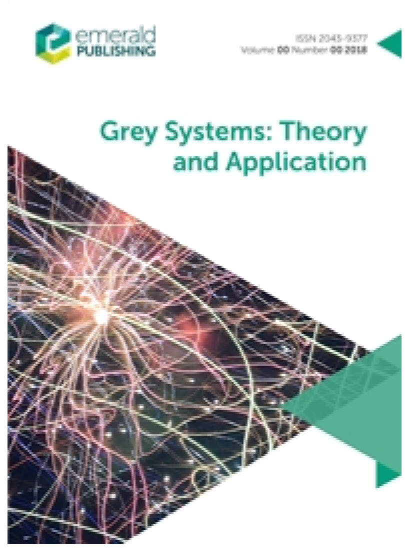

Similarity-based Information Fusion Grey Model for Remaining Useful Life Prediction of Aircraft Engines

Journal:	*Grey Systems: Theory and Application*
Manuscript ID	GS-05-2020-0066.R2
Manuscript Type:	Full Length Paper
Keywords:	GM(1,1), remaining useful life, similarity-based, information fusion, aircraft engine

Similarity-based Information Fusion Grey Model for Remaining Useful Life Prediction of Aircraft Engines

Abstract

Purpose –On-line health monitoring of large complex equipment has become a trend in the field of equipment diagnostics and prognostics due to the rapid development of sensing and computing technologies. The purpose of this paper is to construct a more accurate and stable grey model based on similar information fusion to predict the real-time remaining useful life (RUL) of aircraft engines.

Design/methodology/approach –First, a referential database is created by applying multiple linear regressions on historical samples. Then similarity matching is conducted between the monitored engine and historical samples. After that, an information fusion grey model is applied to predict the future degradation trajectory of the monitored engine considering the latest trend of monitored sensory data and long-term trends of several similar referential samples, and the real-time RUL is obtained correspondingly.

Findings –The results of comparative analysis reveal that the proposed model, which is called similarity-based information fusion grey model (SIFGM), could provide better RUL prediction from the early degradation stage. Furthermore, SIFGM is still able to predict system failures relatively accurately when only partial information of the referential samples is available, making the method a viable choice when the historical whole lifecycle data are scarce.

Research limitations/implications –The prediction of SIFGM method is based on a single monotonically changing health indicator (HI) synthesized from monitoring sensory signals, which is assumed to be highly relevant to the degradation processes of the engine.

Practical implications –The similarity-based information fusion grey model can be used to predict the degradation trajectories and RULs of those online condition monitoring systems with similar irreversible degradation behaviors before failure occurs, such as aircraft engines and centrifugal pumps.

Originality/value – This paper introduces the similarity information into traditional GM(1,1) model to make it more suitable for long term RUL prediction, and also provide a solution of similarity-based RUL prediction with limited historical whole lifecycle data.

Keywords GM(1,1), remaining useful life, similarity-based, information fusion, aircraft engine

Paper type Research paper

1. Introduction

Prognostic and Health Management (PHM) is a promising tool for optimal operation and maintenance solution of equipment which has complex structure and operates at high speeds, high load and high temperature environment, and has been widely applied in many industrial fields such as electricity, aircraft and rail transit, etc.. Prognostics is defined as “the estimation of time to failure and risk for one or more existing future failure modes” by the international standard organization, so remaining useful life (RUL) prediction is one of the main tasks of PHM. Accurate estimation of equipment RUL contributes to maintenance decision-making and potential failure discovery, in which way measures can be taken to effectively avoid unexpected breakdown, extend the service life, and reduce maintenance costs.

Prognostic methods can be broadly categorized into physics-based and data-driven approaches (Peng *et al.*, 2018). Generally, physics-based methods aim to implement RUL prediction by accurately describing the equipment’s failure mechanism and degradation process with limited historical data (Wang *et al.*, 2016; Zhu *et al.*, 2012). It needs to be constructed according to the physical laws by which the equipment follows during degradation and the interrelationships between different parts of the system. However, it is rather challenging and costly to build an accurate physical model for complex systems. The data-driven approach can get around these difficulties, and provide accurate RUL predictions timely and economically due to the rapid growing of sensor technology and real-time computing. Data-driven methods involve using health indicators (HI) for degradation degree indication of the system. Machine learning methods like neural network (Chen and Liu, 2015), support vector machine (SVM) (Li *et al.*, 2015; Chen *et al*, 2018), Bayesian learning (Cao *et al*, 2017), hidden Markov model (Peng and Dong, 2011), particle filter (Zhang *et al.*, 2018) and statistical methods (Bakir *et al.*, 2019) have been widely used in data-driven RUL estimation.

Extensive efforts have been made to develop data-driven prognostic methods for RUL prediction. Zhao *et al.* (2017) proposed a degradation pattern learning method using an improved back propagation neural network for aircraft engine RUL estimation. García Nieto *et al.* (2015) built a hybrid model based on SVM and particle swarm optimization (PSO) to predict the RUL of aircraft engines. Li and Lei (2017) simulated degradation trajectories by applying stochastic process models and estimated the probability density function of RUL. Lasheras *et al.*(2015) considered failure risks by integrating the multivariate adaptive regression splines, principal component analysis (PCA) and classification and regression trees. To improve the performance of PCA, an adaptive canonical variate analysis was proposed to predict the RUL of slowly evolving faults (Li and Duan, 2017; Li and Yang, 2019). The above approaches focus on modeling the degradation process via condition monitoring data and have achieved a satisfied level of accuracy. However, they require a time-consuming training process and large amounts of degradation data to ascertain accurate estimations.

The similarity-based prognostic approaches, which extract information from similar samples to make up the potential data deficiency of the monitored equipment and do not require a sophisticated degradation model, provide a different line of thought. The main idea of similarity-based RUL prediction can be expressed as: if the test sample and referential samples have similar degradation pattern over a certain period of time, they might have a similar RUL (Liu and Wang, 2017). The degradation features/health indices extracted from initial sensor data are applied for similarity matching between the present monitored and historical samples, then the RUL of the present sample can be computed as the weighted average of the RULs of several most similar historical samples with the weights being determined as the similarity score between samples. Liu and Hu (2019) put forward a similarity matching approach to predict the RUL of cutting tools considering Euclidean distance similarity and spatial direction. Zio *et al.* (2010) proposed a RUL prediction approach based on fuzzy similarity analysis, which compared degradation data of the observed system to the trajectory patterns of reference systems. Li and Zhang (2018) presented a RUL estimation method by comparing the similarity of fixed-length health index curves between the test unit and the extracted representative curves. Liang *et al.* (2018) proposed a method to estimate RUL based on the similarity between raw sensory measurements. Huang *et al.* (2019) put forward an improved trajectory similarity-based approach which could provide confidence interval of RUL prediction on the basis of adaptive kernel density estimation (KDE) and β-criterion.

