneity among units, we integrate multitask learning into the above model structure, which aims at sharing knowledge gained in individual units. Specifically, we consider that the model parameter of each unit shares common knowledge (i.e., dependency and heterogeneity) among units and assume it follows a multivariate normal distribution:

, (8)

where and are mean vector and covariance matrix of the multivariate normal distribution and represent the dependency and heterogeneity between units respectively. According to [35], we assume the hyperprior distribution of and is a normal-inverse Wishart (IW) distribution, which is the conjugate prior for the multivariate Gaussian distribution

. (9)

Here, subjects to multivariate normal distribution in which , are corresponding mean vector and covariance matrix respectively and means the level of confidence in the prior distribution. subjects to IW distribution in which is the inverse scale matrix and is the degree of freedom with .

The data-fusion based MTL model can be concluded as follows:

1. and are generated based on (9);
2. For model parameter of each unit , we have from its distribution as , with ;
3. Given that denotes the matrix of observed sensor signals of unit , the HI can be constructed as and then the quadratic model is fitted by

Here, the estimation of the model parameters will be based on all historical units to achieve the multitask learning for HI construction and degradation modeling.

## QOIEM algorithm for parameter estimation

In this section, we propose a QOIEM algorithm to estimate parameters in the data-level fusion based multi-task learning model. There exist some challenges in parameter estimation, i.e., 1) in conventional models, weight coefficient and the degradation model parameters are interdependent. In other words, only one of them can be estimated while the others are obtained through additional engineering knowledge. In our proposed model, however, they are both unknown parameters; 2) contains multiple parameters, which makes it difficult to estimate parameters via derivation of the joint probability density function (PDF).

To address above challenges and estimate the required parameters, i.e., 1) weight coefficient in the linear fusion function, 2) mean vector and covariance matrix of the multivariate normal distribution for in the quadratic degradation model, and 3) the variance of the normal distribution for the noise term , we develop the QOIEM algorithm, and its pseudo codes are provided in Algorithm 1.

|  |  |
| --- | --- |
| **Algorithm 1** QOIEM algorithm | |
|  | **Input**: sensor signals of historical units containing observations until failure |
|  | **Output**: estimated parameters and |
| 1: | **Procedure**: |
| 2: | Initialize parameters ; *j* |
| 3: | **While** |
| 4: | Estimate weight coefficient in the fusion model through quadratic optimization in (12) |
| 5 | Estimate parameters in the degradation model through MAP-EM algorithm: |
| 6: | E step: obtain expected log-likelihood through (13) |
| 7 | M step: update and obtain through following (14) |
| 8: | **If** convergence is achieved, i.e., |
| 9: | **Stop** the loop |
| 10: | **Else** |
| 11: |  |

According to Algorithm 1, we utilize all of the historical units with their completely observed sensor signals () that contain their entire degradation processes until failure as the input of the algorithm, and the procedure of the QOIEM algorithm can be listed as follows, i.e., 1) initialize parameters at the beginning of the iterations, i.e., , where denotes the number of iteration, as , and , where is the identity matrix, 2) estimate the weight coefficient via quadratic optimization by considering the properties for HI construction, 3) expectation (E) step: first obtain the log-likelihood of joint distribution based on complete data, and then calculate the expected log-posterior over the posterior distribution function of latent variable , 4) maximum (M) step:update to through , and 5) check if the and satisfy the convergence condition. Repeat 2) - 5) until the convergence condition is satisfied.

**Quadratic programming to estimate**

Given at iteration , we estimate the weight coefficient in the linear fusion function by considering the following principles for HI construction:

1) *Minimum fitting errors*: The fitting errors between the degradation model and the constructed HI should be minimum [1], as we taking HI as the health status of a unit and it should be well fitted by the degradation model.

2) *Minimum variance at the failure time*: The variance of the constructed HIs for all historicalunits at their failure time should be minimum [24]. These units are under similar operational condition and failure mode, and thus we assume they have the same failure threshold.

3) *Monotonicity*: The constructed HI should follow a monotonic trend at all time points [24], as we suppose the health status of a unit should follow a monotonic trend and it comes to failure when its HI reaches the threshold.

According to these properties for HI construction, we can create the corresponding objective function and constraints. Considering the degradation statuses at time points that are closer to failure time have higher effect on RUL prediction, we set the weight matrix for different time points (see details in Appendix A), and write the weighted residual term from (7) as , where is the posterior expectation of the parameter in the degradation model and can be represented as a function with respect to.

To obtain the posterior expectation and the covariance of , first we calculate the log-likelihood of joint distribution of complete data containing and given based on all historical units, and it can be written as

, (10)

where

and denotes the normalization coefficient for the distribution. Then, the posterior expectation of at iteration can be estimated by setting and the covariance of can be estimated by the inverse of the Hessian matrix , which leads to

,

, (11)

where is the result of the previous iteration, and are the estimated posterior expectation and covariance. Therefore, the unbiased estimation of squared fitting errors can be calculated as .

Meanwhile, the unbiased estimation of the variance of the constructed HI at the failure time can be calculated as . Here,  and denotes the vector of sensor signals of unit at its failure time . Recall that represent sensor signals matrix that contains complete degradation process until failure of a historical unit , and thus its failure time is equal to its last observed time point . , where is the identity matrix and is a unit vector of all ones (see details in Appendix B). Therefore, given the above expressions, we obtain the objective function for HI construction to minimize the fitting errors and the variance at the failure time as follows:

,

where is the tuning coefficient that can be determined by k-fold cross-validation. The first constraint uses to normalize the weights of sensors, and is only composed by 1 and -1 to reflect the increasing or decreasing trend of the corresponding sensor signal. The second constraint ensures the monotonicity of HI with and that denotes the vector of all zeros. Then, we obtain the quadratic programming model for the estimation of the weight coefficient at iteration

, (12)

where

,

,

with

,

.

This standard form of quadratic programming (QP) problem can be solved by existing solvers, e.g., the qpOASES package in Python that provides a parametric active-set algorithm for QP [40], and thus the weight coefficient at iteration can be estimated as .

**EM algorithm to estimate**

After the weight coefficient is obtained, we adopt the EM algorithm to update the parameters in the degradation model. We take as the observed data and as the latent variable. In EM algorithm, we suppose to calculate the expected log-likelihood of complete data in the E step and update the parameters in the M step.

**The E step**: we are supposed to calculate the expected log-posterior based on the joint distribution and hyperprior distribution (i.e., (9)) function according to the theory of maximum a posteriori (MAP) estimation. Specifically, as unknown parameters, should be treated as random variables in its posterior distribution which is proportional to the product of and . Next, to obtain the expected log-posterior , we should transform the log-joint distribution to its expectation over the posterior distribution of .

The joint distribution of complete data can be obtained following (10) by taking as a known value . Next, the equations of the posterior expectation and covariance of latent variable can be calculated following (11) by taking as a known value . According to the estimated posterior expectation and covariance of latent variable , the expected log-posterior of complete data over the posterior distribution of at iteration is

, (13a)

where is the expectation of the log-joint distribution over theposterior distribution of

(13b)

and is log-hyperprior distribution (i.e., (9)) for

, (13c)

where

in which denotes the multivariate gamma function, is the dimension of the distribution and equal to 3 in our quadratic degradation model, and *const* has no relationship with that can be ignored in calculation. In Appendix C, we provide more details about the calculation of in E step.

**The M step**: can be updated through and obtain . The update equations are derived as follows

,

,

. (14)

We repeat the above procedures to update the model parameters until convergence is achieved. The convergence condition is , where is the root-mean-square error calculated by comparing and , and is the prespecified convergence threshold. It should be noted that for hyperparameters in our proposed algorithm, we set the parameters in hyperprior NIW distribution and following common settings. and are supposed to be consistent with the distribution of training data, and thus we take a batch of data for pre-training to detect their proper settings.

## Property of the QOIEM algorithm

In this subsection, we focus on studying the convergence property of the proposed QOIEM algorithm in theory. The ultimate purpose of investigating the convergence of QOIEM algorithm is to prove that the parameters can gradually approach to their true values after iterations. It is equivalent to prove that the log-likelihood function are able to be maximized in each iteration i.e., . We first prove , and then prove .Firstly, in MAP-EM procedure, is considered as a variable and according to the Bayes rule, can be transformed as

(15)

Then, we calculate the equation over the expectation of distribution like the E step. Note that and has no relationship with and thus we can obtain the following equation:

, (16)

where represents . Recall that the proof of convergence is equivalent to prove , and will be maximized in the M step and thus . Therefore, our goal is equivalent to prove the following inequality holds for arbitrary :

. (17)

Based on the expression of expectation, the left side of (17) can be transformed as

(18)

where is the PDF and thus . Then, according to Jensen inequality, the proof of (17) can be obtained as follows:

(19)

where the equal sign is established if and only if . After MAP-EM procedure, the variable is updated as and therefore .

