Aircraft Failure Rate Prediction Method Based on CEEMD and Combined Model

Wenqiang Li and Ning Hou

School of Mechatronics Engineering, Shenyang Aerospace University, Shenyang 110136, China

Correspondence should be addressed to Wenqiang Li; lwqsss@163.com

Received 12 March 2022; Revised 28 April 2022; Accepted 25 May 2022; Published 24 June 2022

1. Introduction

As one of the core and key technologies in many fields, prediction technology has been widely used, such as in the fields of geology [1–4], transportation [5], water conservancy [6], and wind energy [7–16]. It plays an important supporting role in decision making and management. As a highly integrated and complex system, aircraft may break down at any time in harsh working environment for a long time. In order to prevent flight accidents and even air crashes caused by aircraft failures, people are actively looking for more efficient aircraft health management and maintenance decision-making solutions to improve flight safety and reliability. According to statistics, aircraft failures in the field of civil aviation show a downward trend year by year and are becoming more and more safe. However, there are still some factors threatening flight safety. With the rapid development of aircraft manufacturing industry and the continuous improvement of aircraft performance, military aviation has higher and higher requirements for reducing aircraft failures. Scientific prediction of aircraft failures can make scientific decisions on aviation maintenance, diagnosis, and health management. It is an important component of improving maintenance support capabilities. It can effectively improve aircraft safety, enhance reliability and maintainability, and reduce maintenance and insurance-related costs. Failure prediction is generally divided into failure time prediction [17], life prediction [18], and failure rate prediction [19]. Aircraft failure rate is an important parameter of reliability, maintainability, and supportability of aviation equipment. Due to the complexity of aircraft fault cross-linking, strong uncertainty and randomness of fault causes,
few fault samples, and other problems, the modeling and prediction of aircraft failure rate has always been a research hotspot of many scholars, and the research in this area has great theoretical value and practical significance.

In the past few decades, researchers have proposed a variety of models and methods for predicting failure rates, which are mainly divided into two categories, namely, model-based failure rate prediction and data-driven method failure rate prediction [20]. Due to the complex composition of the aircraft system and the cross-linking of each system level, unit level, and component level, there are many interferences and mutual influences in the use process, so it is difficult to accurately model the aircraft failure rate distribution function, and improper modeling will also produce large errors in prediction, thus limiting the effective application of model-based failure rate prediction methods. Therefore, it is difficult to use this method to predict the aircraft failure rate [21]. The application of data-driven prediction methods is more flexible. The main advantage is that the trained model data can be easily updated with new and updated data. Therefore, data-driven failure rate prediction has been widely concerned and applied [22].

Data-driven aircraft failure rate prediction models can be divided into single-item models and combined models. The related literature is shown in Table 1. Single-item models usually include machine learning models, statistical models, and grey models. Al-Garni et al. [23] established an artificial neural network model for predicting the failure rate of Dash-8 aircraft tires, used six years of historical data for model construction and verification, and applied the artificial neural network model to verify the applicability of future tire failure rate prediction. At the same time, the artificial neural network model was applied in the reliability analysis of Boeing 737 tires, and the failure rate prediction model and verification were used based on two years of data [24]. The application of artificial neural network (ANN) is relatively successful, but it is still limited by the slow learning speed and the need for a lot of numerical training, so it is difficult to obtain satisfactory prediction accuracy. Some scholars use improved neural network [25] and genetic neural network [26] to predict the aircraft failure rate, which has more advantages in aircraft failure rate prediction. Because of the strictness of the algorithm, the prediction results are more accurate, but the structure of the model is too complex, and modeling takes more time. At the same time, when the amount of failure rate data is small, the prediction effect of the above model is not ideal. Also, for the failure rate prediction problem, the support vector machine (SVM) prediction method [27] is used for prediction. SVM can use the kernel function to solve non-linear problems, which is suitable for small sample prediction, but it is difficult to implement for large-scale training samples. Yang et al. [28] proposed a seasonal-integrated moving average autoregressive (ARIMA) model and analyzed and compared the forecast results. Seasonally integrated moving average autoregressive SARIMA (0, 1, 1) (0, 1, 1)_{12} model is suitable for aircraft failure rate prediction, which verifies the feasibility and effectiveness of the model. Li and Kang [29] used the autoregressive moving average (ARMA) model to predict the failure rate. Zhang et al. [30] proposed the generalized regression equation based on Weibull distribution to predict the failure of aircraft auxiliary power unit and verified the performance of the proposed model using the three-year dataset provided by China Southern Airlines. When the statistical model encounters large changes in the outside world, it often has large deviations, resulting in low accuracy. The ARIMA model has effectively expanded from stationary time series to non-stationary cases. It is an
effective integration of autoregressive, difference, and moving average models. It is a better time series statistical model and has been widely used. Li et al. [31] aimed at the problem of high failure rate of Cessna 172 basic aircraft in operation, and a GM (1,1) fault prediction method based on grey theory was proposed. This method can effectively predict the failure of Cessna 172 basic aircraft and realize the prediction of aircraft failure. However, this method sometimes has large deviation and failure. Also, the failure rate is predicted by using the non-equidistant grey Verhulst modified model [32]. However, the grey prediction is not suitable for a large number of data and cannot be used for long-term prediction. The grey Verhulst model can solve the problems of less historical data, low sequence integrity, and reliability. It can generate irregular original data to obtain strong sequence. It is suitable for non-monotonic swing sequence. Compared with other grey models, it has certain advantages and can improve the overall prediction accuracy. Although the above single prediction model or method has achieved good prediction results, for the aircraft system, due to the complexity of its system composition and failure mechanism, there is a strong coupling between different components, and the single failure prediction method has its own advantages and disadvantages. Due to its own limitations and applicable conditions, the use of a single model for fault prediction has corresponding limitations and shortcomings, and the existing prediction methods are difficult to achieve the ideal prediction effect. It can overcome the shortcomings of traditional single prediction methods and improve the accuracy of prediction. Therefore, some scholars propose a combined model to predict the failure rate.

