A Data-Fusion Based Prognostic Method for Complex Degrading System

Xiaosheng Si
Department of Automation
Rocket Force University of Engineering
Xi'an, P. R. China
sixiaosheng@gmail.com

Ziqiang Ren
Department of Automation
Rocket Force University of Engineering
Xi'an, P. R. China
renzq_302@163.com

***Abstract*—This paper develops a method where A composite indicator constructed by fusing sensors-data for modeling and life prediction is used to Characterize the degradation state of devices. During constructing the composite indicator based on multi-sensors information, the objective is to minimize the mean squared life prediction error when the composite constructed indicator is used in prognostics. Based on the constructed composite indicator, Wiener process with Bayesian updating for random-effect parameter is used to predict the life of devices. Finally, the C-MAPSS data set demonstrates that the proposed method make the improvement of the life prediction accuracy.**

Keywords-health indicator; Wiener process; Bayesian parameter update; life prediction

I. INTRODUCTION

During the operation of the system or device, due to various factors, such as: external environment, working load, and its own structure, there will be inevitable degradation or even failure. For large structural parts, complex systems, military weapons, aero engines, etc., the failure of the system without warning will bring about major economic, environmental and personnel losses. Therefore, accurate, reliable and efficient inference of the predicted equipment failure time is of great significance to reduce various losses caused by equipment failure and improve the system's reliability[1, 2]. In recent years, aerospace technology has developed rapidly, and its aero-engines, which are key components, play a vital role in the normal operation of aircraft. Due to the complex structure of the engine system and the harsh working environment, problems such as engine safety assessment and maintenance replacement strategy formulation are prominent. The research and application of aeroengine prediction and health management (PHM) technology can provide technical support for its conditional maintenance, effectively ensure flight safety, reduce the incidence of major risk accidents, reduce maintenance costs, and improve Aircraft reliability, where the engine's remaining useful life (RUL) prediction is the most challenging technology in PHM [3].

At present, the life prediction methods mainly consist of two categories: methods based on physical failure models and data-driven methods. Since the degradation mechanism of complex systems is difficult to accurately model, The latter prediction method has become the mainstream in recent years. Hu Changhua et al systematically combed the data-driven method and divided it into three categories: methods of failure data, methods of degraded data, and prediction methods o multi-sources data fusion.

In the existing research, the Wiener process has good mathematical characteristics, which can not only describe the monotonic system degradation process, but also the non-monotonic system. Among them, Ren Shuhong et al analyzed the performance degradation mechanism of civil engines, and established the Brownian motion predicting degradation overrun time with drift coefficient. Zhu Lei et al. modeled the engine with a linear Wiener process, fused the failure data with the degraded data, and used the Bayesian estimation method to update the parameters, and then predicted the RUL of the engine. Zhou Jinglun et al. assumed that the Wiener process model parameters obey the normal distribution to characterize the differences between individual devices in a batch of devices [4].

For devices's degradation modeling and RUL prediction, most of the existing literature focuses on the analysis of single sensor signals under single operating conditions. with the rapid development of sensors technology and computers technology, the monitoring of a unit's health condition is often multiple sensors. The RUL prediction method based on single sensor monitoring data has the problems of low data utilization and low RUL prediction accuracy.

This paper proposes a method to constructs the composite indicator based on multi-sources sensor data, and then determines the sensor fusion coefficient by minimizing the mean square error of real life and predicted life, and then gets the composite indicator. The Wiener process with random parameters was used to predict RUL, and finally the C-MAPSS dataset was used for experimental verification.

II. OVERVIEW OF THE DATASET

C-MAPSS[5] is widely used in numerical simulation experiments to analyze condition monitoring and degradation

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1306100, the NSFC under Grants 61922089, 61773386, 61673311, 61573366, and 61573076, and in part by the Young Elite Scientists Sponsorship Program (YESS) of China Association for Science and Technology (CAST) under Grant 2016QNRC001.