Secondly, according to the restrictions of estimation procedure for weight coefficient in section C, the fitting error of HI (i.e., ) has been minimized in estimation. Based on (10) with respect to , the log-likelihood of joint distribution of complete data increases correspondingly after updating and therefore .

Finally, based on the discussion above, the convergence of the QOIEM algorithm is proved.

## RUL prediction

After model parameters are estimated, we predict the RUL for in-service units. Compared with the historical units that contain the entire degradation processes until failure used in parameter estimation, sensor signals of the in-service units are only available until certain time point before failure. For an in-service unit , we have the incomplete sensor signal matrix ,where is the number of available observations of unit . For RUL prediction, we first calculate the posterior distribution , where is the HI obtained by , as follows

,

where

,

, (20)

and denotes the last observed time of the in-service unit (see details in Appendix D).

Then, given the degradation status of the in-service unit at a specific time , i.e., , where , we obtain the distribution of the degradation status at time as , where and . In addition, we consider the variation of the failure threshold . The distribution of is written as , where the mean value and variance can be estimated by the sample mean and sample variance based on the HIs at the failure time of the historical units. According to the distribution of and , the probability that the degradation status exceeds the failure threshold at time after can be written as

, (21)

where is the RUL of the in-service unit and is the cumulative distribution function (CDF) of the standard normal distribution. Given , the conditional probability can be written as

. (22)

Finally, the RUL of the in-service unit is predicted by solving with variable .

# Numerical Experiments

In this section, we evaluate the effectiveness of our proposed method through numerical experiments. We conduct our proposed method on generated simulation dataset with degradation processes of multiple sensors, and then compare the RUL prediction results using our proposed method and the benchmark method under different levels of missing data for thorough evaluation.

## Data generation

We randomly generate 1000 units with a linear degradation path, i.e., the underlying degradation process for each unit is

(23)

where the random effect parameter  is generated from a bivariate normal distribution

. (24)

We describe the hyperprior distribution of and by a normal-inverse Wishart distribution

. (25)

Note that the probability of is less than , which is thus negligible. If any with is generated, we reject the sample and generate a new one to ensure that the underlying degradation process of any unit is increasing. The failure threshold is set to be and we record the true failure time as according to .

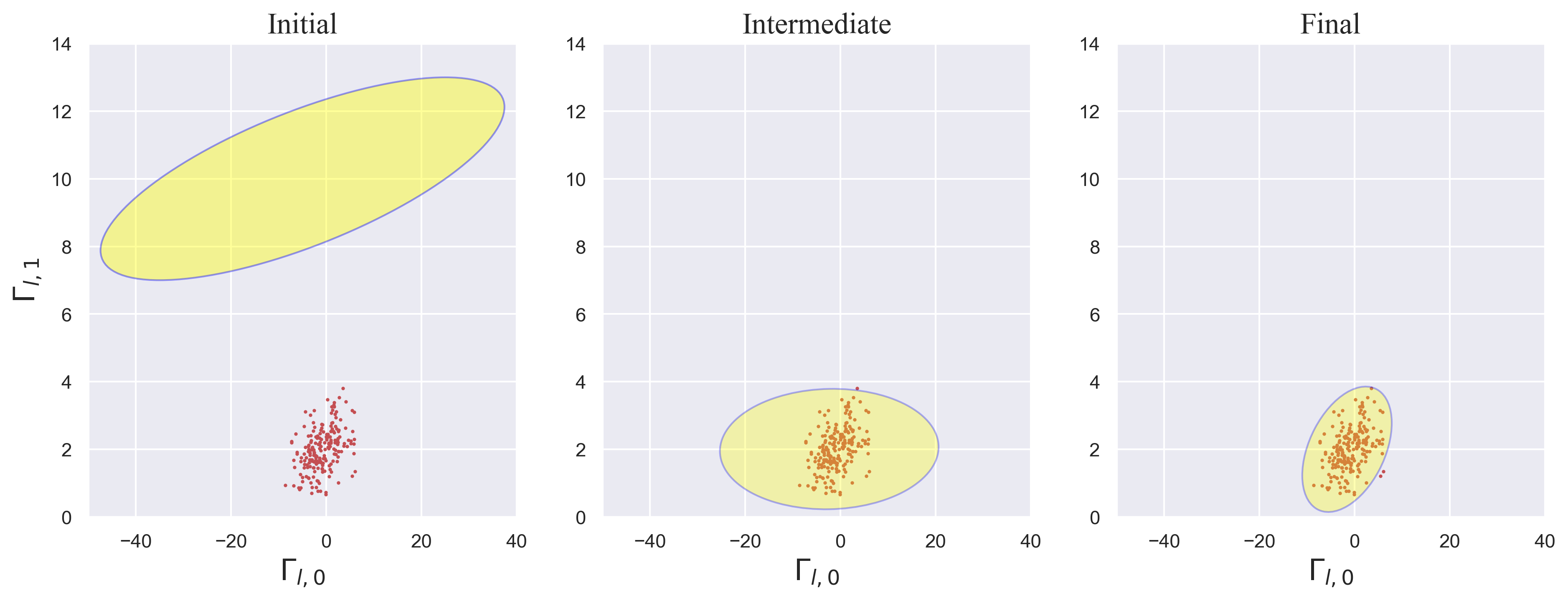


Fig. 4. The convergence process of the QOIEM algorithm by taking the parameter as a demonstration. (Note: is a two-dimensional vector whose true values for all of the historical units are displayed on a plane in the form of red scattered points. The yellow ellipse represents the confidence ellipse of the multivariate normal distribution and its mean and covariance are being updated by the QOIEM algorithm. The plots from left to right use the estimated and at the initial iteration, an intermediate iteration, and the final iteration to draw the ellipse respectively.)

We generate the true HI by adding a random noise to the degradation process , where follows a normal distribution . Each unit has four sensors with the true value of fusion coefficients . Four sensor signals are randomly generated by

,

,

,

, (26)

where , and . Here, the sensor signal is generated from the true HI subtracting the sensor signals and to satisfy . Since the fusion parameter , the sensor signal is not related to the underlying degradation process. Fig. 2 presents an example of the sensor signals and true HIs for three randomly generated units.

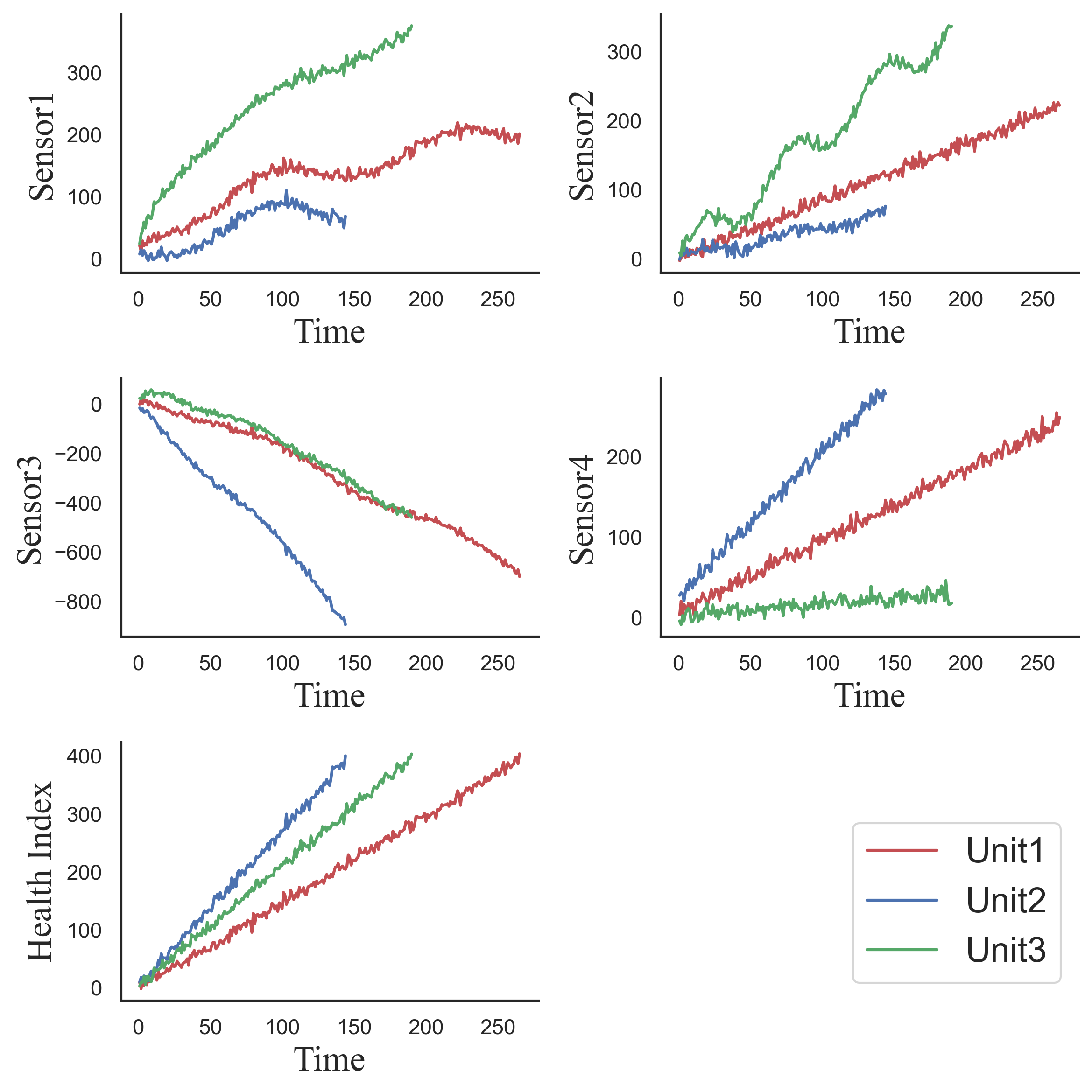


Fig. 2. An example of the sensor signals and true HIs for three randomly generated units.