Although the similarity-based methods have attracted great attention in recent years, they have some limitations. For example, if the similar samples found in the training database are not similar enough to the test sample, the RUL predictions may be inaccurate. In addition, similarity-based RUL algorithms require a large number of samples based on full life cycle data. For large-scale equipment, it may be the case that the machines are not allowed to run to failure and hence only a portion of the degradation data is available. Grey model (GM) offers a new possibility to deal with the small sample size data, which is able to achieve relatively accurate prediction with only 4 data points (Liu and Yang, 2016). Some exploratory efforts on failure prognostics using GM method have been made and satisfactory results have been observed. Li and Miao (2015) proposed a combined model of GM and SVM to forecast the RUL of batteries. Zhou *et al.* (2014) proposed a revised grey model with a forgetting factor to predict the RUL of lithium-ion batteries.

Due to the fact that sensory signals are fluctuated and GM is sensitive to new data information, the applicability of GM in long-term failure prognostics is relatively limited. To compensate for the deficiency of whole lifecycle samples and the instability of GM in long-term prediction, this paper puts forward a similarity-based information fusion grey model (SIFGM) for remaining useful life prediction. In this framework, the degradation of the system is characterized by a synthesized health indicator from monitoring sensory signals, and the degradation trends extracted from similar samples are utilized to assist the degradation trajectory prediction of monitored engine by creating a mapping sequence. By leveraging the advantages of the forecasting capability of GM for small sample data and the opportunities of mutual complementarity provided by the similarity-based method, the SIFGM is demonstrated more accurate and robust in RUL estimation of aircraft engines. Moreover, the SIFGM method is able to make the full use of both whole lifecycle failure data and degradation data with only partial information.

The remainder of the paper is organized as follows. Section 2 reviews the framework of the original similarity-based RUL estimation method. Section 3 describes the modeling procedure of similarity-based information fusion GM(1,1) model. Section 4 demonstrates the SIFGM for remaining useful life prediction on NASA C-MAPSS dataset, both scenarios of whole lifecycle failure data and partial degradation data are considered. Section 5 provides the conclusions of this paper.

2. Framework of similarity-based RUL estimation method

The general similarity-based RUL prediction methodology consists of three essential procedures: HI construction, similarity matching and RUL estimation. Figure 1 shows the flowchart of the method.

(1) HI construction: the originally monitored sensory data is processed after data acquisition, including failure related features selection and data normalization, and then a health indicator (often one-dimensional) is constructed to indicate the hidden health status of the system. Sensory data of historical failure cases are used to obtain a reliable HI and form a referential database, and the test data will be preprocessed using the same steps.

(2) Similarity matching: the HI of the test sample will be compared with the HIs in the referential sample database using a retrieval module, and the similarity degree between the test sample and each training samples in the database is calculated by a specific similarity measure function.

(3) RUL estimation: the RUL of the test sample is normally estimated by applying a weighted average of the RULs of the most similar historical samples in referential database.

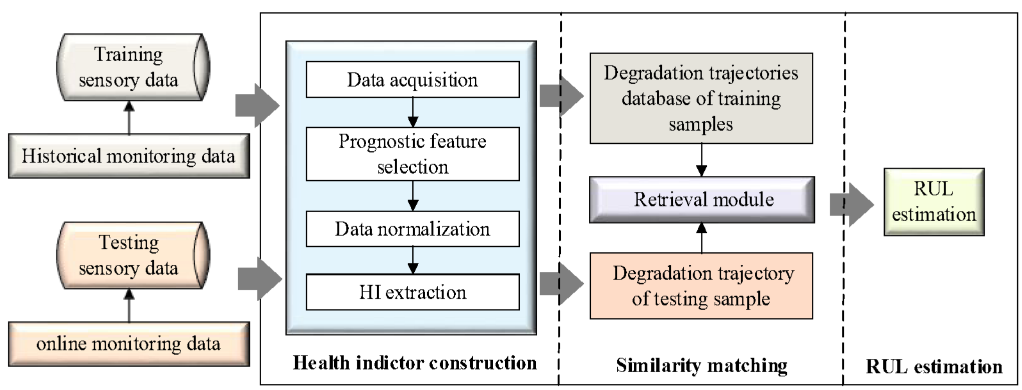

Figure 1. Framework of the conventional similarity-based RUL method

2.1 HI construction

An appropriate HI is the starting point for prognostics analysis which can facilitate prognostics by simplifying the prognostic modeling. Before building a HI, a data pre-processing step is needed in real condition monitoring applications. It is observed that data from multiple sensors exhibit various characteristics. The variation curve of these sensor data can fall into two categories. One category of sensor values demonstrates an increasing or decreasing tendency with time prolonged, while another category of sensor values remains constant or diverges over time. Only the first category of sensor values can effectively reflect the degradation trend of the system. Different sensor selection methods have been proposed in previous research (Wang *et al.*, 2008; Liang *et al.*, 2018; Xu *et al.*, 2014). In this part, we apply the sensor combination utilized in research work by Huang *et al.* (2019). The chosen sensor combination shows an obvious trend of monotonic change and produces satisfying prediction performance along with our proposed method.

Since the scales of sensory measurements are totally different, data normalization is necessary to convert different scales of data into a nominal range for further health indicator extraction. Specifically, z-score normalization is utilized in this work, which can be described as follows:

$$v_i^{'} = \frac{v_i - \mu_i}{\sigma_i}, \tag{1}$$

where v_i ($i = 1, 2, \cdots, N$) is the value of i^{th} sensor data, N is the number of selected sensors from monitoring samples, $v_i^{'}$ is the normalized data of i^{th} sensor data, μ_i and σ_i represent the corresponding mean and standard deviation of the i^{th} sensor data respectively.

After data normalization, all sensory data are in a nominal range and a single dimension HI can be generated. Many dimensionality reduction techniques such as logistic regression (Yan *et al.*, 2004), multiple linear regression (MLR) (Huang *et al.*, 2019), principal component analysis (PCA) (Liu *et al.*, 2019), canonical variate analysis (CVA) (Li and Yang, 2019) are presented to reduce the multi-dimensional features from the original monitoring data to a single dimensional HI. Particularly, MLR is chosen in this paper to indicate the degradation of the system since MLR can preserve the original degradation patterns in the signal data (Wang *et al.*, 2008) and have a relative low computation complexity. The synthesized HI is computed as follows:

$$x = \beta_0 + \sum_{i=1}^{N} \beta_i v_i^{'} + \varepsilon \tag{2}$$

where $V' = (v_1^{'}, v_2^{'}, \cdots, v_N^{'})$ is the normalized N-dimensional feature vector, x is the health indicator, $(\beta_0, \beta_1, \cdots, \beta_N)$ are the $N+1$ model parameters, and ε is the noise term. To obtain the model parameters, one can assign the synthesized health indicator values of healthy conditions with 1 and near failure conditions with 0, respectively. The parameters of the MLR model $(\beta_0, \beta_1, \cdots, \beta_N)$ are then stored for further utilization to testing samples.