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

of engine in various literature. Figure 1 shows a schematic diagram of the aircraft gas turbine engine

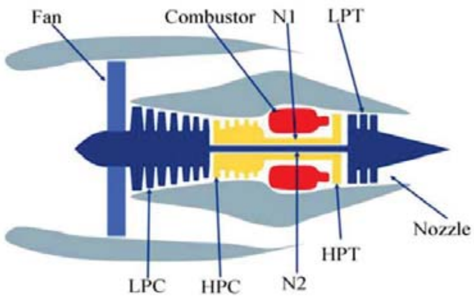


Fig 1. Simplified engine diagram simulated by C-MAPSS

This experiment uses the FD001 dataset in the C-MAPSS dataset to conduct experiments. The FD001 dataset records the degradation and failure data generated by the engine in the case of single-case and single-fault failure scenarios, including 100 training sets, 100 test sets and 100 RUL sets. The training set contains 100 failed engine data: a total of 20,631 sets of monitoring data; the test set contains 100 degraded engine data: a total of 13096 sets of monitoring data; The 100 RUL data correspond to the test data set, which are the RUL of each test engine at the last moment. Among them, each set of monitoring data contains monitoring data of 21. The details of this data set can be found in [6]. A structural block diagram of an engine RUL prediction method based on the multi-source sensor data is shown in fig.2.

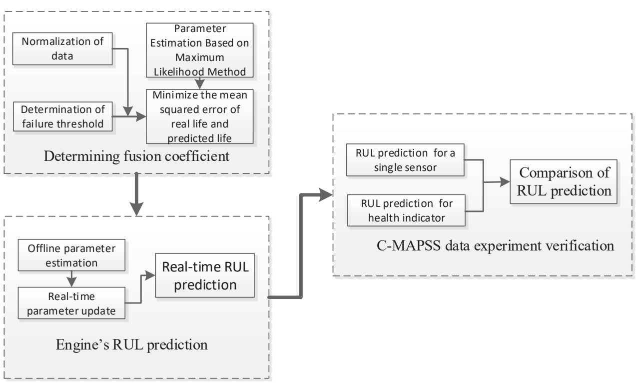


Fig 2. A structural block diagram of an engine RUL prediction

III. METHODOLOGY DEVELOPMENT

This section proposes a method of fusing sensors-data: linearly combining multi-sensors data to construct a composite indicator to characterize the degraded condition of the system [6]. This method can solve the complex problems of data processing and model construction.

A. *Construction of composite Indicator*

The composite indicator of the fusion multi-sensors data constructed in this article is as follows :

$$x_i(t) = \mathbf{Z}_i(t) \bullet \mathbf{W} = \sum_{j=1}^{S} z_{i,j}(t) \bullet w_j \tag{1}$$

Where $x_i(t)$ is the composite health indicator of training engine $i(i = 1, 2, ..., M)$ at the moment t ; $\mathbf{Z}_i(\mathbf{t}) \in R^{1\times S}$, $\mathbf{Z}_i(\mathbf{t}) = [z_{i,1}(t), z_{i,2}(t), ..., z_{i,S}(t)]$ is the sensor monitoring data vector of training engine i ; $z_{i,j}(t)$ is the monitoring data of training engine i for sensor j ; $\mathbf{W} = [w_1, w_2, ..., w_s]^T$, $\mathbf{W} \in R^{S\times 1}$ is a fusion coefficient vector and w_j represents the fusion coefficient of the sensor j , which measures the proportion of the sensor in the data fusion process.

Based on the constructed composite indicator, the failure threshold needs to be determined first. When the failure threshold of the engine cannot be known, the amount of degradation corresponding to the actual failure time is generally taken as the failure threshold. Let $x_i(t_{N_i}) = x_{i,N_i}$ be the composite indicator value of the training engine i at the time t_{N_i} of failure.

In order to reduce the uncertainty of useful life prediction, the engine's failure threshold is determined by minimizing the variance of the deceleration amount corresponding to the different training engine failure times, given the same failure mode and operating conditions [7]. Specifically, the failure threshold can be determined:

$$\min \; Y_P = \sum_{i=1}^{M} \left(P - x_{i,N_i}\right)^2 \tag{2}$$

From the above definition of $x_i(t_{N_i})$, the average threshold of the training engine can be determined as:

$$\overline{P} = \overline{x_{i,N_i}} = \frac{1}{M}\sum_{i=1}^{M} x_{i,N_i} \tag{3}$$

Obviously, the average failure threshold $\overline{P}$ is the optimal solution of equation (2), whereby the composite indicator failure threshold is $P = \overline{P}$.