## Model performance of ideal scenario

We randomly select part of generatedunits as the historical units, and the remaining units as the in-service units. Our aim is to use the historical units to fit our proposed model and predict the RUL of the in-service units for model evaluation.

To explore the required number of historical units for stable and accurate parameter estimation in our proposed model, we estimate the weight coefficient using the QOIEM algorithm based on different number of historical units. Specifically, we randomly select historical units to estimate and repeat the above procedure 50 times to obtain the mean and standard deviation of our estimations. The results are shown in Fig. 3, where X-axis indicates the number of units involved in parameter estimation. Black dashed horizontal lines represent the true values of of each sensor. Red solid lines represent the mean values of the estimated . Blue dotted lines represent corresponding 95% confidence intervals of the estimated . As the number of units involved in parameter estimation increases, the estimated and (red solid lines) are getting closer to their true values (black dashed horizontal lines) and their variances (blue dotted lines) keep decreasing. Therefore, to obtain the stable and accurate estimation of weight coefficient , it is recommended that the number of historical units exceeds 40.

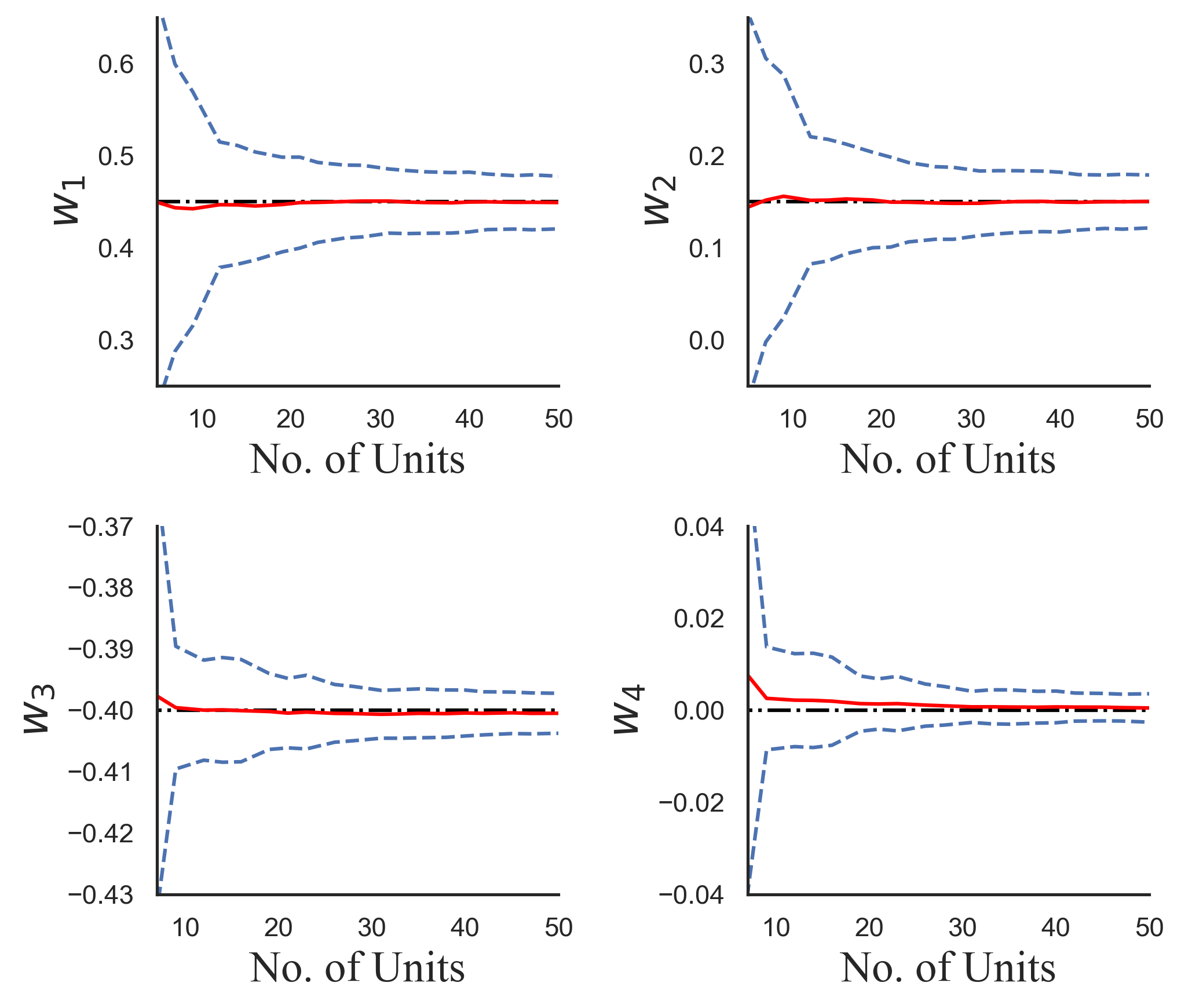


Fig. 3. The estimated fusion coefficients by considering different numbers of units. (Note: X-axis indicates the number of units involved in estimation. Black dashed horizontal lines represent the true values of coefficients of each sensor. Red solid lines represent the mean values of the estimated fusion coefficients by considering different number of units. Blue dotted lines represent the corresponding 95% confidence intervals.)

In our experiments, we choose 500 units from 1000 generated units as the historical units, and the remaining 500 units as the in-service units. Based on the historical units, we use the QOIEM algorithm to estimate the parameters of the degradation model and the weight coefficient in the linear fusion function.

We present the convergence process of the proposed QOIEM algorithm in Fig. 4 by taking the parameter as a demonstration. Recall that the true value of is a two-dimensional vector generated by (19) and (20) for each historical unit . We display the true values of for all of the 500 historical units on a plane in the form of red scattered points. In the QOIEM, we assume and its mean and covariance are updated by the EM algorithm. The yellow ellipses show the confidence ellipses of the multivariate normal distribution using the estimated and at the initial iteration, an intermediate iteration, and the final iteration respectively. It can be seen that the confidence ellipse gradually approaches the true values of as the iteration goes and the final confidence ellipse (yellow ellipse) is dovetail with the true values of (red scatters), which indicates superior convergence performance of the proposed QOIEM algorithm and high accuracy of the estimated parameters. Table Ⅰ shows the true values and estimated values of the model parameters. The estimated values are very close to the true ones, which verifies the efficiency of the proposed QOIEM algorithm.

TABLE I

Comparison of model parameters in numerical experiments.

|  |  |  |
| --- | --- | --- |
| Parameters | True value | Estimated value |
|  |  |  |
|  |  |  |
|  |  |  |
|  | 25 | 24.65 |

After estimating model parameters, we predict the RUL of the in-service units to evaluate our model performance. We define the prediction error of the in-service unit as the absolute difference between the predicted RUL and the true RUL of the in-service unit divided by its true failure time :