The essence of combined model is to regard various individual models as fragments representing different information, disperse the unique uncertainty of individual prediction, and reduce the overall uncertainty through the integration of information, so as to improve the prediction accuracy. Combined models can be divided into two types: multi-model combination and decomposition integration combination. Multi-model combination combines two or more single models to complete the prediction of failure rate. Yang et al. [33] proposed an aircraft failure rate prediction method based on the Holt–Winters seasonal model, analyzed the seasonal time series data, and analyzed the application of the Holt–Winters seasonal model in aircraft failure rate prediction through an example. Wang and Yuan [34] proposed an autoregressive model based on neural network residual correction, and the combined model was used to predict the failure rate of Boeing aircraft. Practice has proved that the model is suitable for short-term failure rate prediction. Celiokmih et al. [35] collected maintenance and failure data of aircraft equipment over a two-year period and employed a multi-layer perceptron (MLP) as a machine learning combination of artificial neural network (ANN), support vector regression (SVR), and linear regression (LR). The algorithm was evaluated, and the results showed that the model improved the prediction accuracy of faults. Gu et al. [36] proposed the interpolation fitting transfer learning (ITF) algorithm of adaptive weight for the prediction of airborne equipment failure rate in different working environments and achieved certain prediction results. Zhang et al. [37] comprehensively used a variety of data analysis techniques such as support vector regression (SVR), multiple regression (MLR), and principal component analysis (PCA) and proposed a comprehensive prediction method. The mathematical relationship between the influencing factors and failure rate was studied when the failure rate was collected from 2000 to 2003, and the prediction results verify the effectiveness of this method. At the same time, many researchers have also proposed some other combination forecasting methods, including random forest combination [38], ARMA-BP combination [39], generalized weighted least squares combination [40], optimal combination [41], etc. The advantage of the multi-model combination method is that it combines the advantages of various prediction models and reduces the prediction error, so it obtains better prediction accuracy and prediction effect. Although it combines the advantages of other single prediction models, it is difficult to optimize the single model involved in the combined model and determine the parameters such as the combined weight coefficient. At the same time, the combined prediction model is too complex and has limitations in dealing with the non-linearity and instability of aircraft failure rate, which needs to be deeply discussed in practical application. Due to the non-linearity and non-stationarity of aircraft failure rate series, some scholars apply the idea of signal decomposition method in the field of signal processing to aircraft failure rate prediction and process the time series before prediction to reduce its non-smoothness. Then, optimize the model and propose a combined model based on decomposition integration. The combined model of decomposition and integration is to decompose the original data to generate different characteristic components. After these different components are generated, the same or different models are used to predict. Finally, the predicted values of each component are superimposed and integrated to form the final predicted values. For example, Wang et al. [42] proposed a fault prediction method based on empirical mode decomposition (EMD) and least squares support vector machine (LS-SVM). The effectiveness and superiority of the proposed method are verified by using the classical fault rate prediction example of Boeing aircraft. Wang and Lu [43] proposed a failure rate prediction method based on EMD and RVM-GM model and selected the failure rate data of Boeing 757-700 aircraft as the research object for multiple sets of data within two years. The MAPE value reached 8%, which effectively improved the prediction accuracy. Xu et al. [44] proposed a prediction method based on correlation vector empirical mode decomposition (RVEMD) and grouped method of data handling (GMDH) reconstruction. 700 kinds of aircraft have been researched on multiple sets of failure rate data within two years, and the MAPE value reached 4.78%. The method based on RVEMD and GMDH reconstruction can well reflect the change law of failure rate and has high prediction accuracy. The decomposition method has the first mock exam for the submodel after the breakdown rate of aircraft, but most of the existing decomposition and integration models adopt unified model
prediction instead of selecting suitable models for each sub-
mode. The complementary set empirical mode (CEEMD) method can effectively improve the smoothness of time series and has high adaptability. It can effectively solve the mode aliasing problem of the EMD method and the problem that the white noise added in the set empirical mode decomposition (EEMD) cannot be neutralized. It can effectively solve the uncertainty of the decomposition level of the variational mode decomposition (VMD) method, which shows the unique advantages of the CEEMD method. The CEEMD method can be used to decompose the aircraft failure rate prediction sequence, and there is little research in this field.

The above researchers have carried out research on the prediction of aircraft failure rate by the combined model. The application of combined model in various sciences shows that the combined model can make full use of the advantages of various methods and models and achieve good prediction results. Compared with the single model, the prediction results are improved to different degrees, and better prediction performance is obtained. However, the combined forecasting model also has certain shortcomings: first, there is a certain blindness in the selection of forecasting models. Since different forecasting models have their own applicable objects, choosing different models will have a certain impact on the subsequent forecasting results. Secondly, the weight coefficients of the combined forecasting model all use the same value, that is, the direct superposition method is used for the forecast values of different forecasting models, and the key role of different weights on the combined model is not analyzed. Finally, the prediction accuracy still needs to be further improved to achieve the ideal prediction effect. Because the aircraft failure rate is affected by various factors such as flight state, climatic conditions, and comprehensive support capabilities and the cause of the failure is highly uncertain, the selection of the modeling method for the combined prediction model of the aircraft failure rate is still a current research focus.

In view of this, in order to solve the existing deficiencies and improve the prediction performance, this paper will focus on the research on the failure rate prediction method of the combined model and propose a combined model to improve the failure prediction accuracy of the aircraft. The major contributions of this research are as follows:

1. The CEEMD method is introduced to decompose the time series of the collected aircraft failure rate. Using the decomposition and integration method, higher prediction accuracy, better directional prediction, and higher robustness can be achieved, and the Hurst index is used to determine the appropriate correlation prediction model for each decomposition component.

2. Select the ARIMA model and the grey Verhulst model as the single-item model of the combined forecasting model to perform combined forecasting of the group failure rate.

3. Use an entropy weight method to solve the coefficients of the combined model to optimize the weight coefficients of each forecasting model, thereby constructing the final combined model, improving the model’s explanatory ability and forecasting performance to meet the actual needs.

The rest of this paper is organized as follows. Section 2 introduces the theoretical introduction based on CEEMD and combined models, analyzes the modeling process in detail, and gives the methods and steps of CEEMD, ARIMA model, and grey Verhulst model. Section 3 uses the experimental example to study the model. Through the collection of experimental data, the established model is used to carry out detailed research and discussion, and the evaluation index of prediction performance is given at the same time. Section 4 compares and analyzes the prediction results and accuracy of the established model with other models in detail. Section 5 is devoted to discussion. Finally, Section 6 gives the conclusions and corresponding suggestions.

2. CEEMD and Combination Models

The aircraft mainly includes many parts such as the power system, flight control system, hydraulic system, fuel system, and communication system. Among them, the fuel system, as an important functional subsystem of the aircraft, undertakes the functions of stable and reliable fuel supply to the engine, control of the position of the center of mass, and cooling of other systems. The aircraft fuel system is composed of a fuel tank, actuator, fuel pipeline, and fuel control system. If the fuel control system fails, the aircraft engine will fail and the aircraft will fail, thus affecting the flight performance and even the occurrence of safety accidents. In this paper, the failure rate of the aircraft fuel control system is predicted, aiming at predicting the failure of the fuel control system through the predictive model, predicting the aircraft that may fail in time, conducting inspection, maintenance, and health management in advance, and improving the safety of the aircraft sturdiness and reliability.

2.1. Modeling Process. In order to study the prediction model of aircraft failure rate, the complementary set empirical mode decomposition (CEEMD) and combined model method are used, including ARIMA model and grey Verhulst model, to solve the coefficients of combined models by the entropy weight method, and finally a combined prediction model is formed. The structure of the prediction model method is shown in Figure 1, which includes the following three parts:

(1) Select typical aircraft systems and fault collection: select typical fuel systems from aircraft systems to carry out research, specifically study the failure rate of the control system in the fuel system to predict, analyze the failure situation of an aircraft fuel control system in detail, and collect the failure rate data; the data are obtained from two aspects: on the one hand, the failure rate is collected from the historical fault data over the years; then, the failure rate of the aircraft fuel control system is calculated, and the time
series of failure rate is obtained, and the failure prediction research is carried out with the percentage of failure rate as the overall failure rate.