In order to solve the problem of different sensor monitoring data dimension and value range, this article uses the method of data standardization to limit the sensor data to (0,1):

$$z^*_{i,j}\left(t_k\right) = \frac{z_{i,j}\left(t\right) - \min(\mathbf{z}_j)}{\max(\mathbf{z}_j) - \min(\mathbf{z}_j)} \tag{4}$$

Where $z^*_{i,j}\left(t_k\right)$ is the normalized value of $z_{i,j}\left(t_k\right)$; $\max(\mathbf{z}_j)$ is the maximum value of all the monitoring data of the sensor j in the M training engines from the start time to the failure time, and so on, $\min(\mathbf{z}_j)$ is the minimum value of all the monitoring data of the sensor j . Thus, the normalized value of the health indicator can be obtained according to formula (1): $x^*_i(t) = \mathbf{Z}^*_i(t) \bullet \mathbf{W} = \sum_{j=1}^{S} z^*_{i,j}(t) \bullet w_j$.

In addition, if the value of the same sensor in the engine appears larger, this value will be updated to one. For the same reason, if the value of the same sensor in the engine appears smaller, this value will be updated to zero.

B. Degradation Modeling of Engine

Use the Wiener process to model the composite indicator degradation process whose expression is as follows:

$$X^*_i(t) = x^*_{i,0} + \alpha_i t + \sigma_i B(t) \tag{5}$$

Where $X^*_i(t)$ is the composite degradation amount of the engine i at time t, $x^*_{i,0}$ is the composite degradation amount of the engine i at time $t_0 = 0$; α_i is the drift coefficient of the engine i, reflecting the degradation rate of different equipment; σ_i is the diffusion coefficient of the engine i, characterizing the process random uncertainty, α_i、σ_i are unknown parameters, which need to be estimated in the text; $B(t)$ is the standard Brownian motion: $B(t) \sim N(0,t)$.

Let $x^*_{i,k} = x^*_i(t_k)(k = 2,...,N_i)$. So, there is $\Delta x^*_{i,k}$, which is as follows:

$$\Delta x^*_{i,k} \sim N\left(\alpha_i \Delta t_k, \sigma_i^2 \Delta t_k\right) \tag{6}$$

The estimation of model parameter based on the maximum likelihood method are as follows:

$$\hat{\alpha}_i = \frac{\sum_{k=1}^{N_i} \Delta x^*_{i,k}}{\sum_{k=1}^{N_i} \Delta t_k} = \frac{x^*_{i,N_i} - x^*_{i,0}}{t_{N_i} - t_0} \tag{7}$$

$$\hat{\sigma}_i^2 = \frac{1}{N_i}\left\{ \sum_{k=1}^{N_i} \frac{\left(\Delta x^*_{i,k}\right)^2}{\Delta t_k} - \frac{\left(\sum_{k=1}^{N_i} \Delta x^*_{i,k}\right)^2}{\sum_{k=1}^{N_i} \Delta t_k} \right\} \tag{8}$$

C. Determination of Fusion Coefficient

For the stochastic degradation process given by equation (5), in the sense of FHT, the life and remaining useful life of the engine can be respectively defined as equations (9) and (10):

$$T_i = \inf\left\{ t : x_i^*(t) \geq P \middle| x_i^*(0) < P \right\} \tag{9}$$

Where T_i is a random variable, which means the life of the engine i; P is the failure threshold of the engine, determined by equation (3).

Let $L_{i,k}$ be the RUL of the engine i at time t_k, then the RUL is defined as:

$$L_{i,k} = \inf\left\{ l_k : x_i^*(t_k + l_k) \geq P \middle| x_i^*(t_k) < P \right\} \tag{10}$$

The time and remaining time of the first threshold are subject to the IG distribution.

$$\left(T_i | \alpha_i, \sigma_i\right) \sim IG\left(\frac{P - x^*_{i,0}}{\alpha_i}, \frac{(P - x^*_{i,0})^2}{\sigma_i^2}\right) \tag{11}$$

$$\left(L_{i,k} | \alpha_i, \sigma_i\right) \sim IG\left(\frac{P - x^*_{i,k}}{\alpha_i}, \frac{(P - x^*_{i,k})^2}{\sigma_i^2}\right) \tag{12}$$