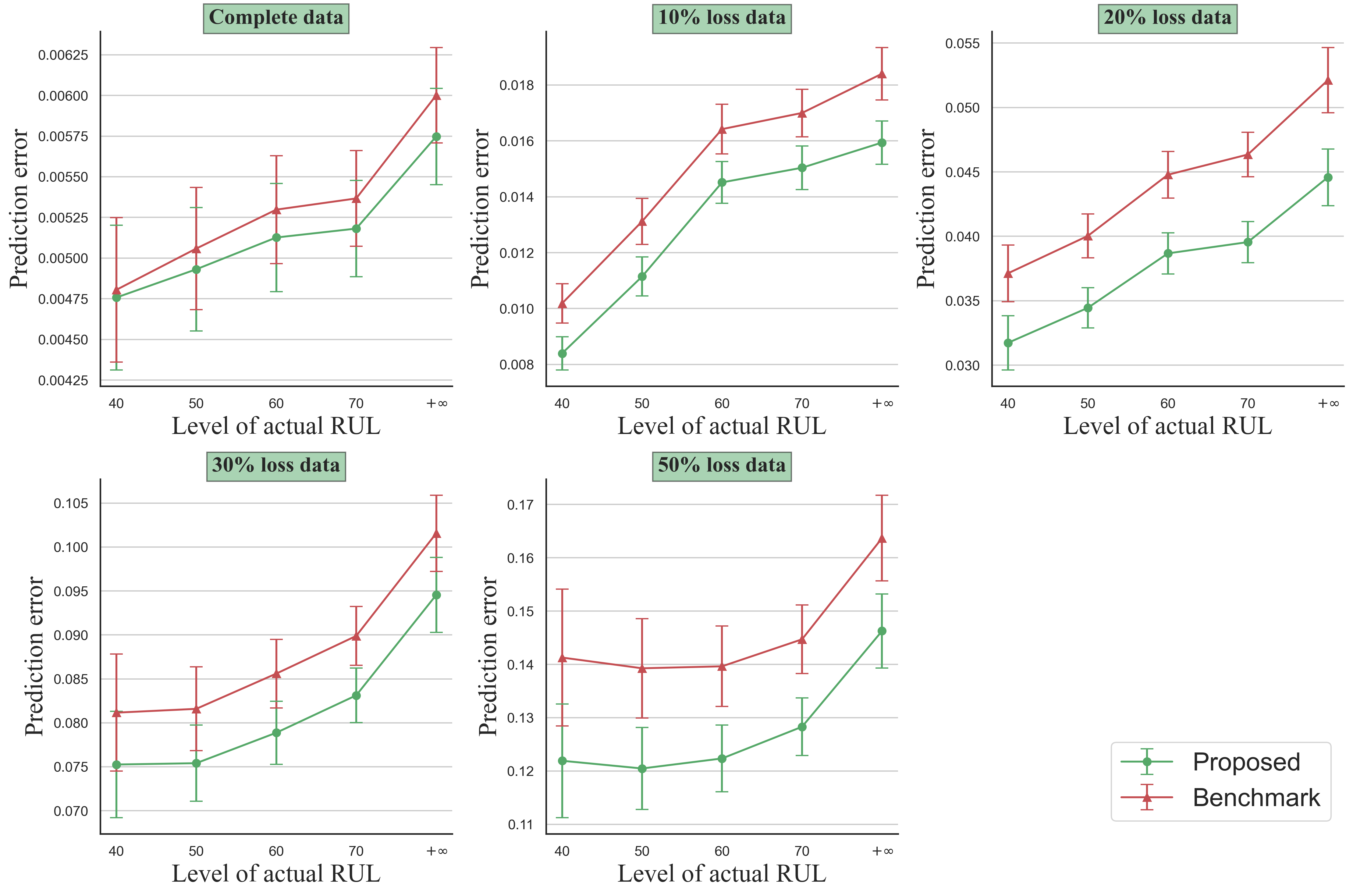


Fig. 6. Comparison of RUL prediction results by the proposed and benchmark methods under various ratios of missing data. (Note: X-axis represents the level of actual RUL, Y-axis represents the mean value of the RUL prediction errors, and the vertical sticks represent corresponding standard errors.)

. (27)

We calculate , with , for all in-service units, and then obtain the corresponding mean values and the standard errors under different levels of actual RUL, where the standard error is defined as the one standard deviation of the prediction errors divided by the square root of the number of units. To show the effectiveness of the proposed method, we compare it with a benchmark method in [1], which constructs a data fusion-based HI for degradation modeling but ignores dependency and heterogeneity among units by assuming the units are independent with each other. Fig. 5 shows the comparison of predicted RUL by the proposed and benchmark methods under different levels of actual RUL. X-axis represents the level of actual RUL, for which level “40” refers to the in-service units that have actual RUL less than or equal to 40 and level “” refers to all in-service units. Y-axis represents the mean value of the RUL prediction errors, and the vertical sticks represent corresponding standard errors. Here, the standard error is the one standard deviation of the RUL prediction errors divided by the square root of the number of units. We can find from this figure that the proposed method (green bars) has consistently lower RUL prediction errors than the benchmark method (red bars) under all levels of actual RUL. The standard errors of the proposed and benchmark methods both keep small under all levels of actual RUL.

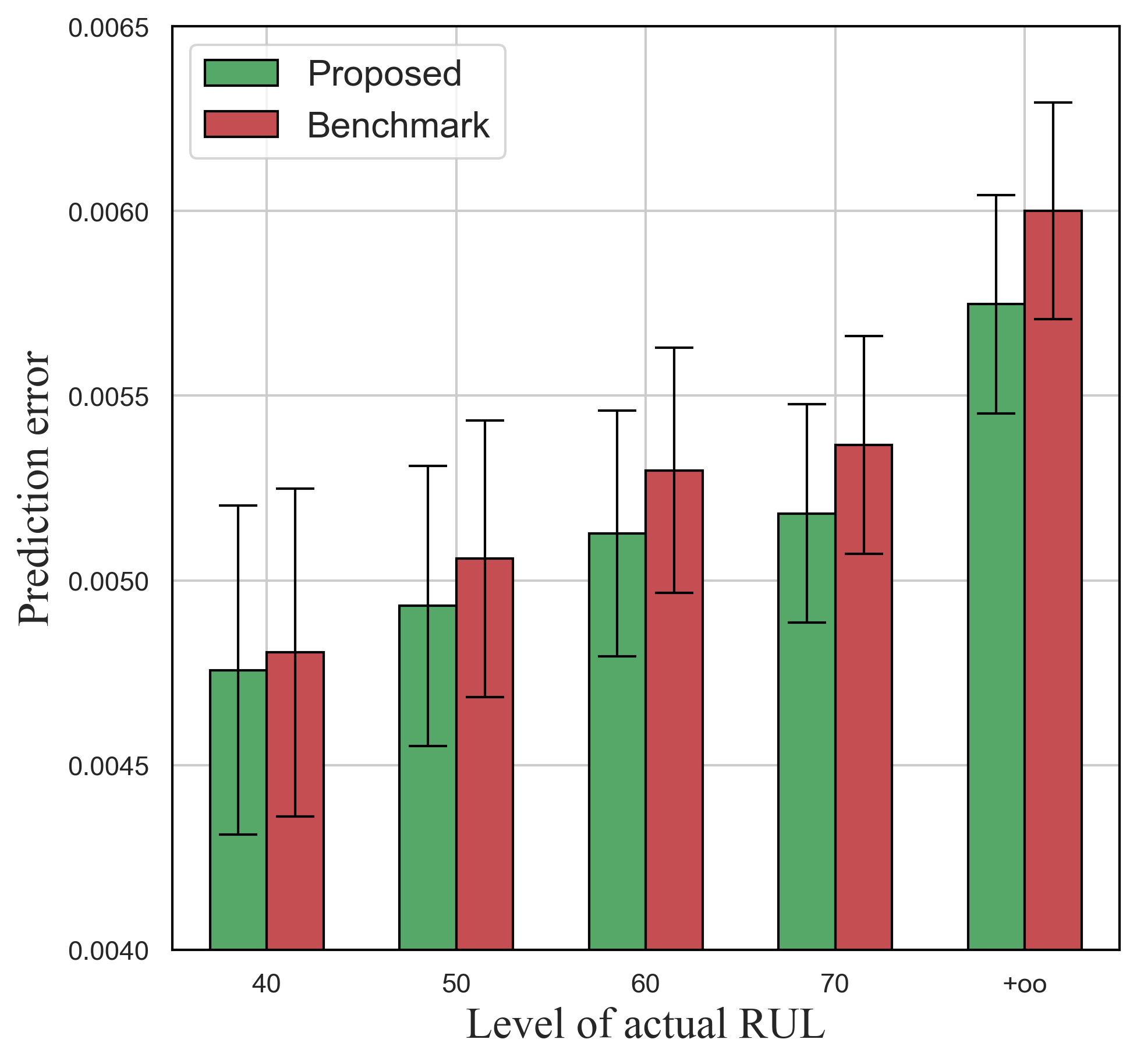


Fig. 5. Comparison of RUL prediction results by the proposed and benchmark methods under different levels of actual RUL. (Note: X-axis represents the level of actual RUL, Y-axis represents the mean value of the RUL prediction errors, and the vertical sticks represent the corresponding standard errors).

## Model performance of the scenarios with missing values

To evaluate the effectiveness of the proposed method for RUL prediction of units with missing values in sensor signals, we generate several scenarios with different ratios of missing values. Specifically, we randomly select a certain number of data from sensor signals of the historical units and assume the selected data are missing. Fig. 6 presents the RUL prediction results by the proposed and benchmark methods for scenarios with 10%, 20%, 30% and 50% missing values, respectively. The proposed method (green lines) has consistently smaller RUL prediction errors than the benchmark method (red lines) for all scenarios with missing values, because the proposed method considers the dependency and heterogeneity between units and shares the common information among units for handling missing data. Besides, the accuracy of the proposed method is relatively high and its superiority is more significant than the benchmark method when the ratio of missing data is larger than 20%, which verifies the efficiency and robustness of the proposed method.