2.2 Similarity matching and traditional similarity-based RUL prediction

In this stage, the HI trajectory of the testing sample is generated by following the same data preprocessing procedure as described in Section 2.1, and is compared with the HIs in referential sample database. The similarity matching process consists of three main elements: sliding window size, similarity measurement function and weight function.

The first step of a trajectory pointwise similarity matching method is to determine the sliding window size L for similarity matching. At operating time t, the length-fixed HI trajectories of the test sample can be expressed as

$$X_{t,T} = (x_{t-L+1,T}, \cdots, x_{t-1,T}, x_{t,T}),$$

Where $x_{t,T}$ denotes the HI of the testing sample at time t. The HI of the j^{th} referential sample at operating time t can be expressed as

$$X_{t,j} = (x_{t-L+1,j}, \cdots, x_{t-1,j}, x_{t,j}),\ j = 1, 2, \cdots, R,$$

where R is the number of available referential samples, and $x_{t,j}$ denotes the HI of j^{th} referential sample at time t. Similarity matching can be processed by any consecutive HI trajectories of a referential sample before its failure. In addition, the sampling interval of the test sample and referential samples should be the same.

Similarity degree is then defined and calculated to measure the similarity between the test sample and referential samples. The similarity measurement is the function of two length-fixed HI trajectories of the operating and referential samples, respectively. Normally, the similarity measurement is a distance measure function, which can be Euclidean distance, cosine distance, probability function or membership function. In this study, the distance of the two samples is measured by the Euclidean distance, which is expressed as:

$$d_{T,j} = \left(\sum_{i=1}^{L} (x_{t-i+1,T} - x_{t-i+1,j})^2 \right)^{1/2} \tag{3}$$

where $d_{T,j}$ is the distance measure of degradation trajectory between the test sample and j^{th} referential sample in the time range $[t-L+1,t]$.

In order to avoid introducing dissimilar samples, the similarity degree is calculated only when $d_{T,j} < d_{th}$, where d_{th} is the maximum acceptable similarity measurement distance. The similarity degree $s_j(t)$ between the test sample and the j^{th} referential sample at time t is defined as:

$$s_j(t) = (1 + d_{T,j})^{-1} \tag{4}$$

The weight function is then defined based on similarity measurement. The weight of the j^{th} referential sample is given by

$$w_j(t) = s_j(t) \Big/ \sum\nolimits_{i=1}^{h} s_j(t) . \tag{5}$$

For traditional similarity-based prediction approaches, the RUL of the test sample at time t is the weighted average of the RULs of several most similar referential samples, namely

$$RUL_T(t) = \sum_j w_j \cdot RUL_j(t) . \tag{6}$$

3. Proposed similarity-based information fusion grey model for RUL estimation

It can be seen from Section 2 that the training samples in traditional similarity-based RUL estimation should be whole lifecycle data, and these methods focus only on the failure time prediction rather than the HI trajectory of the testing sample. Knowing the development of a fault allows one to not only predict the failure time but also evaluate the risk of operation at every time instance, which would be of interests to site engineers. The proposed similarity-based information fusion grey model for RUL estimation starts with the HI trajectory construction of testing sample. The framework of the RUL estimation module with respect to the proposed method is given in Figure 2. After implementing the aforementioned similarity retrieval module, the development tendencies of the most similar samples are computed by GM(1,1). The HI trajectory of testing sample is predicted by a metabolism information fusion GM(1,1) which combines the both information from historical samples and the online sample.

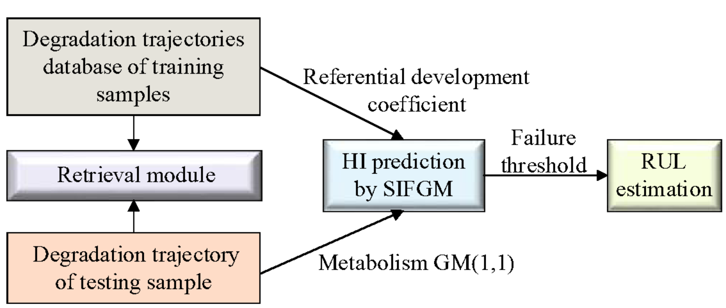

Figure 2. Framework of the proposed similarity-based RUL method

3.1 GM(1,1)

GM series models are the basic of grey prediction theory, especially the original even grey model proposed by Deng is broadly used (Liu *et al.*, 2016). GM(1,1), which means a single variable first order grey model, extracts evolution law of the sequence by a single variable differential equation of first order.

Definition 1. Assume a non-negative sequence of raw data is

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)),$$

then $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n))$ is called first order accumulative generation sequence of $X^{(0)}$, where

$$x^{(1)}(i) = \sum_{j=1}^{i} x^{(0)}(j), \tag{7}$$

and $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \cdots, z^{(1)}(n))$ is called the background value of sequence $X^{(0)}$,where

$$z^{(1)}(i) = 0.5\left(x^{(1)}(i-1) + x^{(1)}(i)\right). \tag{8}$$

Definition 2. Build a differential equation of $X^{(1)}$, the equation

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{9}$$

is defined as the whitening form of GM(1,1), where the parameter a is called the development coefficient and b is called the grey action quantity.

The parameter vector $\Phi = [a, b]^T$ of GM(1,1) model can be obtained by the least square estimation $\Phi = (B^T B)^{-1} B^T Y$, in which

$$Y = \left[x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n)\right]^T$$

$$B = \begin{bmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(n) \\ 1 & 1 & \cdots & 1 \end{bmatrix}^T,$$

Taking $x^{(1)}(1) = x^{(0)}(1)$ as the initial value, then the time response equation of GM(1,1) is

$$\hat{x}^{(1)}(i) = (x^{(0)}(1) - b/a)e^{-a(i-1)} + b/a \tag{10}$$

To obtain the predicted value of the primitive data at time k, the inverse accumulated generating operation is used to establish the following grey model:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) = (1-e^{a})(x^{(0)}(1) - b/a)e^{-a(k-1)} \tag{11}$$

3.2 SIFGM for RUL prediction

After a suitable prognostic feature is constructed and several similar referential samples are selected, the SIFGM prognostic technique can be applied to predict the RUL of the running machine. The GM(1,1) is trained using a small number historical data of the test sample, which do not consider measurement and process noise. The GM(1,1) is very sensitive to the latest modeling data subject to the modeling mechanism, on the one hand, it could better reflect the latest trend of the test sample, on the other hand, the outputs may be affected by measurement and process noise, which lead to a poor prediction in long-term. It has been proved that information fusion could improve the accuracy and robustness of traditional GM(1,1) method (Yang *et al.* 2019), which inspires the idea of SIFGM. The SIFGM will not only learn from the historical information of the test sample, but also learn from the future information of the similar referential samples from the observation point.

The schematic of an information fusion GM(1,1) is illustrated in Figure 3. As shown in the figure, the final predicted trajectory of SIFGM is between the GM(1,1)'s prediction and the integrated referential trajectory which is generated from the selected similar samples in the training database. The latest tendency of the test sample is captured using GM(1,1), and the predictive failure time is closer to the true failure time by considering similar referential information.