The PDF of the engine life t can be obtained from the PDF formula of the IG distribution:

$$f(t;\alpha_i,\sigma_i) = \sqrt{\frac{(P - x^*_{i,0})^2}{2\pi t^3 \sigma_i^2}} exp\left(\frac{-(P - x^*_{i,0} - \alpha_i t)^2}{2\sigma_i^2 t}\right) \tag{13}$$

From the numerical characteristics of the IG distribution, the mathematical expectation of the engine life can be obtained as follows:

$$E(T_i) = \frac{P - x^*_{i,0}}{\alpha_i} \tag{14}$$

Let T_i be the real life of the engine i. In order to improve the accuracy of life prediction, the fusion coefficient should minimize the mean squared error of the predicted life and real life of the engine.

$$\min \ Y_{T_i} = \sum_{i=1}^{M} \left(\frac{(P - x^*_{i,0}) \cdot (t_{i,N_i} - t_{i,0})}{x^*_{i,N_i} - x^*_{i,0}} - T_i \right)^2 \tag{15}$$

Bringing the formula (7) and the formula (14) into the equation (15), the fusion coefficient can be obtained. Because the function structure of equation (15) is complex and difficult to solve, the fusion coefficient $\mathbf{W} = [w_1, w_2, ..., w_s]^T$ can be obtained by nonlinear programming method, multi-dimensional search method or intelligent optimization method. The experimental study in this paper uses a nonlinear programming method.

D. Prediction of Engine RUL

For the specific engine, the model of the composite indicator is as follows:

$$X^*(t) = x^*_0 + \alpha t + \sigma B(t) \tag{16}$$

a) Prior parameter estimation:

Assuming that there are historical life data for M same type of engine, the initial and failure values of their composite health indicator data sets are $\{x_1^*(t_0), x^*_2(t_0), ..., x^*_M(t_0)\}$ and $\{x^*_1(t_{N_1}), ..., x^*_M(t_{M,N_M})\}$, respectively.

To describe the individual differences between different engines, let α be random drift coefficient, and assuming its prior distribution is $\alpha \sim N\left(\mu_{\alpha 0}, \sigma^2_{\alpha 0}\right)$. To determine the hyperparameters $\mu_{\alpha 0}$ and $\sigma^2_{\alpha 0}$ in the prior distribution of α, the method is as follows:

$$\begin{aligned} \mu_{\alpha 0} &= \frac{1}{M}\sum_{i=1}^{M} \alpha_i \\ \sigma^2_{\alpha 0} &= \frac{1}{M-1}\sum_{i=1}^{M} (\alpha_i - \mu_{\alpha 0})^2 \end{aligned} \tag{17}$$

At the same time, let σ be the diffusion coefficient and represent the common characteristics between similar engines, which is determined by the following formula:

$$\sigma^2 = \frac{1}{M}\sum_{i=1}^{M} \hat{\sigma}_i^2 \tag{18}$$

b) Real-time parameter update

For the specific running engine, use the monitoring data of the running process to update the random parameter α of the model, so that the modeling result is more in line with the actual running condition of the engine. Assume that at time t_k, the data of all the sensors of the engine and the fusion coefficient determined above can be utilized, and the monitoring data of the composite indicator is obtained according to formula (1): $\mathbf{x}^*_{1:k} = \left\{ x^*(t_r), 0 = t_1 \le t_r \le t_k \right\}$.

Based on the prior distributions $\alpha \sim N\left(\mu_{\alpha 0}, \sigma^2_{\alpha 0}\right)$ and $\mathbf{x}^*_{1:k}$ of α, the posterior distribution of the model parameter α is still Gaussian [8]: $\alpha \mid \mathbf{x}^*_{1:k} \sim N\left(\mu_{\alpha k}, \sigma^2_{\alpha k}\right)$.