# Case Study

## Dataset description

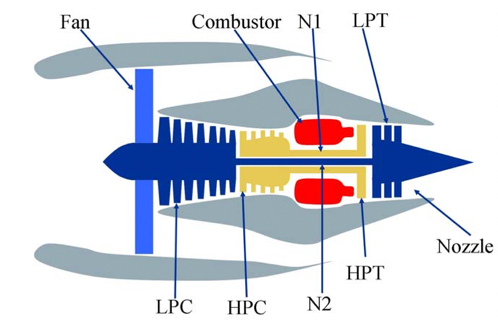


Fig. 7. Simplified diagram of engine simulated by C-MAPSS [39].

In real case study, we choose the degradation-based dataset of aircraft turbofan engines to validate our proposed model and algorithms. The dataset is originated from the C-MAPSS [27], which has been adopted in many PHM related researches, such as [1] and [41].

To demonstrate the C-MAPSS data in more details, the schematic diagram of an aircraft turbofan engine is displayed in Fig. 7, and the detailed information about the 21 available degradation sensor signals of aircraft turbofan engines under the same operational condition and failure mode is shown in Table Ⅱ. The dataset includes 100 historical units (i.e., ) and 100 in-service units (i.e., ). The data of historical units include the complete observations of degradation processes for each unit until failure and there are 20,631 observations (i.e., ) in total. Meanwhile, the data of in-service units only have the observations up to some time point before failure and there are 13,096 observations (i.e., ) in total. Each observation contains 21 sensor signals from different monitors in aircraft turbofan engines. We use the historical units for model construction and parameter estimation, and the in-service units for RUL prediction. We compare the predicted RUL to the provided true RUL for model validation. The preprocessing procedure are implemented as in [1]: Considering the degradation signals should exhibit a consistent increasing (or decreasing) trend, 11 candidate sensors are selected out of 21 sensors, including T24, T50, P30, Nf, Ps30, Phi, NRf, BPR, htBleed, W31, and W32. Next, all sensor signals are preprocessed by conducting z-score normalization and logarithm transformation.

TABLE Ⅱ

Detailed information of the 21 sensors in the case study.

|  |  |  |
| --- | --- | --- |
| 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 | rpm 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 |

## Results and comparison

TABLE Ⅲ

Estimated weight coefficient in the fusion model for each sensor.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | T24 | T50 | P30 | Nf | Ps30 | phi |
| Value | 0.1012 | 0.1126 | -0.0831 | 0.0094 | 0.1407 | -0.1293 |
| Name | NRf | BPR | htBleed | W31 | W32 |  |
| Value | 0.0276 | 0.1176 | 0.0873 | -0.0944 | -0.0965 |  |

TABLE Ⅳ

Estimated in the degradation model.

|  |  |
| --- | --- |
| Parameter | Value |
|  |  |
|  |  |
|  | 0.0090 |

We first estimate model parameters using data of historical units. Table Ⅲ lists the estimated weight coefficient in the fusion model for each sensor. The positive and negative of the weight reflects the increasing or decreasing trend of each sensor, and the sum of the absolute value of each sensor weight is 1. The estimated value of parameter in the degradation model is given in Table Ⅳ.

Fig. 8 shows the sensor signals and the constructed HI of a randomly selected historical unit. In the figure, the data of each sensor is fitted by the quadratic polynomial degradation model that proposed in [24]:

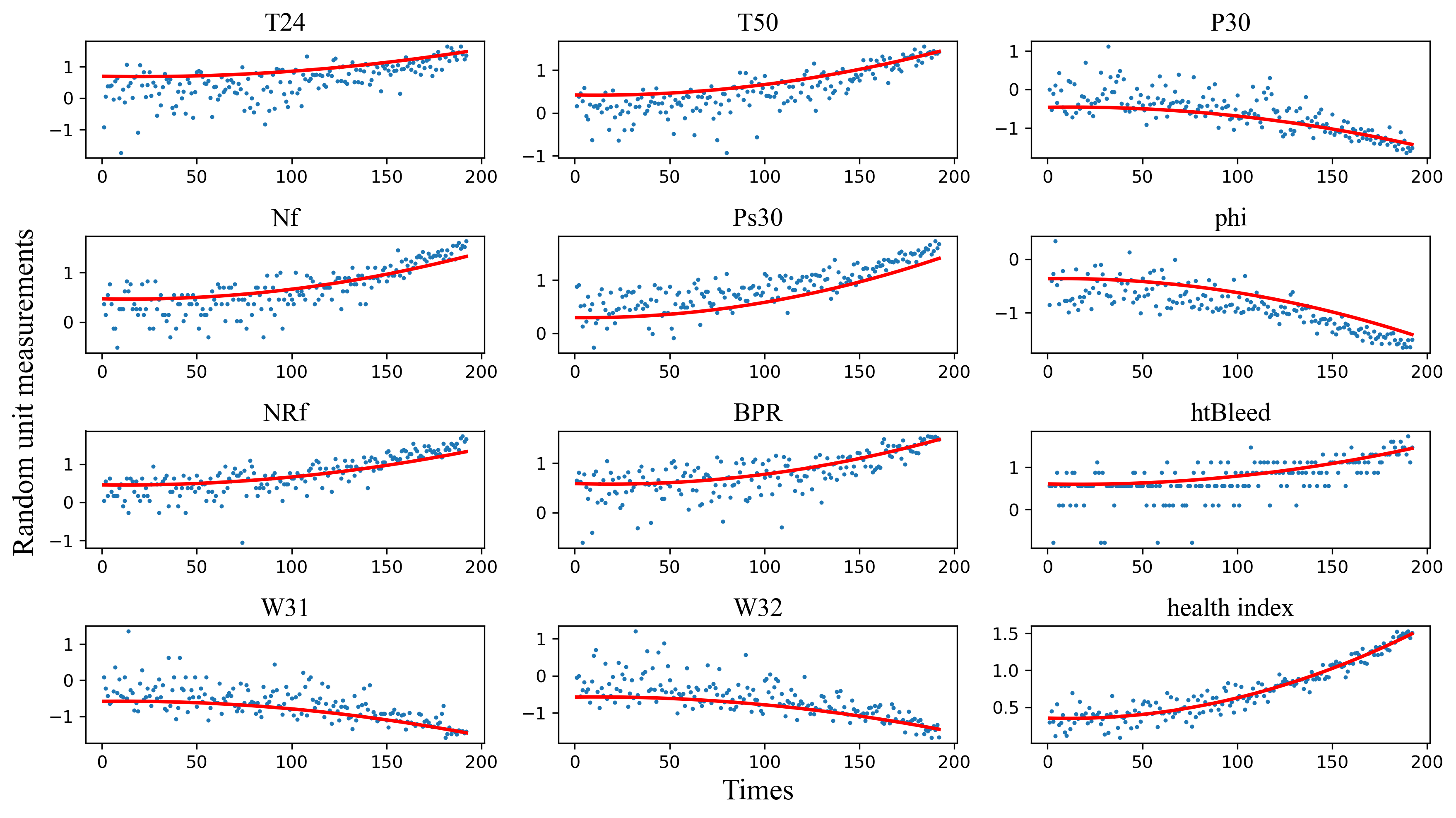


Fig. 8. Sensor signals and the constructed HI of a randomly selected historical unit. (Note: the blue dots are sensor signals and the constructed HI. The red solid lines are the fitted degradation statuses.)

(28)

where and represent the index of unit and sensor respectively, are random-effect parameters and is the noise term. is the vector of model parameters and follows a multivariate Gaussian distribution . Parameters of the degradation model are estimated through EM algorithm similar to (9)-(14) but without data fusion. Next, to characterize the degradation process of each single sensor signal of historical unit , we calculate the posterior distribution following (15) and obtain the degradation curves of each single sensor signal shown in figure, where denotes the matrix of . It can be seen that the HI has better monotonicity and smaller fluctuations than any single sensor signal, which indicates the proposed method for HI construction is reasonable and effective.

Next, we predict the RUL of the in-service units and compare RUL prediction errors of the proposed method and benchmark method. Fig. 9 shows the comparison of the RUL prediction errors using our proposed method (green bars) and the benchmark method (red bars) under different actual RUL levels. It can be seen that our proposed method has consistently smaller RUL prediction errors than the benchmark method under all RUL levels. It indicates that the proposed method can make better use of the common information among units and achieve more accurate parameter estimation results through the QOIEM algorithm.