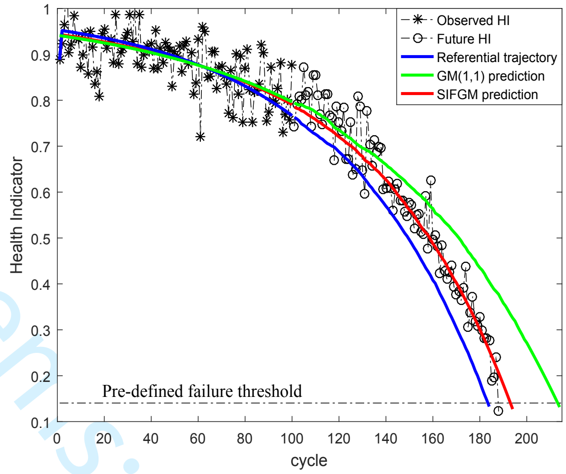

Figure 3. Illustration of the SIFMG prediction

Definition 3. Suppose $X_{r_1}^{(0)}, X_{r_2}^{(0)}, \cdots, X_{r_h}^{(0)}$ are the selected referential sequences of testing sample $X^{(0)}$ by the similarity matching procedure, $a_{r_1}, a_{r_2}, \cdots, a_{r_h}$ are the corresponding development coefficients of referential sequences $X_{r_1}^{(0)}, X_{r_2}^{(0)}, \cdots, X_{r_h}^{(0)}$ calculated by GM(1,1) from the end point of similarity matching to the end of each sequence. The integrated referential development coefficient $-a_r$ is defined as the weighted average of the referential development coefficients,

$$a_r = w_1 a_{r_1} + w_2 a_{r_2} + \cdots + w_h a_{r_h}, \tag{12}$$

where $w_j (j = 1, 2, \cdots, h)$ is the weight of referential sequence $X_{r_j}^{(0)} (j = 1, 2, \cdots, h)$ determined by similarity degree.

From the time response equation of GM(1,1), it is observed that the simulated curve of GM(1,1) is determined by three parameters: development coefficient, grey action quantity and initial value. Though integrated referential development coefficient $-a_r$ reflects the information of all selected referential sequences, to further determined the curve shape of referential development sequence, additional information is needed. So we creatively construct a mapping sequence according to the latest section of the monitored sample in SIFGM.

Theorem 1. Given a testing sample $X^{(0)}$, if the first point and the end point of are $X^{(0)}$ adopted to construct the mapping sequence $X_m^{(0)}$, and $X_m^{(1)}$ is denoted as the first order accumulative generation sequence of $X_m^{(0)}$, then

$$x_m^{(1)}(i) = (x^{(1)}(1) - b_m / a_m) e^{-a_m (i-1)} + b_m / a_m, \tag{13}$$

$$x_m^{(0)}(i) = x_m^{(1)}(i) - x_m^{(1)}(i-1) = (1 - e^{-a_m})(x^{(1)}(1) - b_m / a_m) e^{-a_m (i-1)}, \tag{14}$$

in which $a_m = a_r$ and $b_m = \dfrac{a_m (x^{(1)}(1) - x^{(1)}(n) e^{a_m (n-1)})}{1 - e^{a_m (n-1)}}$, and $X^{(1)}$ is the first order accumulative generation sequence of $X^{(0)}$.

Proof. Consider that the evolution trend of GM(1,1) models is determined by the development coefficient, so the development coefficient of the mapping sequence is assumed to be the same as $-a_r$. When the development coefficient of mapping sequence is determined by $a_m = a_r$, the other parameters of the mapping sequence can be obtained from the differential equations. Since the general solution of the grey differential equation in GM(1,1) can be represented as (15),

$$x^{(1)}(t) = ce^{-at} + b/a\,, \tag{15}$$

where c is a constant, a and b are parameters calculated by the least square estimation method.

From the description of (15), the mapping sequence we created is a particular solution. To make the mapping sequence useful for HI trajectory prediction, the initial condition is set to $x_m^{(1)}(1) = x^{(1)}(1)$ and $x_m^{(1)}(n) = x^{(1)}(n)$. For the general solution in (15), constant c is determined by the initial value, so

$$c_m = x^{(1)}(1) - b_m/a_m\,. \tag{16}$$

Then the unknown parameter b_m can be obtained as (17)

$$b_m = \frac{a_m(x^{(1)}(1) - x^{(1)}(k)e^{a_m(n-1)})}{1 - e^{a_m(n-1)}}\,. \tag{17}$$

Definition 4. $\hat{X}_f^{(0)}$ is denoted as the simulated value of the SIFGM,

$$\hat{x}_f^{(0)}(i) = (1-\alpha)\hat{x}^{(0)}(i) + \alpha x_m^{(0)}(i)\,. \tag{18}$$

where α is called the fusion coefficient, and the value range of the fusion coefficient α is $[0,1]$.

The value of fusion coefficient reflects the degree of reliance on referential information, and is depended on the impact of the noise to the sensory data. When $\alpha = 0$, the SIFGM(1,1) model deteriorates into the normal GM(1,1); when $\alpha = 1$, the SIFGM(1,1) model deteriorates into predicting the changes of its mapping sequence. Generally, the greater the sensory data is affected by the noise, the larger the value of fusion coefficient. In engineering practice, the value of fusion coefficient could be estimated by testing the fusion coefficient for the best prediction performance using several sequences in the historical database.

The future HI trajectory of the test sample at time t is given by a metabolism SIFGM model, the predictive HI value of the test sample at time $t+1$ is

$$\hat{x}_f^{(0)}(t+1) = (1-\alpha)\hat{x}^{(0)}(t+1) + \alpha x_m^{(0)}(t+1)\,. \tag{19}$$

The idea of metabolism SIFGM is taking $\hat{x}_f^{(0)}(t+1)$ as the new information of the original sequence $X^{(0)}$, and removing the first data sample, so the updated sequence is $X^{(0)} = (x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(k), x^{(0)}(t+1))$. Repeat the above metabolism procedure until the $\hat{x}_f^{(0)}(t+l)$ first exceed the pre-defined failure threshold, then the RUL of the test sample at observation time k is determined as l.

In this study, we use the information of the similar sequences' development trend to realize multi-step ahead predictions until the failure threshold is reached, which makes it possible to use both the run-to-failure data and degradation data without failure as references.

4. Case study: C-MAPSS dataset #1

To comprehensively evaluate the prediction performance of the proposed SIFGM method, this study adopts two scenarios using C-MAPSS dataset #1, which are whole lifecycle data and partially degradation data without failure. Some comparisons with other approaches are also included.