$$\mu_{\alpha k} = \frac{\mu_{\alpha 0}\sigma^2 + x^*_k \sigma_{\alpha 0}{}^2}{t_k \sigma_{\alpha 0}{}^2 + \sigma^2}; \sigma^2_{\alpha k} = \frac{\sigma^2 \sigma_{\alpha 0}{}^2}{t_k \sigma_{\alpha 0}{}^2 + \sigma^2} \quad (19)$$

c) RUL prediction

Since $\alpha \mid \mathbf{x}^*_{1:k} \sim N\left(\mu_{\alpha k}, \sigma^2_{\alpha k}\right)$, when considering the random characteristics of the drift parameter, the RUL's PHF of the engine obtained by the full probability formula is [8]:

$$f_{L_k}(l_k) = \int f_{L_k|\alpha}(l_k|\alpha) \bullet p(\alpha \mid \mathbf{x}^*_{1:k}) d\alpha$$
$$= \sqrt{\frac{\left(P - x^*_k\right)^2}{2\pi l_k{}^3 \left(l_k \sigma_{\alpha k}{}^2 + \sigma^2\right)}} exp\left(-\frac{\left(P - \mu_{\alpha k} l_k - x^*_k\right)^2}{2 l_k \left(l_k \sigma_{\alpha k}{}^2 + \sigma^2\right)}\right) \quad (20)$$

According to equation (20), the mathematical expectation of the engine to predict the remaining life is as follows:

$$E(l_k) = \int_0^\infty f_{L_k}(l_k) \cdot l_k dl_k = \frac{\sqrt{2}\left(P - x^*_k\right)}{\sigma_{\alpha k}} D\left(\frac{\mu_{\alpha k}}{\sqrt{2}\sigma_{\alpha k}}\right) \quad (21)$$

where $D(b) = \exp\left(b^2\right) \int_0^b \exp\left(b^2\right) db$ is the Dawson score for the real number b.

According to (20) and (21), the PDF and the mean value of the engine's RUL at any time can be determined, thereby realizing online prediction of the engine's RUL.

IV. Experimental Studies

By using the composite indicator method proposed in this article, the fusion coefficients of the 21 sensors determined are determined and the corresponding monitoring data of the composite indicator is obtained, as shown in Figure 3.

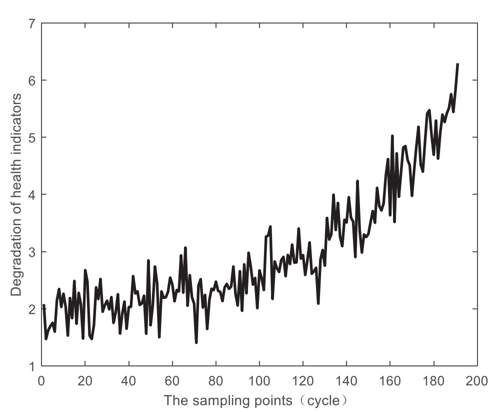


Fig 3. Monitoring data of composite indicator

As shown in Figure 3, there is a large random fluctuation in the monitoring data. Therefore, the data is pre-filtered by exponential smoothing. When filtering is performed, the window width is set to 10 in the exponential smoothing method, and the filtering result is as shown in Figure 4.

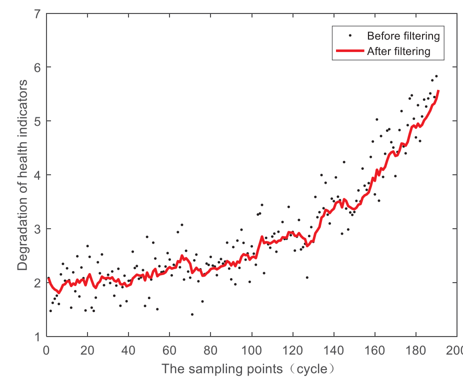
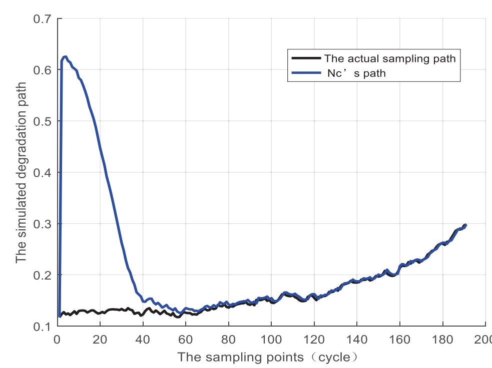
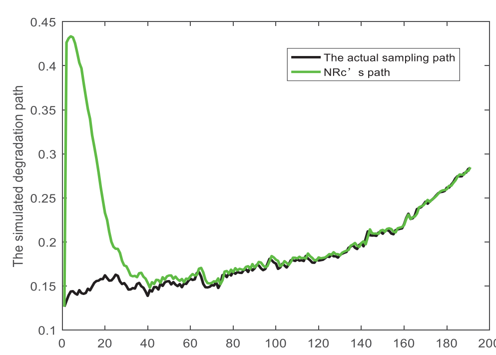