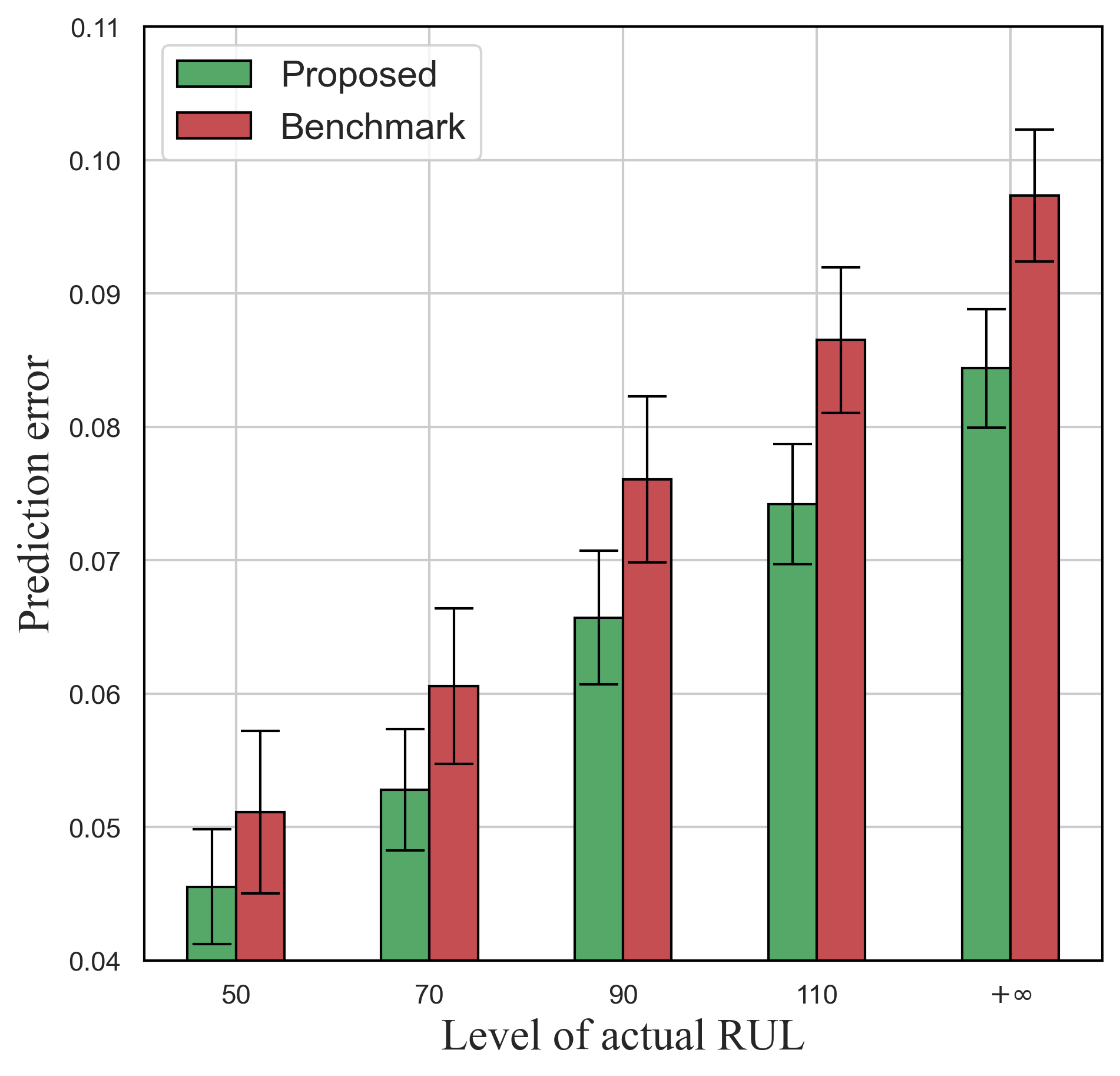


Fig. 9. Comparison of RUL prediction results by the proposed and benchmark methods under different levels of actual RUL. Here X-axis represents the level of actual RUL, Y-axis represents the mean value of the RUL prediction errors, and the vertical sticks represent corresponding standard errors.

TABLE Ⅴ

RUL prediction errors based on HI (the proposed method) and each single sensor signal.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | T24 | T50 | P30 | Nf | Ps30 | phi |
| Error | 0.117 | 0.099 | 0.104 | 0.151 | 0.093 | 0.107 |
| Name | NRf | BPR | htBleed | W31 | W32 | HI |
| Error | 0.210 | 0.091 | 0.115 | 0.127 | 0.120 | 0.084 |

To evaluate the effectiveness of constructing the HI, we further compare the proposed method with the benchmarks that use each single sensor signal for degradation modeling and prognostics respectively (i.e., (23)). These benchmarks ignore data fusion without the construction of HI. Table Ⅴ shows the RUL prediction errors based on each single sensor signal and the constructed HI. It can be seen that the benchmarks that only use a single sensor signal induce relatively larger RUL prediction errors than the proposed method. Meanwhile, we find the sensor with smaller RUL prediction error has higher weight coefficient in the fusion model of our proposed method (see Table Ⅲ), which means our fusion model for HI construction and its result are credible.

## Model performance for handling missing data

In order to reflect various modes of missing data in real engineering cases, we design three different modes of signal losses to validate the performance of the proposed method as follows:

(A) **Drop data point**: Randomly select a certain number of signal data in historical units and assume they are missing;

(B) **Drop time step**: Randomly select a certain number of time steps in historical units, and assume all sensor signals at the selected time steps are missing. Note that considering engineering practice, the last observed time steps of the units are not dropped.

(C) **Compound drop method**: Combine the above modes A and B. First, we drop a certain ratio (30% in our experiments) of signal data as mode A, and then we drop time steps as mode B by setting various drop ratios.

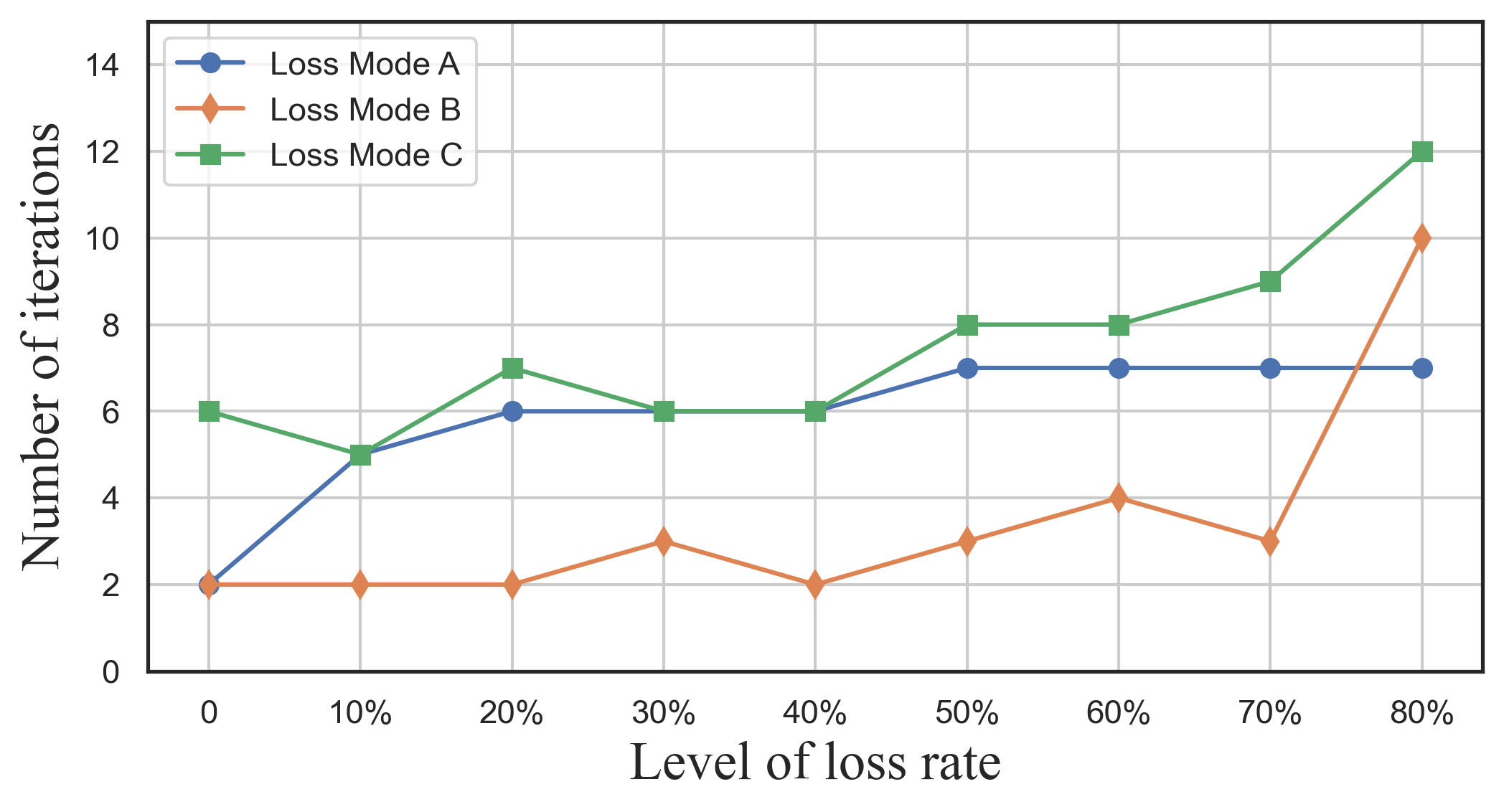


Fig. 10. Number of iterations in the QOIEM algorithm under loss mode A, B and C with the ratios of missing data from 10% to 80%.