(2) Implementation of complementary set empirical mode decomposition and combination algorithm. Firstly, IMF fluctuation term and RF trend term of fault rate series are separated by CEEMD, and the single prediction model is determined by Hurst index calculation. Secondly, the ARIMA prediction model of each IMF component and the grey Verhulst prediction model of RF trend term component are established. Finally, the combined model based on each ARIMA and grey Verhulst model is realized to predict the failure rate value (as the test data parameter) and failure rate value (as the output parameter). Based on the combined modeling theory of entropy weight method, the separated coefficients are solved to optimize the weight coefficients of each prediction model, and the combined prediction model is deduced to obtain the final prediction results, which provides an effective model for predicting the aircraft failure rate.

(3) Accuracy analysis and comparison of the model: in order to evaluate the prediction performance of the combined model, it is compared with other models. Using a variety of accuracy indicators as evaluation criteria, compare the accuracy of other model validation data with the results of the proposed combined prediction model, carry out systematic comparative analysis and research, and carry out specific application and verification to achieve the corresponding prediction effect.
2.2. CEEMD. CEEMD is an improved algorithm based on EMD and EEMD; the empirical mode decomposition (EMD) algorithm proposed by Huang et al. in 1998 is a method for dealing with non-linear signals, which has shortcomings such as modal aliasing and end-point effects. In view of the shortcomings of EMD, Huang et al. proposed ensemble empirical mode decomposition (EEMD) in 2009, which added white noise sequences on the basis of EMD to overcome the modal aliasing that appeared in EMD. However, in each decomposition process, adding dissimilar white noise makes EEMD have the problems of noise residue, many iterations, and slow operation efficiency. In 2010, Yeh et al. [45] proposed CEEMD based on EEMD. Different from EMD and EEMD, CEEMD uses adding different white noise to the signal and EMD processing, respectively, to remove the adverse effect of adding white noise to EEMD, reduce reconstruction error as much as possible, overcome the noise residue, and improve the calculation efficiency, and the effect will be better when it is used for signal decomposition. $S(n)$ is defined as the original sequence, and $E_k(\cdot)$ and $IMF_k$ represent the K-th modal components generated by EMD and CEEMD, respectively. The specific steps are as follows:

- Step 1: CEEMD makes I test on the signal, and the first modal component is obtained by EMD algorithm and recorded as

$$IMF_1(n) = \frac{1}{I} \sum_{i=1}^{I} IMF_1^i(n).$$

- Step 2: when $k=1$, the current unique margin is obtained and recorded as

$$r_1(n) = S(n) - IMF_1(n).$$

- Step 3: conduct I tests ($i = 0, 1, 2, \ldots, I$) and decompose continuously $r_1(n) + \epsilon_i E_1(\nu(n))$ until the first modal component is obtained. On this basis, calculate the second modal component:

$$IMF_2(n) = \frac{1}{I} \sum_{i=1}^{I} E_i(r_1(n) + \epsilon_i E_1(\nu(n))).$$

- Step 4: for each of the remaining stages, that is, $k = 2, 3, \ldots, K$, according to Step 3 above, calculate the $k$-th residual signal and the $(k+1)$-th modal component:

$$r_k(n) = r_{k-1}(n) - IMF_k(n),$$

$$IMF_{k+1}(n) = \sum_{i=1}^{I} E_i(r_k(n) + \epsilon_i E_k(\nu(n))).$$

- Step 5: repeat Step 4 until the residual signal can no longer be decomposed, then the algorithm stops, and the final residual signal is

$$R(n) = S(n) - \sum_{k=1}^{K} IMF_k.$$ 

Therefore, the original signal sequence is finally decomposed into $S(n) = \sum_{k=1}^{K} IMF_k + R(n)$.

2.3. Hurst Index. There is a known time series $x_1, x_2, \ldots, x_n$. The average value is $\overline{x}_n = 1/n \sum_{t=1}^{n} x_t$, and the standard deviation is $S_n = \sqrt{1/n \sum_{t=1}^{n} (x_t - \overline{x}_n)^2}$. For $1 \leq t \leq n$, the cumulative deviation is $x_{t,n} = \sum_{i=1}^{t} (x_i - \overline{x}_n)^2$, and range is $R = \max|x_t| - \min|x_t|$. The Hurst index refers to the coefficient $H$ satisfying the following equation [46], where $C$ is a constant. By specifying a value of $C$ (e.g. 0.5), the Hurst index can be obtained.

2.4. ARIMA Model. The autoregressive integrated moving average (ARIMA) model was proposed by Box and Jenkins [47]. ARIMA consists of three parts, namely, AR, I, and MA, in which AR is the autoregressive model, I represents the difference, and MA is the moving average model. As a kind of time series analysis method, the ARIMA model can predict future values according to historical values. ARIMA is the most widely and commonly used time series prediction. Its modeling is easier than more complex methods. Because of its simplicity and stability, it has been widely used in forecasting, social, economic, and engineering fields. The general expression of the ARIMA model is

$$\Phi(L) (1 - L)^d y_t = \epsilon_t + \Theta(L) \epsilon_t.$$  

Among them, $\Phi(L) = 1 - \lambda_1 L - \lambda_2 L - \ldots - \lambda_p L^p$ is the $P$-order autoregressive coefficient polynomial, $\Theta(L) = 1 + \theta_1 L + \theta_2 L + \ldots + \lambda_q L^q$ is the $P$-order moving average coefficient polynomial, $L$ is the lag operator, $\lambda$ and $\theta$ are the estimated values of the respective variables, $y_t = c + y_{t-1} + \mu_t$ is a d-order single integer sequence, $c$ is a constant, $\mu_t$ is a stationary sequence, $t = 1, 2, \ldots, T$, and $\epsilon_t$ is a white noise sequence with a mean of 0 and a variance of $\sigma^2$. The ARIMA model prediction steps are as follows:

- Step 1: stationarity test. The stability of the data is preliminarily judged according to the sequence diagram, autocorrelation function diagram, and partial autocorrelation function diagram of the time series.

- Step 2: smoothing. Stationarize the non-stationary time series data until the processed data can pass the stationarity test.

- Step 3: model identification and order determination. Establish a corresponding time series model according to the identified features. After smoothing, if the partial autocorrelation function is truncated and the autocorrelation function is tailing, the AR model is established; if the autocorrelation function is tailing, the autocorrelation function is tailing; if the function is truncated, the MA model is established. If both the partial autocorrelation function and the autocorrelation function are tailing, the sequence is suitable for the ARMA model.
Step 4: model parameter estimation. Estimate the model parameters and test the estimated values to determine the reliability of the parameter estimates.

Step 5: model checking. In the white noise test, it is assumed that the model judges whether the residual sequence is a white noise sequence.

Step 6: prediction of the model.

After building the ARIMA(p, d, q) model, a predictive model is formed to predict future values based on historical values.