Fig 4. Monitoring data of composite indicator after exponential smoothing

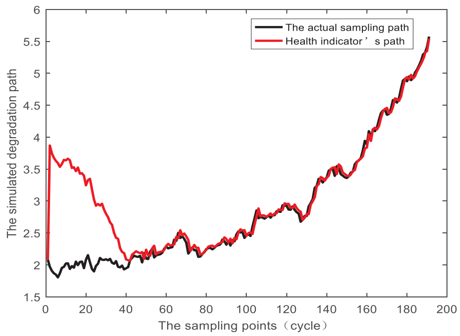
By analyzing the monitoring data, it is found that among the 21 sensors, not all sensor monitoring data can characterize the degraded condition of the engine. Therefore, this section selects the monitoring data of the two sensors with the best degradation trend among the 21 sensors: Nc sensor and NRc sensor. For the monitoring data of Nc sensor and NRc sensor, this experiment also uses the exponential smoothing method for filter preprocessing, and also are modeled by linear Wiener process.Figure 5 shows the comparison between the model degradation path and the real degradation path.



Nc's sensor



NRc's sensor



Composite indicator

Fig 5. The degradation path and fitted path

It can be seen from Fig. 5 that the degradation path of the engine model is far from the actual degradation path at the beginning, but the gap between the model degradation path and the actual degradation path becomes small. The main reason is that the error obtained from the historical data is relatively larger than the actual parameters of the engine degradation process. With the increase of real-time monitoring data, the more the existing degradation information is acquired, the more accurate the parameter estimation is. The degree of model fit is also higher. The engine's RUL prediction can be achieved based on the results of the degradation modelling, which is shown in Fig. 6.

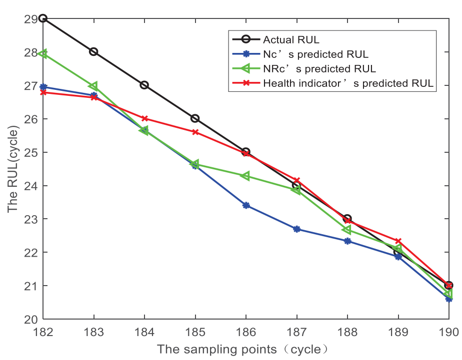
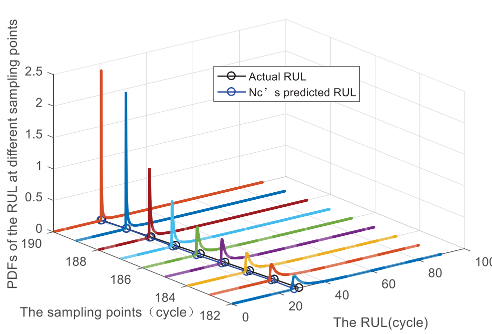
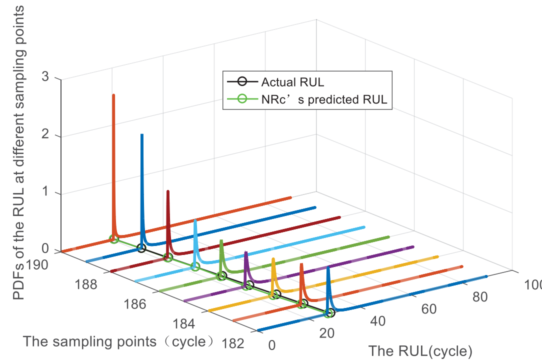


Fig 6. The prediction result of the RUL

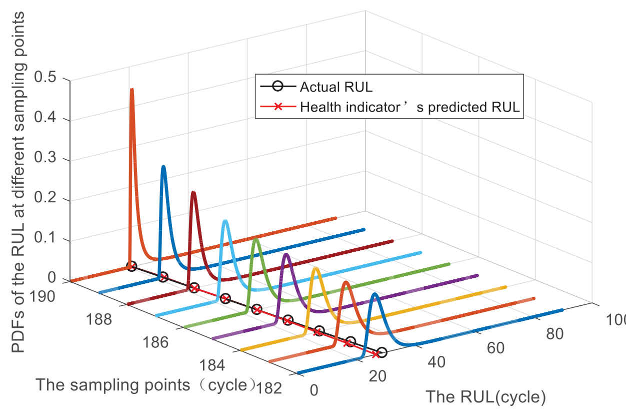
As indicated by Fig. 6, with the accumulation of monitoring data, the RUL prediction accuracy of the proposed method is higher than that method based on the single sensor. The results of the RUL probability density function predicted by the single sensor method and the composite health indicator method with the running time are shown in Fig. 7.