4.1 Data description

The data used in this experiment are from the C-MAPSS simulation program provided by the prognostics center of excellence of National Aeronautics and Space Administration (NASA), which aims to solve the lack of run-to-failure test data for data-driven prognostics (Saxena *et al.*, 2008). The data set includes 4 sets of data under different operating states and fault modes, in which 21 sensors (as shown in Table 1) and 3 input parameters are used to record the entire engine degradation process. Each data set includes a training set, a test set and a residual life subset. Each dataset contains information of the residual life of multiple engines in a homogeneous state. These engines are in normal state at the beginning, and then failure occurs at a certain moment and show obvious degradation of performance, and the degradation continues to accumulate until the system failure. Because of the degradation process integrity of the training data set, this paper adopts the samples in first group of training data set for demonstration. The first group of training data set is the assembly of High-Pressure Compressor (HPC) failure generated in the same operating state, including the monitoring samples of 100 compressors.

Table 1 Sensor data description

No.	Description	Units
1	Total temperature at fan inlet	°R
2	Total temperature at LPC outlet	°R
3	Total temperature at HPC outlet	°R
4	Total temperature at LPT outlet	°R
5	Pressure at fan inlet	psia
6	Total pressure in bypass-duct	psia
7	Total pressure at HPC outlet	psia
8	Physical fan speed	rpm
9	Physical core speed	rpm
10	Engine pressure ratio (P50/P2)	--
11	Static pressure at HPC outlet	psia
12	Ratio of fuel flow to Ps30	pps/psi
13	Corrected fan speed	rpm
14	Corrected core speed	rpm
15	Bypass Ratio	--
16	Burner fuel-air ratio	--
17	Bleed Enthalpy	--
18	Demanded fan speed	rpm
19	Demanded corrected fan speed	rpm
20	HPT coolant bleed	lbm/s
21	LPT coolant bleed	lbm/s

The health condition of the aircraft engine is monitored by 21 sensors equipped on the engine itself. Since some signals retain their constant values (i.e., No. 1, 5, 6, 10, 16, 18, and 19), these sensory data are discarded first. Then, the remaining 14 sensors are further analyzed. The sensor combination utilized in research work by (Huang *et al*. 2019) is applied here, namely the results of sensor selection are No. 2, 4, 7, 8, 11, 12 and 15. The selected seven sensor signals have shown certain degradation trends, then data normalization is applied to convert different scales of multi-sensor into a nominal range, finally HI is constructed by multiple linear regression (MLR) as mentioned in Section 2.1. The degradation trajectories database of training samples is shown in Figure 4.

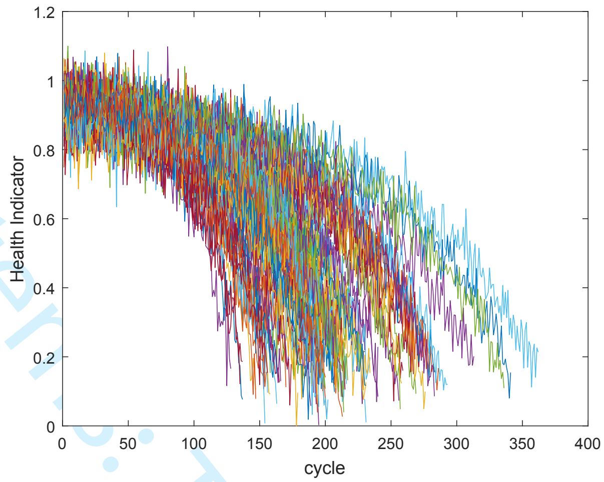

Figure 4. Degradation trajectories database of training samples

4.2 Prognostic performance criteria

To comprehensively evaluate the performance of the prediction approaches, two evaluation criteria for each prediction trajectory are chosen in this paper, i.e. mean absolute deviation (MAD) and mean prediction score (MPS).

Mean absolute deviation (MAD) is a common criterion used in prediction models, which is defined as

$$\mathrm{MAD} = \frac{1}{K}\sum_{t=t_0}^{K}\left|\hat{R}_t - R_t\right| \tag{20}$$

where K is the actual RUL from the prognostic starting point t_0, $\hat{R}_t$ is the predictive RUL at time t, and R_t is the actual RUL at time t.

Another assessing criterion, prediction score, is an asymmetric penalty function with late predictions penalized more heavily than early predictions, which is defined by the PHM community as

$$s_t = \begin{cases} e^{-d_t/13} - 1, for\ d_t \le 0 \\ e^{d_t/10} - 1, for\ d_t > 0 \end{cases}, \tag{21}$$

where $d_t = \hat{R}_t - R_t$ is the predictive error at time point t, and s_t denote the prediction score at time point t. The mean prediction score of a test sample from the prognostic starting point t_0 is defined as

$$MPS = \frac{1}{K}\sum_{t=t_0}^{K} s_t\ . \tag{22}$$

where K is the actual RUL from the prognostic starting point t_0. The MPS is an average score over the whole testing sample from the prognostic starting point t_0 to the engine failure.

4.3 Results and discussion

To demonstrate the effectiveness of developed SIFGM method, exponential regression (ER) method and the original grey model (GM(1,1)) are chosen here to conduct a comparative analysis. This choice is mainly driven by the fault evolution behavior of the aircraft engine dataset derived by C-MAPSS, which is suggested to follow an exponential function. More specifically, the exponential function applied here is

$$x = x_0 - \lambda_1 \exp(\lambda_2 t + \lambda_3) \tag{23}$$

where x_0 is the initial HI of the engine reflected the non-zero initial degradation, $\lambda_1, \lambda_2, \lambda_3$ are the parameters obtained by solving a nonlinear least squares curve fitting problem, and x is the synthesized HI.

The developed SIFGM method and its counterparts need to predict the degradation trajectory of the test sample first, and the RUL is further determined by a pre-defined failure threshold, which is set as the average minimum HI value of all training samples in the database. Note that three key parameters affecting the performance of the proposed method, where the sliding window size L and the number of the most similar samples h are related to similarity matching process and the fusion coefficient α is related to the prediction process. All parameters are tested using 100 training samples. It should be noticed that the closer to the end point of the sequence, the more accurate the prediction. So for testing a suitable L, all predictions start from 100th cycle to avoid unfair comparison. Since the final results is not only related to L, but also the number of the most similar samples h, 4 different values of h are chosen in Figure 5. Figures 5-7 present the average MAD and MPS for the 100 training samples given different values of L, h and α.