2.5. Grey Verhulst Model. The grey system theory was first proposed by Professor Deng Julong to deal with the "small sample, poor information" system. The grey prediction method is based on the grey system theory and is a method for fuzzy prediction of the system by establishing approximate differential equations. It includes GM(1, 1), GM(1, n), GM (0, n), Verhulst, and other models. The grey Verhulst model is an important part of the grey system theory and is a single sequence first-order non-linear dynamic model [48].

The grey Verhulst model prediction steps are as follows:

Step 1: carry out the accumulation and generation operation.

Let the original sequence be known as $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$ and perform a cumulative generation (1-AGO) on the original sequence to obtain a new sequence $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$, where

$$x_k^{(1)} = \sum_{i=1}^{k} x_i^{(0)}, \quad (k = 1, 2, \ldots, n). \quad (7)$$

Step 2: generate a sequence of adjacent mean values. $x^{(1)}$ is generated as the next-to-neighbor mean $z^{(1)} = (z_1^{(1)}, z_2^{(1)}, \ldots, z_n^{(1)})$, where $z_k^{(1)} = 1/2 (x_k^{(1)} + x_{k-1}^{(1)})$, $k = 2, 3, \ldots, n$.

Then, the grey Verhulst model is [49]

$$x^{(0)} + az^{(1)} = b (z^{(1)})^2. \quad (8)$$

The corresponding whitening equation is

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = bx^{(1)^2}. \quad (9)$$

In formulas (7) and (8), a and b are parameters and t is time.

Step 3: establish an approximate time response sequence.

$$B = \begin{bmatrix} -z_1^{(1)} & (z_2^{(1)})^2 \\ -z_2^{(1)} & (z_3^{(1)})^2 \\ \vdots & \vdots \\ -z_n^{(1)} & (z_n^{(1)})^2 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix}.$$ 

Then, the least squares estimation parameters of the grey Verhulst model are listed as

$$\hat{a} = (B^T B)^{-1} B^T Y. \quad (10)$$

Then, the solution of the whitening equation (time response function) is

$$x^{(1)}(t) = \frac{ax_0^{(1)}}{bx_0^{(1)} + (a - bx_0^{(1)})e^{at}}. \quad (11)$$

The time response sequence for the grey Verhulst model is

$$\tilde{x}^{(1)}_{k+1} = \frac{ax_0^{(1)}}{bx_0^{(1)} + (a - bx_0^{(1)})e^{at}}. \quad (12)$$

$x_0^{(1)}$ will be taken as $x_1^{(0)}$; then, (11) becomes

$$\tilde{x}^{(1)}_{k+1} = \frac{ax_1^{(0)}}{bx_1^{(0)} + (a - bx_1^{(0)})e^{at}}. \quad (13)$$

Step 4: accumulate and restore the prediction model. $\tilde{x}^{(1)}_{k+1}$ carries out the cumulative reduction and reduction operation, and the prediction model of the grey Verhulst can be obtained as

$$\tilde{x}^{(0)}_{k+1} = \tilde{x}^{(1)}_{k+1} - x_k^{(1)}. \quad (14)$$

2.6. Solving the Coefficients of the Combined Model by the Entropy Weight Method. Based on CEEMD, the failure rate sequence is decomposed into IMF term and RF residual term of each order and then predicted according to the characteristics of each component sequence, and the predicted values of the components are added to obtain the final prediction result. However, this method is likely to cause omission or loss of some information, thereby affecting the accuracy of prediction. Therefore, it is necessary to take into account the physical meaning behind each component and systematically mine the implicit information hidden in the IMF component and the RF remainder to determine the importance of each factor. At the same time, different single-item models also have different predictive capabilities, so the weight coefficient of the single-item models should not simply take the average of the number of single-item models in combined forecasting but give different weights according to the predictive ability of different single-item models; single-item models with strong predictive performance are assigned larger weights. Considering the above two reasons, an intelligent combination model based on entropy weight method index weight calculation is constructed, so that the prediction of the combination model is more accurate, the generalization ability of the model is strengthened, and the purpose of improving the prediction accuracy is finally achieved.

The basic idea of the entropy weight method is to use the discrete degree of the index, that is, the information entropy, to measure the importance of the index [50]. The weight of each index is determined by applying the entropy weight
method. Based on the objective data, the index weight is determined according to the entropy value of the data. The weight of the indicators can represent the differential impact of objective data on the results. By using the entropy weight method, calculate the weights of the predicted values corresponding to IMF1, IMF2, IMFm, and RF decomposed by CEEMD. The entropy weight method determines the weight steps as follows:

Step 1: data normalization.
To quantify the high-frequency and low-frequency data decomposed by CEEMD, assuming that there are $m$ evaluation indicators and $n$ evaluation objects, the original data matrix is $A = (a_{ij})_{mn}$; after normalization, the matrix is obtained on $B = (b_{ij})_{mn}$, $b_{ij} = a_{ij}/\max\{a_{ij}\}$.

Step 2: calculation of entropy.
The entropy of the $i$-th index is $h_i = -k \sum_{j=1}^{n} f_{ij} \ln f_{ij}$, where $f_{ij}$ is the standardized value of the index proportion of the $j$-th evaluation object under the $i$-th index, $f_{ij} = b_{ij}/\sum_{j=1}^{n} b_{ij}$, and $k$ is the adjustment coefficient, $k = 1/\ln n$.

Step 3: calculation of entropy weight.

After calculating the entropy of the $i$-th indicator, the entropy weight of the $i$-th indicator is calculated as $\omega_i = 1 - h_i/m - \sum_{i=1}^{m} h_i$, $(0 \leq \omega_i \leq 1, \sum_{i=1}^{m} \omega_i = 1)$, and $\omega = [\omega_1, \omega_2, \ldots, \omega_m]^T$ is the weight of the corresponding prediction values of IMF1 and IMF2. IMFm and RF were decomposed by CEEMD. The ARIMA model was set from IMF1 to IMF2. The predicted value at time $t$ is $\hat{X}(t) = \sum_{i=1}^{m}(\omega_i x_i(t))$, $\sum_{i=1}^{m} \omega_i = 1$.

The execution steps of the combined model are shown in Figure 2. First, the data of the aircraft failure rate are collected in detail, with the failure rate percentage data as the benchmark, and the collected dataset is divided into two parts: training data and test data. Secondly, using complementary ensemble empirical mode decomposition (CEEMD) aircraft failure rate data, the ARIMA model and grey Verhulst model are used for training, and the failure rate is predicted to obtain the corresponding actual value and predicted value of failure. Again based on the ARIMA and grey Verhulst forecasting models, the predicted failure rate values are treated as input variables to the combined model, which forms the model to generate input and output datasets. Finally, the entropy weight method is used to solve the coefficient, the basis of the combination weight is determined, and the final combination function is obtained, and the corresponding accuracy evaluation index is used for comparison and analysis, and the overall prediction of the failure rate is completed.