(a) The estimated PDFs of Nc's RUL



(b) The estimated PDFs of NRc's RUL



(c) The estimated PDFs of composite indicator's RUL

Fig 7. The estimated PDFs of the RUL

Based on the predicted RUL probability density function, the mean square error of the RUL prediction can be calculated. The comparison of the RUL mean square error predicted based on the single sensor and the composite health indicator method is shown in Fig. 8.

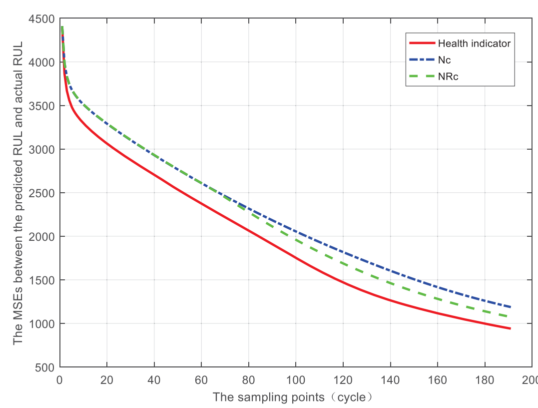


Fig 8. The MSEs between the predicted RUL and actual RUL

It can be seen from Fig. 8 that the RUL mean square error based on the single sensor and the composite health indicator method is getting smaller and smaller over time, but the error predicted based on the composite health indicator method is significantly smaller than that based on the single sensor method.

V. CONCLUSION

In this article, the RUL prediction method based on single sensor data has the problems of low data utilization and low prediction accuracy. This article combines multi-sensor data into a composite indicator to characterize engine degradation performance. In order to determine the fusion coefficient, a method for minimizing the predicted mean square error using real life and predicted life is proposed. After determining the fusion coefficient, based on the constructed composite indicator, Wiener process with Bayesian updating for random-effect parameter is applied to predict the RUL of engine. The simulation experiment based on engine C-MAPSS dataset shows that the proposed method make the improvement of the life prediction accuracy and overcome the problem of sensor selection in the application of engineering practice.

REFERENCES

[1] Z. Huang, Z. Xu, and W. Wang, "Remaining Useful Life Prediction for a Nonlinear Heterogeneous Wiener Process Model With an Adaptive Drift," *IEEE T. Reliab.*, vol. 64, no. 02, pp. 687-700, 2015.

[2] D. H. Zhou, M. H. Wei, and X. S. Si, "A Survey on Anomaly Detection, Life Prediction and Maintenance Decision for Industrial Processes," *Acta Automatica Sinica,* vol. 39, no. 06, pp. 711-722, 2013.

[3] G. Jian, S. Hui, and H. Fang, "The design of the prognostics and health management system for special vehicle," P*rognostics & System Health Management Conference*, 2017.

[4] A. Chehade, C. Song, and K. Liu, ""'A data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes," *J. Qual. Technol.*, vol. 50, no. 02, pp. 150-165, 2018.

[5] Saxena A and Goebel. "C-MAPSS data set," *NASA Ames Prognostics Data Repository*, 2008.

[6] H. Yan, K. Liu, and X. Zhang, "Multiple Sensor Data Fusion for Degradation Modeling and Prognostics Under Multiple Operational Conditions," *IEEE Trans. Rel.,* vol. 65, no. 03, pp. 1416-1426, 2016.

[7] S. B. Gao, "Estimation Research in Remaining Useful Life Prediction Driven by Degradation Data," *Dalian University of Technology*, 2016.

[8] X. Si, W. Wang, and Hu C, "Remaining Useful Life Estimation Based on a Nonlinear Diffusion Degradation Process," *IEEE Trans. Rel.*, vol. 61, no. 01, pp. 50-67, 2012.