Fig. 10 presents the number of iterations in the QOIEM algorithm for parameters estimation under loss mode A, B and C with the ratios of missing data from 10% to 80%. The number of iterations can reflect the convergence performance in our QOIEM algorithm, where we set as the breakout condition. As is shown in the figure, the number of iterations keeps low when the missing rate is under 60% and raises as the missing rate increases but keeps no more than 12, which indicates under different levels of missing data, our proposed QOIEM algorithm can still have great convergence. Moreover, the iteration times of loss mode B (orange line) is relatively low, which is owing to our quadratic degradation model can well fit the degradation sensor signals and thus some missing time steps have little impact on the parameter estimation in our proposed method.

TABLE VI

RUL prediction errors based on HI (the proposed method) and each single sensor signal under loss mode a.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Missing rate: 0.1** | | | | | | |
| Name | T24 | T50 | P30 | Nf | Ps30 | phi |
| Error | 0.122 | 0.100 | 0.107 | 0.151 | 0.095 | 0.108 |
| Name | NRf | BPR | htBleed | W31 | W32 | HI |
| Error | 0.203 | 0.096 | 0.151 | 0.133 | 0.126 | 0.092 |
| **Missing rate: 0.2** | | | | | | |
| Name | T24 | T50 | P30 | Nf | Ps30 | phi |
| Error | 0.147 | 0.110 | 0.122 | 0.164 | 0.098 | 0.110 |
| Name | NRf | BPR | htBleed | W31 | W32 | HI |
| Error | 0.219 | 0.110 | 0.182 | 0.152 | 0.140 | 0.105 |
| **Missing rate: 0.3** | | | | | | |
| Name | T24 | T50 | P30 | Nf | Ps30 | phi |
| Error | 0.170 | 0.117 | 0.142 | 0.176 | 0.113 | 0.115 |
| Name | NRf | BPR | htBleed | W31 | W32 | HI |
| Error | 0.217 | 0.130 | 0.202 | 0.171 | 0.149 | 0.111 |
| **Missing rate: 0.5** | | | | | | |
| Name | T24 | T50 | P30 | Nf | Ps30 | phi |
| Error | 0.227 | 0.143 | 0.156 | 0.208 | 0.139 | 0.127 |
| Name | NRf | BPR | htBleed | W31 | W32 | HI |
| Error | 0.288 | 0.182 | 0.244 | 0.244 | 0.191 | 0.134 |

Table VI shows RUL prediction errors based on HI (the proposed method) and each single sensor signal with missing data under loss mode A. It can be seen that by conducting data fusion to construct HI, the RUL prediction errors are able to be decreased compared with each single sensor signal because the constructed HI contains more information for degradation modeling. Meanwhile, as the missing rate increases, the accuracy of RUL prediction by HI is more prominent than other results predicted by single sensor signals, which indicates that our proposed method gains more robustness for accurate RUL prediction under different ratios of missing data. The results under loss modes B and C are similar to loss mode A, and thus omitted here.

Fig. 11 shows the comparison of RUL prediction results by the proposed and benchmark methods under loss modes A, B and C with different levels with missing data from 10% to 80%. It can be seen that under these situations, the proposed method (green lines) still has lower RUL prediction errors than the benchmark method (red lines) because the proposed method shares the common information among units to effectively handle the issue of missing data. Specifically, a) under loss mode A, the RUL prediction errors of the two methods both increase significantly with the increasing loss rate of missing data. b) Under loss mode B, the RUL prediction errors of the two methods have no significant change. The possible explanation is that the quadratic model fits the data well, and thus the accurate parameter estimation of the quadratic model can still be achieved with the absence of some time steps. C) Under loss mode C, because the proposed method captures the dependency and heterogeneity among units under the same operational condition and failure mode, the RUL prediction errors of the proposed method keep stable when the loss rate of missing data is increasing. On the contrary, the RUL prediction errors of the benchmark method increases with the increasing loss rate of missing data. In conclusion, the above analysis verifies the robustness of the proposed method when missing data exist.

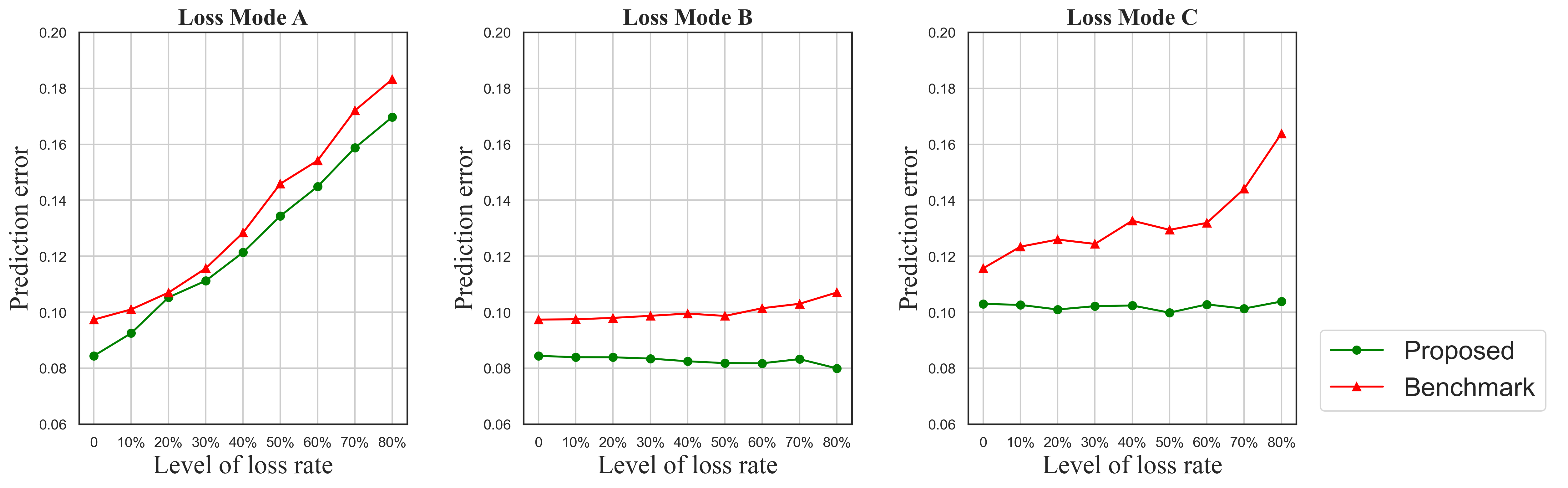


Fig. 11. Comparison of RUL prediction results by the proposed and benchmark methods under loss modes A, B and C and with different levels of missing data from 10% to 80%.

# Conclusion

In real production processes, prognosis of machines plays an important role in the maintenance of system stability. However, multiple sensor signals with missing data pose challenges to RUL prediction of machines. To address this issue, this paper proposes a data-level fusion based multi-task learning method for prognostics of machines. Specifically, we adopt the quadratic polynomial model to fit the degradation processes of machines based on HIs that are constructed by linear fusion model. Considering both the dependency and heterogeneity among machines under the same operational condition and failure mode, we innovatively integrate the MTL into the polynomial degradation model, and develop a unique QOIEM algorithm for parameter estimation. Finally, given the constructed HI and degradation model parameters, the RUL can be predicted correspondingly. We conduct the proposed method to numerical experiments and real case study on C-MAPSS data for model validation. As a result, the proposed method has advantages of superior robustness for accurate RUL prediction under different ratios of missing data compared with state-of-the-art benchmarks. In addition, our method has great convergence for parameter estimation.

For future work, we considered two related topics. *First*, the HI in the proposed method is constructed by a linear fusion model, future works will focus on non-linear fusion for systems with high complexity. *Second*, as our method only consider one type of failure mode and operational condition for RUL prediction, in future research we will consider predicting RUL under multiple failure modes and operational conditions.

# Appendix A

# Setting time-step weight matrix

Recall that for the formula , represents the vector of residual term of unit , where residual term is the difference between the constructed HI and the degradation model at time . The degradation statuses at time points that are closer to failure time has higher effect on RUL prediction. Therefore, we set the weight for residual term , and should follows

. (A1)

In addition, each historical unit shares the same importance so that should be normalized:

. (A2)

We assume the follows an arithmetic series and can be presented as:

. (A3)

Finally, we obtain the weight coefficient matrix .