3. Experimental Example

3.1. Experimental Example Dataset

3.1.1. Composition and Failure Factor Analysis of Aircraft Fuel Control System. As an important part of the aircraft system, the aircraft fuel control system plays a key role. Its composition and fault factors are shown in Figure 3. The aircraft fuel control system generally includes two parts: the fuel supply control system and centroid position control system. The fuel supply control system controls the specific fuel consumption of each fuel tank according to the fuel consumption of the engine, so as to meet the fuel consumption demand of the engine; the centroid control system controls the centroid position of the aircraft according to the needs, reduces the control burden of the control surface, and ensures that the centroid position of the aircraft is basically unchanged. The fuel supply control system and centroid control system can calculate the aircraft mass and centroid position information in real time through the integration of mass and centroid position control system, which can be provided to the aircraft dynamic system and other subsystems to meet their needs. The fuel supply control system is composed of a converter, fuel control distributor, and fuel tank control unit. The centroid position control system includes a converter, fuel control switch, and fuel tank control unit. According to statistical analysis, there are many factors causing the failure of aircraft fuel control system. The factors affecting the failure rate include the following aspects. (1) Flight mission: including the length of flight, the number of take-off and landing, and the number of missions. If the flight mission is heavy, the probability of failure will be greater. (2) External environmental conditions: mainly including weather conditions, ambient temperature, humidity, and other factors. Abnormal weather conditions and temperature and humidity will have a certain impact on the failure probability of fuel control system. The worse the weather conditions meaning that the environment beyond the
reasonable temperature and humidity range, the greater the probability of failure. (3) Product quality: devices with good product quality, high system integration, and high reliability will reduce the failure rate of fuel control system to a certain extent. At the same time, products with poor quality and low stability will also increase the failure rate of fuel control system. (4) Maintenance quality: when the fuel control system is in use, it needs regular maintenance and corresponding maintenance to improve the system reliability and reduce the failure rate. In some previous aircraft failure rate prediction studies, many factors such as flight time, ambient temperature, ambient humidity, product quality, and support ability need to be comprehensively considered, which increases the complexity of the prediction model. At the same time, there are many coupling relationships between these influencing factors, and some influencing factors are difficult to obtain accurately in practice. Based on this, this paper will only consider the main failure rate of aircraft fuel control system to be predicted, without considering other influencing factors.

3.1.2. Fault Data Acquisition and Analysis. The failure data of aircraft fuel control system are tested and evaluated by collecting the failure data of aircraft fuel control system for 48 months from March 2014 to February 2018. The data are divided into two subsets: the first 80% of the data samples are used as the training dataset, and the last 20% are used for the test dataset. The training dataset is used for model training, and the test dataset is used to evaluate the performance of the established prediction model to verify the effectiveness of the proposed model. Using the data of the first 38 months of the sample from March 2014 to April 2017, the prediction model is established as the modeling input, the prediction is verified by using the failure rate of the last 10 months from May 2017 to February 2018, and then the prediction results are compared with the real data. The collected failure rate time sequence is shown in Figure 4, and the statistical parameters of the dataset including average value, standard deviation, maximum value, minimum value, skewness, kurtosis, and median value are obtained at the same time. As shown in Table 2, combined with Figure 3 and Table 2, it can be seen that the failure rate data of the fuel control system fluctuate in a smaller range, which is a small sample, non-linear, and non-stationary curve. The maximum failure rate is 5.7%, the maximum failure rate data does not exceed 6%, the minimum value is 2.1%, and the minimum failure rate is not less than 2%. The whole failure rate basically fluctuates between 2% and 6%, and the average failure rate is 3.7%. For the aircraft fuel control system, the failure rate of nearly 6% is relatively high which directly affects the operating state of the engine and has a certain impact on the flight safety of the aircraft. Therefore, it is necessary to effectively predict the failure rate of aircraft fuel control system, improve the
operation state of aircraft engine, and ensure the safety and reliability of the aircraft system.

3.2. Prediction Performance Evaluation Indicators. To evaluate the prediction performance of the prediction model, we need to measure and analyze the prediction results, and the emphasis and adaptability of different evaluation indicators are also different. At present, a variety of evaluation indicators are proposed [51]. In order to evaluate the model and test the effectiveness of the method from many aspects, this paper selects the following seven evaluation criteria to measure the error between the predicted value and the real value, specifically sum of squares (SSE), mean absolute error (MAE), root mean square error (RMSE), mean square percentage error (MSPE), normalized root mean square error (NRMSE), consistency index (IA), and mean absolute percentage error (MAPE) are used as reference. Using the above evaluation indicators to comprehensively measure and evaluate the prediction effect, it has a systematic and comprehensive prediction ability compared with general indicators. Suppose the predicted value is \( \hat{y} = \{ \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n \} \), the true value is \( y = \{ y_1, y_2, \ldots, y_n \} \), \( \bar{y} \) is the average value of the true value, and the corresponding prediction model capability formula is obtained as shown in Table 3.

<table>
<thead>
<tr>
<th>Table 2: Statistical parameters of failure rate dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>48-month failure rate data</td>
</tr>
</tbody>
</table>

The SSE calculates the sum of squares of the errors between the fitted data and the corresponding points of the original data. The smaller its value, the better the selection method and fit of the model and the better the data prediction effect. When the predicted value is in good agreement with the actual value, the MAE is the expected value of the absolute error loss, and its smaller value means higher prediction accuracy. RMSE is also one of the comprehensive indicators of error analysis, representing the degree of dispersion of the predicted value, and the best fit is RMSE = 0. MSPE reflects the dispersion degree of error. The smaller the value, the smaller the error fluctuation to a certain extent. The smaller the NRMSE, the higher the accuracy. IA represents the change trend and degree of consistency between the predicted value and the measured value, and the closer it is to 1, the higher the trend and consistency are. The MAPE can be treated as a benchmark, and it is generally considered that when the MAPE is less than 10%, the prediction accuracy is higher.

3.3. Prediction Model Implementation

3.3.1. CEEMD. Using the collected fuel control system failure rate data for the 38 months before the period from March 2014 to April 2017, a prediction model was established as the modeling input, and the CEEMD model was used to decompose the data. After the decomposition of the obtained CEEMD, the timing is shown in Figure 5.

It can be seen from Figure 4 that the failure rate data of the first 38 groups of data are decomposed into 5 components, including 4 IMF fluctuation terms and 1 RF trend term. Among them, the frequency and amplitude of IMF1 and IMF2 components are still at a high level, but the fluctuation is significantly lower than the original data, and the smoothness is significantly improved. The frequency and amplitude of the trend items IMF3 and IMF4 gradually decreased, and the sequence gradually became stable. The trend term RF shows a plateau in the failure rate over the first 38 months. The failure rate data decomposed by CEEMD are more regular, and the change trend of the overall failure rate range after decomposition is consistent with the change trend of the actual aircraft system failure rate. Each component still contains the characteristic information of the original failure rate sequence, which reduces the interference of abnormal data in the modeling process, smooths the curve, facilitates data prediction, and can effectively improve the accuracy of data prediction. The obtained fluctuation terms and trend terms are also more representative for reflecting the real physical components of the original data, and based on this, the next step of predictive modeling analysis is carried out. Use Section 2.3 to calculate the Hurst index of the five components decomposed by CEEMD to obtain the Hurst index of each component, as shown in Table 4.