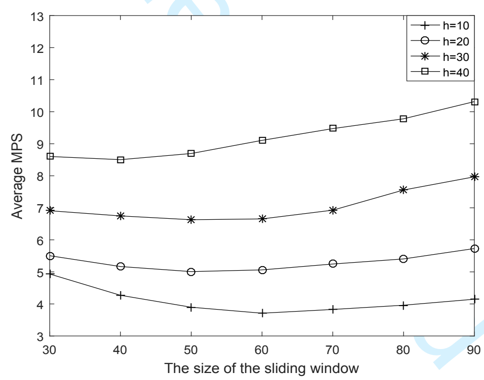

Figure 5 *Average MPS of SIFGM approach with different values of sliding window size L (RUL predictions start from 100 cycles)*

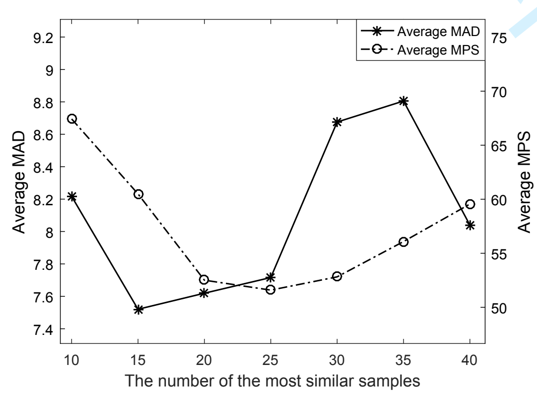

Figure 6 *Average MAD and MPS of SIFGM approach with different numbers of the most similar samples h (RUL predictions start from 60 cycles)*

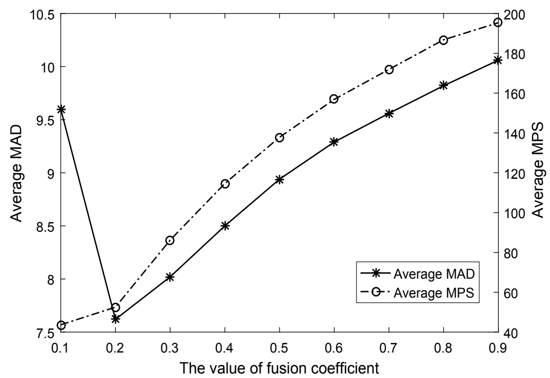

Figure 7. *Average MAD and MPS of SIFGM approach with different values of fusion coefficient* α *(RUL predictions start from 60 cycles)*

As shown in Figure 5, the recommended sliding window size is taken $L = 60$ considering different values of h. Though the preferable number of the most similar samples h seems to be 10 in Figure 5, taking $h = 20$ or $h = 25$ performs better based on the adopted two criteria when RUL predictions start from 60 cycles as shown in Figure 6. Due to the fact that it is hard to define which value is better from Figure 6, and the final results of these two values are only slightly different, the number of the most similar samples is taken as $h = 20$ for convenience. This also suggests that appropriate quantity of the most similar samples could provide more stable prediction in the early degradation stage. Based on the determined L and h, the values of fusion coefficient α is tested, and the recommended value is $\alpha = 0.2$, shown in Figure 7. The performance of the proposed approach is then evaluated in two scenarios.

4.3.1 RUL prediction results of whole lifecycle data

In this part, the degradation trajectories database is generated from 100 training samples using whole lifecycle data. The RUL of the 100 testing samples are predicted from 60^{th} cycle until the engine failure, and the latest 60 observing data points are fed into the retrieval module. The maximum iteration of all models is set to 240 cycles since 96% of the machines operate less than 240 cycles from 60^{th} cycle. The maximum acceptable similarity measurement distance d_{th} is set to 1. The prediction performance of the proposed SIFGM method is evaluated by average MAD and average MPS of 100 samples. It's worth noting that the average prediction error measured by scores defined in (21), however, is not intuitive, because several large errors among the many predictions will completely dominate the final score. For example, a late prediction of 60 cycles will give a penalty score of 402.43, which could happen in early failure prediction of those rather long sequences of the testing samples (Wang *et al.*, 2008). As a comparison, a late prediction beyond 30 cycles gives only 19.09 penalty score and an early prediction beyond 30 cycles gives only 9.05 penalty score for all comparable models. Figures 8 and 9 illustrate the comparison of the trajectory prediction and RUL estimation between the proposed method and its counterparts.

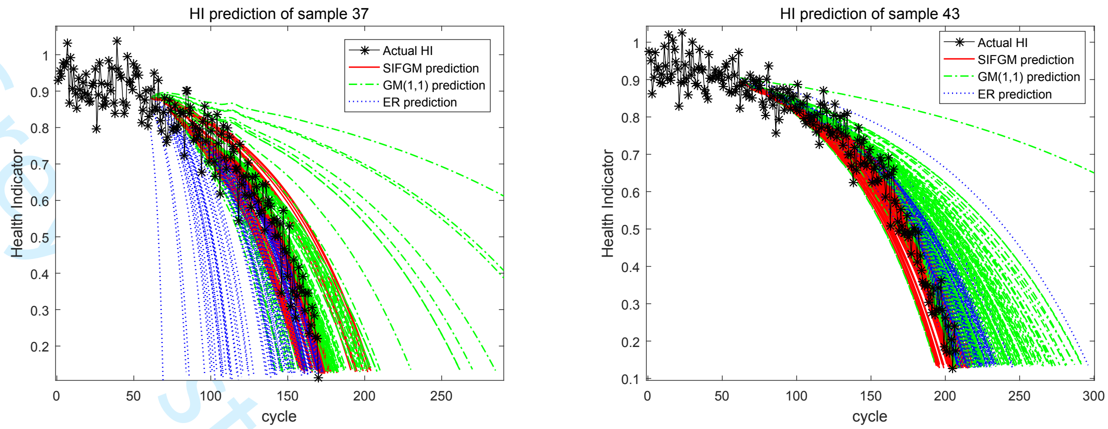

Figure 8. Illustration of the trajectory prediction comparison (whole lifecycle data)

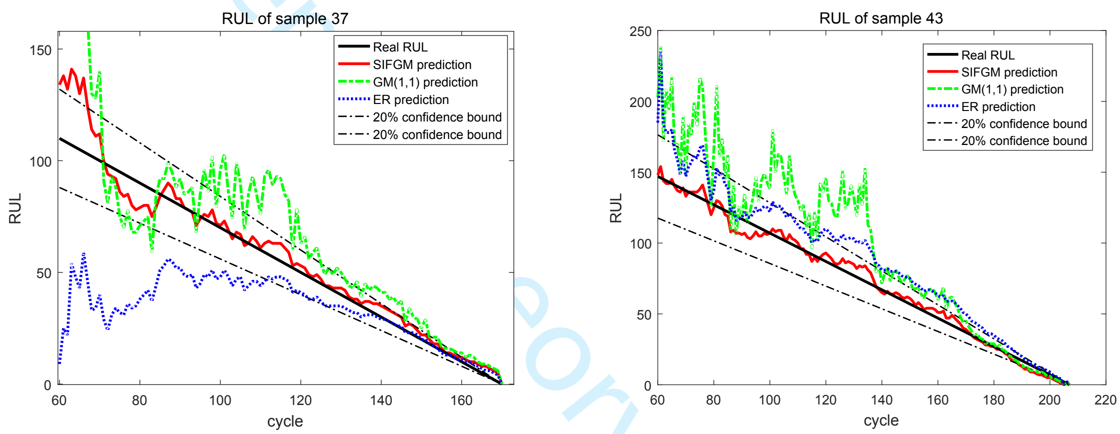

Figure 9. Illustration of the RUL estimation comparison (whole lifecycle data)

Figure 8 presents all predictive trajectories of two samples from 60th cycle until the engine failure using three models, and Figure 9 shows the corresponding RUL at each cycle. Based on above figures, some results and discussion are summarized.