# Appendix B

# The variance of HIs in the failure threshold

We present the calculation of the variance in the failure threshold of the constructed His based on historical units. Recall that and denotes the vector of sensor signals of unit at its failure time . Recall that represents sensor signal matrix that contains complete data of degradation process until failure of the historical unit , and thus its failure time is equal to its last observed time point . The estimated failure threshold of the constructed HI (called failure threshold below) can be considered as . Then, the mean of the failure threshold can be expressed as , where denotes the column vector with all ones. The mean of squared failure threshold can be expressed as .Therefore, the unbiased estimation of the variance in the failure threshold is calculated as

. (B1)

(B1) can be transformed to , in which and is the identity matrix.

# Appendix C

# The expected log- posterior

In the E step of the QOIEM algorithm, we calculate the expected log-posterior based on the joint distribution and hyperprior distribution (i.e., (9)) .

Note that the joint distribution is proportional to according to the Bayes rule. is the combination of multiple Gaussian distributions , with , thus the distribution of observed data is

. (C1)

is the probability density function (PDF) of multivariate Gaussian distribution , which is

, (C2)

where is the dimension of the distribution, which is equal to 3 in our quadratic degradation model. Then, the joint distribution of complete data can be written as (10) and

. (C3)

As is discussed in above text, maximum a posterior (MAP) requires the log-joint distribution  and log- prior distribution (i.e., (13c)). can be transformed as

, (C4a)

where

(C4b)

For the calculation of expected log-likelihood , according to the posterior distribution with expectation and covariance given in (11) and use to replace as a known variable, we can obtain the expectation of expressions with in (C4b):

(C5)

and thus obtain in (13b).

Finally, the  can be acquired by adding (i.e., (13b)) and (i.e., (13c)).

# Appendix D

# The posterior distribution

Recall that is the parameter of the degradation model and is the constructed HI of the in-service unit . We have the prior distribution of , i.e., . Since with and the likelihood function of can be given based on the conditional distribution , the posterior distribution of is obtained as

(D1)

where

.

Therefore, the posterior distribution is a multivariate normal distribution with mean vector and covariance matrix .

References

[1] 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.

[2] X. Zhang, L. Tang, and J. Decastro, “Robust fault diagnosis of aircraft engines: a nonlinear adaptive estimation-based approach,” *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 3, pp. 861–868, May 2013.

[3] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, “Residual life predictions from vibration-based degradation signals: a neural network approach,” *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 694–700, Jun. 2004.

[4] A. Rai and S. H. Upadhyay, “Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering,” *Mech. Syst. Signal Process.*, vol. 93, pp. 16–29, Sep. 2017.

[5] Hoke, A. Brissette, D. Maksimović, A. Pratt, and K. Smith, “Electric vehicle charge optimization including effects of lithium-ion battery

degradation,” in *Proc. IEEE Veh. Power Propulsion Conf.*, Sep. 2011,

pp. 1–8.

[6] F. Yang, D. Wang, Y. Zhao, K.-L. Tsui, and S. J. Bae, “A study of the relationship between coulombic efficiency and capacity degradation of commercial lithium-ion batteries,” *Energy*, vol. 145, pp. 486–495, Feb. 2018.

[7] P. Wang, B. D. Youn, and C. Hu, “A generic probabilistic framework for structural health prognostics and uncertainty management,” *Mech. Syst. Signal Process.*, vol. 28, pp. 622–637, Apr. 2012.

[8] M. S. Kan, A. C. C. Tan, and J. Mathew, “A review on prognostic techniques for non-stationary and non-linear rotating systems,” *Mech. Syst. Signal Process.*, vol. 62–63, pp. 1–20, Oct. 2015.

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

[10] 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, Apr. 2017.

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

[12] N. Gebraeel, A. Elwany, and J. Pan, “Residual life predictions in the absence of prior degradation knowledge,” *IEEE Trans. Reliab.*, vol. 58, no. 1, pp. 106–117, Mar. 2009.

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

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

[15] M. Hajiha, X. Liu, and Y. Hong, “Degradation under dynamic operating conditions: modeling, competing processes and applications,” *J. Qual. Technol.*, vol. 53, no. 4, pp. 347–368, Aug. 2021.

[16] X. Fang, N. Z. Gebraeel, and K. Paynabar, “Scalable prognostic models for large-scale condition monitoring applications,” *IISE Trans.*, vol. 49, no. 7, pp. 698–710, Jul. 2017.

[17] X. Li, J. Qian, and G. Wang, “Fault prognostic based on hybrid method of state judgment and regression,” *Adv. Mech. Eng.*, vol. 5, p. 149562, Jan. 2013.

[18] R. Moghaddass and M. J. Zuo, “An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process,” *Reliab. Eng. Syst. Saf.*, vol. 124, pp. 92–104, Apr. 2014.

[19] R. Gouriveau and N. Zerhouni, “Connexionist-systems-based long term prediction approaches for prognostics,” *IEEE Trans. Reliab.*, vol. 61, no. 4, pp. 909–920, Dec. 2012.

[20] E. G. Julie and S. T. Selvi, “Development of energy efficient clustering protocol in wireless sensor network using neuro-fuzzy approach,” *Sci. World J.*, vol. 2016, pp. 1–8, 2016.

[21] R. E. Silva *et al.*, “Proton exchange membrane fuel cell degradation prediction based on adaptive neuro-fuzzy inference systems,” *Int. J. Hydrog. Energy*, vol. 39, no. 21, pp. 11128–11144, Jul. 2014.

[22] J. Xu, Y. Wang, and L. Xu, “PHM-oriented integrated fusion prognostics for aircraft engines based on sensor data,” *IEEE Sens. J.*, vol. 14, no. 4, pp. 1124–1132, Apr. 2014.

[23] F. Wang, J. Du, Y. Zhao, T. Tang, and J. Shi, “A deep learning based data fusion method for degradation modeling and prognostics,” *IEEE Trans. Reliab.*, vol. 70, no. 2, pp. 775–789, Jun. 2021.

[24] 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.

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

[26] Z. Li, J. Wu, and X. Yue, “A shape-constrained neural data fusion network for health index construction and residual life prediction,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 5022–5033, Nov. 2021.

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

[28] X.-S. Si, “An adaptive prognostic approach via nonlinear degradation modeling: application to battery data,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5082–5096, Aug. 2015.

[29] S. S. Y. Ng, Y. Xing, and K. L. Tsui, “A naive bayes model for robust remaining useful life prediction of lithium-ion battery,” *Appl. Energy*, vol. 118, pp. 114–123, Apr. 2014.

[30] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, “Kalman filtering with intermittent observations,” *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[31] S. Mohamed and T. Marwala, “Neural network based techniques for estimating missing data in databases,” *16th Annu. Symp. Pattern Recognit. Assoc. South Afr.*, Jan. 2005.

[32] O. Anava, E. Hazan, and A. Zeevi, “Online time series prediction with

missing data,” in *Proc. ICML*, 2015, pp. 2191–2199

[33] R. J. Aliaga, “Real-time estimation of zero crossings of sampled signals for timing using cubic spline interpolation,” *IEEE Trans. Nucl. Sci.*, vol. 64, no. 8, pp. 2414–2422, Aug. 2017.

[34] R. Caruana, “Multitask learning,” *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997.

[35] K. Yu, V. Tresp, and A. Schwaighofer, “Learning Gaussian processes

from multiple tasks,” in *Proc. 22nd Int. Conf. Mach. Learn.*, Bonn,

Germany, 2005, pp. 1–8.

[36] H.-P. Wan and Y.-Q. Ni, “Bayesian multi-task learning methodology for reconstruction of structural health monitoring data,” *Struct. Health Monit.*, vol. 18, no. 4, pp. 1282–1309, Jul. 2019.

[37] 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.*, pp. 1–16, 2021.

[38] T. Shireen, C. Shao, H. Wang, J. Li, X. Zhang, and M. Li, “Iterative multi-task learning for time-series modeling of solar panel pv outputs,” *Appl. Energy*, vol. 212, pp. 654–662, Feb. 2018.

[39] T. S. Kim and S. Y. Sohn, “Multitask learning for health condition identification and remaining useful life prediction: deep convolutional neural network approach,” *J. Intell. Manuf.*, vol. 32, no. 8, pp. 2169–2179, Dec. 2021.

[40] H. J. Ferreau, C. Kirches, A. Potschka, H. G. Bock, and M. Diehl, “QpOASES: a parametric active-set algorithm for quadratic programming,” *Math. Program. Comput.*, vol. 6, no. 4, pp. 327–363, Dec. 2014.

[41] E. Ramasso and A. Saxena, “Performance benchmarking and analysis of prognostic methods for cmapss datasets.,” *Int. J. Progn. Health Manag.*, vol. 5, no. 2, pp. 1–15, 2014.

1. This work was supported in part by the National Science Foundation of China under Grant 72101148; in part by the Shanghai Sailing Program under Grant 21YF1420100, and in part by the National Science Foundation of Shanghai under Grant 22ZR1433000.

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

   Yuhui Wang is with the department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China (email: yh.wang@sjtu.edu.cn)

   Xiaochen Xian is with the department of Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA (email: xxian@ufl.edu) [↑](#footnote-ref-1)