The self-similarity of time series can be measured by the Hurst index. When \( 0 \leq H < 0.5 \) means negatively correlated, the time series has inverse persistence. When \( H = 0.5 \) means that the time series is purely random, the current state does not affect the future state, and the time series is irrelevant. When \( 0.5 \leq H \leq 1 \) represents a positive correlation, it means that the time series has persistent behavior and long
memory, so the larger the Hurst index, the higher the self-similarity of the time series. It can be seen from Table 4 that the Hurst exponent values of IMF1, IMF2, IMF3, and IMF4 are between 0.5 and 0.9, indicating that the corresponding components are periodic nonstationary. The ARIMA model can effectively analyze the correlation of periodic non-stationary data series and is suitable for time series. It is reasonable to choose the ARIMA model as the prediction model of IMF1 to IMF4 stationary characteristics. However RF has a large Hurst exponent, and its exponent value is greater than 0.9, indicating that the RF component has stable changes and certain non-linear characteristics. Selecting the grey Verhulst model has a good predictive ability for non-linear sequences and is used as a prediction model for RF components. Also, from the perspective of fluctuation, the aircraft failure rate will increase significantly at the beginning and will gradually decline in the future. When the failure rate reaches saturation, the growth rate is close to zero, and there is a negative growth with time. The failure rate shows an “s” downward trend with the development of time and the sample size of failure rate is small, which is more suitable for the grey Verhulst model and has certain advantages over the ARIMA model. The grey model has good prediction effect, so the RF term decomposed by CEEMD is used for prediction. Therefore, the modeling method is more reasonable.

3.3.2. ARIMA Model. According to the basic principle in Section 2.4, after decomposing the collected 38 groups of failure rate data by CEEMD, the obtained IMF1 to IMF4 data sequences are converted into stationary sequences through several differential operations, and the order of each model can be determined. With the help of SPSS software, it can be obtained that IMF1 to IMF4 are transformed into stationary series after one difference operation, at the same time, the appropriate parameter values are calculated by SPSS software to screen P and Q, and the optimal parameter values are selected according to AIC, so as to determine the optimal model. Finally, it is determined that the optimal models of failure rate prediction model are IMF1-ARIMA (3, 1, 1), IMF2-ARIMA (2, 1, 3), IMF3-ARIMA (2,1,4), and IMF4-ARIMA (2, 1, 2), and so far the order and parameter selection of each prediction model is completed, the prediction

<table>
<thead>
<tr>
<th>Components</th>
<th>Hurst index</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMF1</td>
<td>0.5682</td>
</tr>
<tr>
<td>IMF2</td>
<td>0.5694</td>
</tr>
<tr>
<td>IMF3</td>
<td>0.8842</td>
</tr>
<tr>
<td>IMF4</td>
<td>0.8853</td>
</tr>
<tr>
<td>RF</td>
<td>0.9123</td>
</tr>
</tbody>
</table>

Figure 5: Timing diagram after CEEMD.
3.3.3. Grey Verhulst Model. According to the basic principle of Section 2.5, the collected RF data of 38 groups of failure rates after CEEMD are decomposed, and the corresponding prediction model data are obtained by using the grey system modeling software of China Southern Airlines (V7.0) to calculate $a = -0.06$, $b = -0.01$, the corresponding grey Verhulst model is $\frac{dx^{(1)}}{dt} - 0.06x^{(1)} = -0.01(x^{(1)})^2$, and the corresponding time is $\hat{x}^{(1)}_{k+1} = -0.19/ -0.04 = 0.02e^{-0.06k}$. A dependent prediction model is established, the RF component is modeled based on the above model, the subsequent 10 sets of data are predicted, and the predicted value of the RF component is obtained.

3.3.4. Entropy Weight Method. The entropy weight method weight coefficient solution and construction process is shown in Figure 6.

Most of the previous models use simple addition to aggregate the predictions of all models into a final prediction, while the proposed model utilizes the grey Verhulst model for RF components simultaneously after utilizing the ARIMA model for each prediction model of the IMF1-IMF4 components. After modeling and predicting, the corresponding predicted values are obtained, which are based on the predicted values of 5 groups of corresponding failure rates. Assuming that the ARIMA model is used to predict the failure rate of IMF1 as $\hat{y}_{IMF1}$, the ARIMA model is used to predict the failure rate of IMF2 as $\hat{y}_{IMF2}$, the ARIMA model is used to predict the failure rate of IMF3 as $\hat{y}_{IMF3}$, the ARIMA model is used to predict the failure rate of IMF4 as $\hat{y}_{IMF4}$, and the grey Verhulst model predicts the failure rate obtained by RF as $\hat{y}_{RF}$; on the basis of not increasing the complexity, based on the entropy weight method combined modeling theory, the corresponding weight coefficients are obtained, as shown in Table 5.

All IMF variables decomposed by CEEMD participate in the construction of the model, and the weight coefficients of the second, third, and fourth IMF components are only small, indicating that the contribution rate of the IMF component to the final result is small, which may represent the fluctuation of failure rate caused by some random factors. The weight coefficients of the first and fifth IMF components are relatively large, indicating that this component may represent some important factors affecting the failure rate. The final prediction model is obtained by multiplying the corresponding predicted value by the corresponding weight coefficient, and the combined model expression is $\hat{y}_t = \omega_1\hat{y}_{IMF1} + \omega_2\hat{y}_{IMF2} + \omega_3\hat{y}_{IMF3} + \omega_4\hat{y}_{IMF4} + \omega_5\hat{y}_{RF}$. Through the combined model, the predicted values of the last 10 groups of test data are obtained. Compared with the simple addition method, the entropy weight method has good interpretability and better prediction accuracy.

4. Comparison and Analysis of Combined Model Results

Using ARIMA model, CEEMD-ARIAM model, CEEMD-grey Verhulst model, and the proposed CEEMD and combined prediction model to predict 10 sets of failure rate data from May 2017 to February 2018, the change trends of actual failure rate and corresponding failure rate prediction values of the four models are shown in Figure 7.

In Figure 7, it can be seen that the fitting trend of the 10 sets of failure rate data from 2017/05 to 2018/02 after using a single ARIMA model is not high compared with the actual failure, and the prediction effect of the failure rate by the ARIMA model is not good. Ideally, in the whole failure rate prediction period, the deviation of the front section is large, and the deviation of the rear section is small, but the overall fitting degree is not high. The fitting degree of the CEEMD-ARIMA model tends to be reasonable, which is
significantly different from that of the ARIMA model. In the front and middle stages, the fitting effect is significantly better than that of the single ARIMA model, and the fitting degree has been significantly improved, but the overall effect is not ideal. The CEEMD-Verhulst model is used for prediction. The predicted value of the failure rate and the actual value have the largest fitting deviation among the four models. The fitting degree is not high in the entire failure rate prediction period, and the overall fitting effect compared with other three models is not ideal. That CEEMD and combined model are used to predict the failure rate, and compared with other models, it has the highest fitting degree, the deviation between the actual value and the predicted value is the smallest, and the prediction effect is good, which can meet the prediction needs of aircraft failure rate.