(1)The HI trajectories predicted by the proposed SIFGM method is much closer to the true HI trajectory compared with ER and GM(1,1), and the distribution of SIFGM prediction curves is more concentrated, which proves the robustness of the SIFGM prediction in one aspect.

(2)In the early stage of engine operation, some HI trajectories predicted by GM(1,1) failed to reach the failure threshold at the maximum iteration, while some HI trajectories predicted by ER reach the failure threshold too fast. This is because the GM(1,1) and ER can only extract the information from the historical data, and the fluctuations in the raw data impact greatly on the results. In contrast, SIFGM method adopts the latest information of the test sample and draws on the development trends of similar referential samples at the same time, which leads to a better and more stable prediction.

(3)In the very late stage of engine operation, ER performs better than the other two methods, since ER utilizes all the data points while GM(1,1) is more sensitive to the new information and SIFGM is an extension of GM(1,1). This also shows that three models are effective in short-term prediction near failure, but only SIFGM can provide more reliable long-term predictions.

(4)Almost all RULs predicted by the proposed SIFGM method are well located within the $\pm 20\%$ confidence boundaries of the actual RUL, while only half of predictions by ER are in this area and the majority of predictions by GM(1,1) are out of the boundaries.

The quantitative evaluation results of the aforementioned algorithms are detailed in Table 2. Average MAD and MPS are used to evaluate the performance of three prediction models. The prognostic performance is greatly improved by applying the similar referential information from the perspective of both evaluation criteria. Owing to the fact that average MAD and MPS considers all cycles of RUL prediction, it also reflects the robustness of SIFGM prediction compared with its counterparts. The predictive RUL of the proposed diagnostic and prognostic framework could provide valuable information for spare parts organizing and maintenance scheduling in advance, and prevent serious abnormal conditions and catastrophic failures simultaneously.

Table 2 The performance of the proposed SIFGM and its counterparts (whole lifecycle data)

	SIFMG	GM	ER
Average MAD	7.37	23.94	25.42
Average MPS	2.88	10.55	8.64

4.3.2 RUL prediction results of partial degradation data

For practical applications, large amounts of run-to-failure data are difficult to obtain. Indeed, it is partial degradation data rather than entire run-to-failure data that are commonly available for engineers. In order to test the performance of the proposed method to predict RUL when only part of the whole life cycle data is available for referential machines, part of the whole life data of the original samples are cut off after 180 cycles, resulting in 48 sequences of whole life cycle data and 52 sequences of degradation data without failure. HIs are calculated based on the truncated data and are then stored in the training library for similarity matching, while other experimental conditions are the same as those in 4.3.1. Figure 10 illustrates the comparison of the RUL estimations among the proposed method and its counterparts.

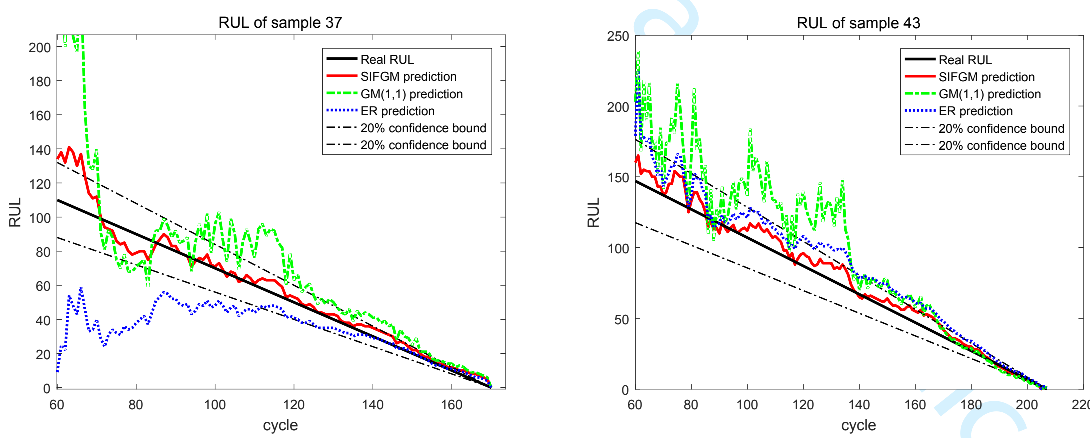

Figure 10. Illustration of the RUL estimation comparison (partial degradation data)

Based on the results shown in Figure 10, the prediction performance of the proposed model does not change much, though the fluctuation of predictive RULs using partial degradation data is larger than that of the results using the whole life fault data. The conclusion that SIFGM shows superiority over its counterparts as summarized in 4.3.1 still holds true, demonstrating that the SIFGM is able to effectively utilize the partial information of the referential samples, while the traditional similarity-based RUL methods fail to do so. The quantitative evaluation results are detailed in Table 3. Compare the results in Table 2 and 3, the performance of the SIFGM is superior to other models, the prediction error is less than half of other models' results. Though the errors become larger in partial degradation data, the predictive RULs are still valuable and stable for parts organizing and maintenance scheduling in advance.

Table 3 The performance of the proposed SIFGM and its counterparts (partial degradation data)

	SIFMG	GM	ER
Average MAD	9.37	23.66	25.41
Average MPS	4.16	10.40	8.59

5. CONCLUSIONS

A trajectory prognostic method based on grey forecasting model and similarity information fusion for RUL prediction is proposed in this paper. The main contribution of this study is to extend the framework of the RUL estimation by introducing metabolism grey model and information fusion in the RUL estimation module. With this improvement, both the latest trend of the test sample and long-term trend of the similar referential samples can be adopted in HI trajectory prognostic, resulting in a more accurate and stable RUL prediction. The case study in two scenarios demonstrates the effectiveness and applicability of the proposed framework for RUL estimation of aircraft engine, which also provides a viable choice of similarity-based RUL prediction when the historical whole lifecycle data are scarce.

To obtain a more accurate prediction, the parameters of the SIFGM method were tested using several discrete values. However, the parameter combination is probably not optimal. More computational experiments will be necessary to make conclusions on the performance of different combinations. The fusion coefficient is taken as a fixed value from beginning to the end, which fails to reflect the reliance degree of monitoring sample on referential information in different operation stages. Future research will discuss the method which can adjust the fusion coefficient dynamically. Theoretical demonstration of the robustness of SIFGM and the application of SIFGM to other real cases need to be further explored as well.

References

Bakir, A., Zaman, M., Hassan, A., and Hamid, M.(2019), "Prediction of remaining useful life for mech equipment using regression", *Journal of Physics Conference*, Vol. 1150, 1-7.

Cao, Y., Liu, S., Fang, Z., and Dong, W.(2017), "Grey Bayesian network model for reliability analysis of complex system", *2017 IEEE International Conference On Grey Systems and Intelligent Services*, pp. 289-294.

Chen, X., and Liu, M.(2015), "Gear remaining useful life prediction based on Grey Neural Network", *2015 IFTOMM World Congress Proceedings*, pp. 1-5.