In order to select the best prediction model, we compared the error indexes of the validation data in different models, selected 10 groups of validation samples for error calculation, and obtained the corresponding SSE, MAE, RMSE, MSPE, NRMSE E, IA, and MAPE which are shown in Figures 8–10.

As can be seen from Figure 8, the CEEMD combined model performs best among all models, mainly reflecting the following: the SSE value of the model is 0.84, MAE is 0.22, and RMSE is 0.28, which is the smallest compared with ARIMA model, CEEMD-ARIMA model, and CEEMD grey Verhulst model. The SSE, MAE, and RMSE values of the CEEMD-grey Verhulst model are the largest, indicating that the prediction accuracy of the CEEMD-grey Verhulst model is not high. Compared with the CEEMD-grey Verhulst model, the SSE, MAE, and RMSE indexes of this model are reduced by 87.2%, 67.1%, and 65.4%, and the model has higher prediction accuracy than other models.

It can be seen from Figure 9 that the MSPE of this model is 0.17 and the NRMSE is 0.12, which is lower than that of other models, and the MSPE index of this model is reduced by 68.5% compared with the ARIMA model. Compared with the index, the decrease range is larger, and the difference between the two is 0.2. At the same time, the IA of this model is 0.82. The consistency index (IA) of this model is larger than that of other models, and compared with other models, it is closer to 1. Compared with the ARIMA model, the IA index has increased by 26.3%. From the above indexes, the prediction performance of the model is better.

It can be seen from Figure 10 that the MAPE values of the ARIMA model, the CEEMD-ARIAM model, and the CEEMD-grey Verhulst model are all higher than 10%, which are two to three times the value of the CEEMD-combination model. The prediction of ARIMA model and CEEMD grey Verhulst model is reasonable, while the prediction of CEEMD-ARIAM model is good, but the overall prediction performance is not high, while the MAPE value of the CEEMD combined model is 8.36%. Compared with other models, it is the smallest and less than 10%, which also shows that the prediction accuracy of the combined model is high. The results of the above seven evaluation indexes show the prediction effect of the combined model. Therefore, the CEEMD combined model proposed in this paper is suitable for aircraft failure rate prediction and improves the performance and accuracy of aircraft failure rate prediction. In order to better evaluate whether the prediction accuracy of the proposed prediction model is significantly better than

<table>
<thead>
<tr>
<th>Weight factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.211</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.198</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.038</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>0.049</td>
</tr>
<tr>
<td>$\omega_5$</td>
<td>0.504</td>
</tr>
</tbody>
</table>
other models, this study introduces a directional statistic $D_{stat}$ to evaluate the directional prediction ability of the model.

$$D_{stat} = \frac{1}{n} \sum_{i=1}^{n} \alpha_i \times 100\%,$$

where $\alpha_i = \begin{cases} 1, & (y_i - y_{i-1}) \times (\hat{y}_i - \hat{y}_{i-1}) \geq 0; \\ 0, & \text{otherwise}, \end{cases}$ $y_i$ is the actual value, and $\hat{y}_i$ is the predicted value. The larger the value of $D_{stat}$, the higher the prediction accuracy and the better the direction prediction performance. Also, we use the Diebold–Mariano (DM) test to test the pros and cons of the model from a statistical point of view [52]. At the same time, in order to further verify the accuracy of the CEEMD-combined model, the EMD-ARIMA model [53], the BP model [54], and the triple exponential smoothing model [55] were used for comparison and analysis. The different models and the accuracy indicators are shown in Table 6.

Table 6 provides a comprehensive comparison in terms of prediction accuracy metrics between the proposed model and the EMD-ARIMA model, the BP model, and the triple exponential smoothing model. Judging from the accuracy of different evaluation indicators, the IA value of the proposed model is higher than that of other models, while the other six indicators are lower than those of other models. It shows that compared with other models, the prediction model can predict the aircraft failure rate more accurately and has better prediction performance. It can be seen from the directional statistics that the combined model has the highest value of $D_{stat}$ among the four models, indicating that the model has better performance in the direction of failure rate prediction. The EMD-ARIMA model and the BP model have higher values of $D_{stat}$ than the cubic exponential smoothing model. In addition, the combined model is also much better than other prediction models, and the evaluation result of $D_{stat}$ has increased from 72.32%–84.78% to 99.52%. Using the evaluation index of $D_{stat}$ as the evaluation standard related to maintenance decision-making has more practical guidance and constructive significance for the direction judgment of corresponding decision makers in aircraft fault prediction and maintenance. From the perspective of statistics, the prediction accuracy of this model is significantly better than that of other models. The DM test is used to evaluate the proposed model, and the statistical data and $P$ values (in parentheses) of the target model and benchmark model are obtained as shown in Table 7.

It can be seen from Table 7 that at the 5% significance level, the proposed CEEMD combined model is significantly better than the EMD-ARIMA model, the BP model, and the triple exponential smoothing model. The DM test results further confirm the superiority of the model from a statistical point of view.

Because the aircraft failure rate has the characteristics of uncertainty and instability, it is necessary to evaluate the uncertainty and stability of the established model. The uncertainty of the prediction model mainly includes input uncertainty, parameter uncertainty, and model uncertainty, which will have a certain impact on the model. In order to verify the uncertainty of the proposed model, the coefficient of variation is used for verification. Coefficient of variation is a statistical index to measure the dispersion degree and variation degree of each observation value. If the ratio of standard deviation to average is taken as the coefficient of variation, $CoV = \sigma/\mu$, $\sigma$ is the standard deviation, and $\mu$ is the sample mean value [56]. The coefficient of variation can eliminate the influence of different units or averages on the comparison of two or more degrees of variation. It is a dimensionless absolute value, indicating the dispersion of random variables around its mean. The coefficient of variation of different models is obtained by calculation as shown in Figure 11.

As can be seen from Figure 11, the coefficient of variation of the proposed CEEMD combined model is 0.23, which is smaller than that of the BP model, EMD-ARIMA model, cubic exponential smoothing model, CEEMD Verhulst...
model, CEEMD-ARIMA model, and ARIMA model. In terms of uncertainty quantification, the CEEMD-combination method provides more satisfactory results (CoV) than the other six model methods, indicating that the model has less uncertainty and can meet the actual needs. The stability of each model is compared by the Monte Carlo method. Monte Carlo sampling is a general method to compare the performance of prediction methods. Its most remarkable feature is that it ensures the accurate comparison of prediction methods and avoids the biased results that may be obtained accidentally. In this paper, the proposed model is tested by this method and the average error values of different models are obtained as shown in Figure 12. It can be seen from Figure 12 that the average error values of the CEEMD-combined model, the EMD-ARIMA model, the BP model, and the triple exponential smoothing model are, respectively, 1.37, 3.41, 2.39, and 3.52. Compared with the other three methods, the CEEMD-combined model shows better characteristics of significant error difference and avoids accidentally obtaining biased results, indicating that the proposed prediction method has better stability.

The program execution time is one of the indexes in modeling and calculation processing, which determines the training and prediction time of the model. The execution time of the proposed CEEMD combined model is 0.21 s which compares with 0.33 s for EMD-ARIMA, 0.82 s for BP, and 0.41 s for the triple exponential smoothing model, and the time taken is shorter. It shows that the proposed model has a low computational time complexity and can achieve a faster learning process and a satisfactory and acceptable prediction effect, and its prediction model is more efficient.

Table 6: Accuracy evaluation indicators of different models.

<table>
<thead>
<tr>
<th>Accuracy index</th>
<th>EMD-ARIMA model</th>
<th>BP model</th>
<th>Triple exponential smoothing model</th>
<th>CEEMD-combined model</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>1.71</td>
<td>0.98</td>
<td>1.63</td>
<td>0.84</td>
</tr>
<tr>
<td>MAE</td>
<td>0.42</td>
<td>0.34</td>
<td>0.54</td>
<td>0.22</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.63</td>
<td>0.51</td>
<td>1.22</td>
<td>0.28</td>
</tr>
<tr>
<td>MSPE</td>
<td>0.41</td>
<td>0.38</td>
<td>2.43</td>
<td>0.17</td>
</tr>
<tr>
<td>NRMSE</td>
<td>0.22</td>
<td>0.21</td>
<td>0.21</td>
<td>0.12</td>
</tr>
<tr>
<td>MAPE (%)</td>
<td>0.56</td>
<td>0.79</td>
<td>0.45</td>
<td>0.82</td>
</tr>
<tr>
<td>D_{slae} (%)</td>
<td>84.78</td>
<td>80.45</td>
<td>72.32</td>
<td>99.52</td>
</tr>
</tbody>
</table>

Table 7: DM test indicators of different models.

<table>
<thead>
<tr>
<th>Target model</th>
<th>EMD-ARIMA model</th>
<th>BP model</th>
<th>Triple exponential smoothing model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEEMD-combined model</td>
<td>−2.209 (0.042)</td>
<td>−2.184 (0.029)</td>
<td>−2.946 (0.003)</td>
</tr>
</tbody>
</table>
According to the above overall results, it can be seen that the CEEMD-combination model is superior to other models in several evaluation indicators, and the proposed model has better prediction performance and meets the needs of aircraft failure rate prediction.

5. Discussion

In this study, based on the collected aircraft failure rate information and related data, the information decomposition of the failure rate is carried out by using the complementary ensemble empirical mode decomposition (CEEMD) method. Complementary integrated empirical mode decomposition (CEEMD) is a relatively novel data preprocessing method, which can effectively overcome the advantages of mode aliasing and white noise interference, effectively reduce the non-smoothness of time series, greatly improve the smoothness of the original sequence, and make the sequence have a certain regularity. Use the Hurst index to determine the appropriate correlation prediction model for each decomposition component and select the integrated moving average autoregression (ARIMA) model and the grey Verhulst model as the single model of the combined forecasting model to carry out the combined prediction of group failure rate, so as to avoid the randomness and blindness of the selection of prediction model, fully consider the applicable objects of different models, and improve the prediction accuracy. Fully analyze the key role of different weights on the combination model and use the entropy weight method to solve the combination weight, so as to avoid the impact of unified weight on the prediction effect. Therefore, it has strong applicability.

Due to the characteristics of complex cross-linking and random faults of the aircraft system, there are few data sources of fault rate information samples and lack of effective fault characteristics and diversified faults, so it is difficult to troubleshoot the faults of the system. It is difficult to accurately predict the failure rate. Due to the particularity of the prediction object and the complexity of the conditions, the prediction model is not fixed. The single-term ARIMA model not only considers the dependence of the failure rate data on the time series but also considers the instability of random fluctuations of the data, which is suitable for the prediction of aircraft failure rate. The grey model can take advantage of the advantages of less historical sample data and high prediction accuracy required for the grey system prediction. At the same time, it does not need to have typical distribution characteristics and law requirements for the original data. It is suitable for the prediction of aircraft failure rate with less original information and lack of data, but the prediction accuracy is not high in the case of more data.

6. Conclusion

In order to provide good repair and maintenance decision making, establish an aircraft health management mechanism, and efficiently predict the aircraft failure rate, an aircraft failure rate prediction method based on the fusion of CEEMD and combined model was established. The prediction model is constructed with the failure rate of aircraft system failure as the original time series, through the study of performance index, uncertainty, and instability. The case study results of actual aircraft failure rate data show that the proposed prediction method can better predict the change rules of aircraft failure rate and effectively improve the prediction accuracy of aircraft failure rate. The excellent performance of the proposed prediction model is fully demonstrated and the prediction effect is verified. The prediction method provides a certain idea and useful reference for solving the prediction problem of fault rate and effectively improves the ability of aircraft system fault diagnosis, prediction, and comprehensive support.

The accuracy index of the prediction model has a SSE of 0.84, a MAE of 0.22, a RMSE of 0.28, a MSPE of 0.17, and an NRMSE of 0.12. IA is 0.82 close to 1, and the mean absolute percentage error (MAPE) is 8.36%. Controlling below 10% indicates high accuracy, significantly narrowing the error fluctuation range and making the predicted value closer to the actual value. At the same time, the value of $D_{stat}$ is 99.52%, and the DM test method indicates that the CEEMD-combination model used to predict the aircraft failure rate is more reasonable. At the same time, the model has less influence of uncertainty and has the advantage of stability, high efficiency, and remarkable prediction effect.

Although the model achieves better predictions than several traditional methods, there are still some challenges that require further research. (1) This study only uses historical data to predict the future failure rate. In fact, the aircraft system is a complex system and failures are affected by flight time, flight sorts, ambient temperature, ambient humidity, maintenance capabilities, guarantee conditions, and human operations. It has a certain impact on the aircraft failure rate. These factors lead to strong non-linearity, uncertainty, and instability in the failure rate. In the future, a
multi-input prediction model will be established by combining the above influencing factors to further improve the prediction performance and prediction effect. (2) It is not perfect and systematic to evaluate the performance of the model only through several commonly used evaluation indicators, and the reflected indicators and performance are single. In the follow-up, more evaluation indicators should be developed and applied to evaluate the model which can be extended to more indicators to evaluate the model as a whole and improve the overall evaluation ability of the model. (3) The signal decomposition method is adopted, and only the basic CEEMD method and a basic case are used for research. In order to make the research more convincing, in the follow-up research and application, the prediction algorithm research based on multiple combination models (EMD combination model, EEMD combination model, VMD combination model, and CEEMDAN combination model) can be carried out to gradually improve the pre-failure rate measurement accuracy and the prediction ability and universality of the model. At the same time, the prediction method can be applied to the prediction of aircraft failure time, remaining life, and safety monitoring to ensure the safety and reliability of the aircraft and guarantee the flight safety of the aircraft.

Data Availability

The maintenance data used to support the findings of this study have not been made available because sharing the data might compromise data privacy. Moreover, the authors are not allowed to share these data due to security concerns.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Doctoral Startup Foundation of Shenyang Aerospace University (grant no. 19YB30) and Scientific Research Funding Project of the Education Department of Liaoning Province (General Program) (grant no. LJKZ0202).

References


[52] D. Pradeepkumar and V. Ravi, “Forecasting financial time series volatility using particle swarm optimization trained...