Chen, Z., Cao, S., and Mao, Z.(2018), "Remaining Useful Life Estimation of Aircraft Engines Using a Modified Similarity and Supporting Vector Machine (SVM) Approach", *Energies*, Vol. 11 No. 1, pp. 28.

Fan, J., and Wang, Y.(2014), "Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis", *Information Sciences*, Vol. 259,pp. 369-379.

García Nieto, P. J., García-Gonzalo, E., Sánchez Lasheras, F., and De Cos Juez, F. J.(2015), "Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability", *Reliability Engineering and System Safety*, Vol. 138,pp. 219-231.

Huang, C., Huang, H., Peng, W., and Huang, T.(2019), "Improved trajectory similarity-based approach for turbofan engine prognostics", *Journal of Mechanical Science and Technology*, Vol. 33 No. 10, pp. 4877-4890.

Lasheras, F., Nieto, P., de Cos Juez, F., Bayón, R., and Suárez, V. (2015), "A hybrid PCA-CART-MARS-based prognostic approach of the remaining useful life for aircraft engines", *Sensors (Switzerland)*, Vol. 15 No. 3,pp. 7062-7083.

Li, J., Zhang H., Yang Z., Li X., and Zhang H., (2018), "Remaining useful life estimation based on similarity of degenerate curves", *Applied Science and Technology*, Vol. 45 No. 5,pp. 82-86, 90.

Li, N., Lei, Y., Guo, L., Yan, T., and Lin, J.(2017), "Remaining Useful Life Prediction Based on a General Expression of Stochastic Process Models", *IEEE Transactions On Industrial Electronics*, Vol. 64 No. 7, pp. 5709-5718.

Li, X., Duan, F., Loukopoulos, P., Bennett, I., and Mba, D.(2018), "Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor", *Control Engineering Practice*, Vol. 72, pp. 177-191.

Li, X., Miao, J., and Ye, J.(2015), "Lithium-ion battery remaining useful life prediction based on grey support vector machines", *Advances in Mechanical Engineering*, Vol. 7 No. 12, pp. 1-8.

Li, X., Yang, X., Yang, Y., Bennett, I., Collop, A., and Mba, D.(2019), "Canonical variate residuals-based contribution map for slowly evolving faults", *Journal of Process Control*, Vol. 76, pp. 87-97.

Li, X., Yang, X., Yang, Y., Bennett, I., and Mba, D.(2019), "A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines", *Applied Soft Computing Journal*, Vol. 82, pp.105564.

Li, X., Yang, Y., Bennett, I., and Mba, D.(2019), "Condition monitoring of rotating machines under time-varying conditions based on adaptive canonical variate analysis", *Mechanical Systems and Signal Processing*, Vol. 131,pp. 348-363.

Liang Z., Jiang. H., Zhou B.(2018), "Multi-variable similarity-based information fusion method for remaining useful life prediction", *Computer Integrated Manufacturing Systems*, Vol. 24 No. 4, pp. 813-819.

Liu, S., Yang, Y., and Forrest, J. (2016), Grey Data Analysis: Methods, Models and Applications. Singapore, Springer.

Liu, Y., Hu, X., and Zhang, W.(2019), "Remaining useful life prediction based on health index similarity", *Reliability Engineering and System Safety*, Vol. 185, pp. 502-510.

Liu, Z., Wang, Q., Song, C., and Cheng, Y.(2017), "Similarity-Based Difference Analysis Approach for Remaining Useful Life Prediction of GaAs-Based Semiconductor Lasers", *IEEE Access*, Vol. 5, pp. 21508-21523.

Peng, W., Shen, L., Shen, Y., and Sun, Q.(2018), "Reliability analysis of repairable systems with recurrent misuse-induced failures and normal-operation failures", *Reliability Engineering and System Safety*, Vol. 171, pp. 87-98.

Peng, Y., and Dong, M.(2011), "A hybrid approach of HMM and grey model for age-dependent health prediction of engineering assets", *Expert Systems with Applications*, Vol. 38 No. 10, pp. 12946-12953.

Ruiz-cárcel, C., Cao, Y., Mba, D., Lao, L., and Samuel, R. T.(2015), "Statistical process monitoring of a multiphase flow facility", *Control Engineering Practice*, Vol. 42, pp. 74-88.

Saxena, A., Goebel, K., Simon, D., and Eklund, N.(2008), "Damage propagation modeling for aircraft engine run-to-failure simulation", IEEE conference on prognostics and health management, pp. 1-9.

Wang, R. Z., Zhang, X. C., Tu, S. T., Zhu, S. P., and Zhang, C. C.(2016), "A modified strain energy density exhaustion model for creep-fatigue life prediction", *International Journal of Fatigue*, Vol. 90, pp. 12-22.

Wang, T., Yu, J., Siegel, D., and Lee, J.(2008), "A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems", *2008 International Conference on Prognostics and Health Management*, pp. 1-6.

Wook, S., Lee, C., Lee, J., Hyun, J., and Lee, I.(2005), "Fault detection and identification of nonlinear processes based on kernel PCA", *Chemometrics and Intelligent Laboratory Systems*, Vol. 75, pp. 55-67.

Xu, J., Wang, Y., and Xu, L.(2014), "PHM-Oriented Integrated Fusion Prognostics for Aircraft Engines Based on Sensor Data", *IEEE Sensors Journal*, Vol. 14 No. 4, pp. 1124-1132.

Yan, J., Ko, M., and Lee, J.(2004), "A prognostic algorithm for machine performance assessment and its application", *Production Planning & Control*, Vol. 15 No. 8, pp. 796-801.

Yang, X., Fang, Z., Yang, Y., Mba, D., and Li, X.(2019), "A novel multi-information fusion grey model and its application in wear trend prediction of wind turbines", *Applied Mathematical Modelling*, Vol. 71, pp. 543-557.

Zhang, L., Mu, Z., and Sun, C.(2018), "Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Exponential Model and Particle Filter", *IEEE Access*, Vol. 6, pp. 17729-17740.

Zhao, Z., Liang, B., Wang, X., and Lu, W.(2017), "Remaining useful life prediction of aircraft engine based on degradation pattern learning", *Reliability Engineering & System Safety*, Vol. 164, pp. 74-83.

Zhou, Z., Huang, Y., Lu, Y., Shi, Z., Zhu, L., Wu, J., and Li, H.(2014), "Lithium-ion battery remaining useful life prediction under grey theory framework", *Proceedings of 2014 Prognostics and System Health Management Conference*, pp. 297-300.

Zhu, S. P., Huang, H. Z., He, L. P., Liu, Y., and Wang, Z.(2012), "A generalized energy-based fatigue-creep damage parameter for life prediction of turbine disk alloys", *Engineering Fracture Mechanics*, Vol. 90, pp. 89-100.

Zio, E., and Di Maio, F.(2010), "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system", *Reliability Engineering and System Safety*, Vol. 95 No. 1, pp. 49-57.