Harmonic Reducer Performance Prediction Algorithm Based on Multivariate State Estimation and LargeVis Dimensionality Reduction

YIYANG ZHAO[1], PEIXING LI[2], HAOLUN DING[1], JIABIN CAO[1], AND WEIXIN YAN[1]
[1]School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Corresponding author: Weixin Yan (e-mail: xiaogu4524@ sjtu.edu.cn).

ABSTRACT As an important transmission component of industrial robots, the harmonic reducer determines the positioning accuracy, bearing capacity and service life of the robot end-effector. Predicting the performance can grasp the working status in advance and avoid major losses caused by uncertain factors such as component damage. The current paper focuses on a harmonic reducer performance prediction algorithm based on Multivariate State Estimation Technique (MSET) and LargeVis dimensionality reduction. Firstly, an accelerated life test platform is designed to collect multi-dimensional parameters that can characterize the operating state of the harmonic reducer throughout the life cycle. Afterwards, as far as the MSET method is concerned, the fault warning threshold is set according to the residual between the constructed memory matrix of the health state data and the actual observed value. Finally, utilizing LargeVis to reduce the dimensionality of multi-dimensional features, combining with Mahalanobis distance to construct a health index degradation model, and then selecting Long Short-Term Memory (LSTM) network to predict the downward trend of the harmonic reducer. The analysis of the accelerated life test data of the harmonic reducer demonstrates that the proposed method can send out the fault warning signal 18 minutes in advance in the sample with a life of 5.7 hours, and has a strong ability to predict the degradation trend of the harmonic reducer.

INDEX TERMS Harmonic reducer, performance prediction, Multivariate State Estimation Technique, LargeVis, features dimensionality reduction, Long Short-Term Memory network.

I. INTRODUCTION

Industrial robots are the most common type of robots, which are widely used in intelligent manufacturing. In the era of Industry 4.0, industrial robots also play an indispensable role in it [1]. Industrial robots can replace people to do many things, such as painting of cars and tightening of screws, which require high stability, low failure rate and low vibration of industrial robots [2]-[5].

Reducers are an important component of industrial robots [6], and their life and stability are directly related to the operating conditions of industrial robots. Precision reducers are utilized in all joints of industrial robots to improve the overall stiffness and output torque of the robot. There are two main types of precision reducers used in robots, RV reducers and harmonic reducers. RV reducers are large in size compared to harmonic reducers and are mainly used in large joints of industrial robots. Harmonic reducer is small in size, large transmission ratio, small mass and high transmission efficiency and transmission accuracy, which is widely used in wrist and small arm hand joints of industrial robots. The performance of harmonic reducer is directly related to the safety and stability of industrial robots.

The model diagram of harmonic reducer is shown in Figure 1. Harmonic reducer mainly consists of flexible wheel, rigid wheel and wave generator. As the use time increases, the harmonic reducer motion accuracy decreases

and fatigue damage occurs. Damage to the harmonic reducer occurs in the robot will bring a large economic loss. At present, the factory uses regular inspection and maintenance to troubleshoot the harmonic reducer. There is a certain lag in this troubleshooting method, and it is important to establish a harmonic reducer health assessment model to predict its performance [7].

FIGURE 1. Harmonic reducer model diagram

Researchers have conducted some theoretical and experimental studies on the performance evaluation of mechanical structures such as reducers. Zhang [8] investigated the reliability evaluation method of harmonic reducers using transient finite elements and accelerated life tests. Qian [9] proposed a multi-failure mode time-varying reliability analysis method based on a two-loop Kriging model to analyze the time-varying reliability of reducers. Liu [10] proposed a simple index indicating the failure severity based on the ratio between the characteristic frequency component of the failure and the currently sampled vibration signal. The fault characteristic frequency components are identified by applying wavelet packet decomposition and Hilbert transform to the original signal. Shiau [11] proposed a nonparametric regression-based accelerated life stress model for describing the accelerated degradation curve. An estimate of the mean time to failure under normal conditions of the product was obtained by analyzing the relationship between the stress level and the acceleration factor. Crk [12] proposed a multivariate multiple regression function parametric analysis method for applied stresses using a degradation trajectory-based approach. Doksum [13] proposed a variable stress accelerated life test model based on the concept of cumulative degradation with a Wiener process and inverse Gaussian distribution, which has more flexibility for nonparametric life models. Li [14] built a performance margin model based on the hysteresis and transmission error of harmonic reducer as key parameters, and analyzed and quantified the multi-source uncertainty to construct a reliability model. Jun [15] proposed a new n-ARIMA based equipment performance degradation model method, and used a non-smooth autoregressive integrated moving

average model to build an equipment performance degradation model to verify the performance degradation of OTM650. Riascos-Ochoa [16] proposed a Levy framework to simulate the effect of degradation sources on the degraded system, and obtained the moments of reliability function, probability density and lifetime. The performance of harmonic reducers is often expressed through multiple state parameters, such as vibration, transmitted torque, temperature, etc. These performance evaluation methods, designed using a single state parameter, do not reflect the performance of harmonic reducers.

Some scholars have also developed performance evaluation models through neural networks. Ma [17] proposed a particle swarm optimization based reliability life prediction analysis for harmonic reducers. Wang [18] proposed a hybrid prediction method for remaining service life prediction of rolling bearings using a correlated vector machine regression with different kernel parameters to represent the bearing test dataset and adaptively estimate the remaining service life using an exponential degradation model combined with Frechet distance. Prudhom et al [19]. showed that a simple fault severity index using short-time fourier transform (STFT) information can be proposed based on the notation over definition of time-frequency transform instead of fault indicators. Wang [20] proposed a bearing performance evaluation method based on topological representation and hidden Markov model, where a topological network of original features is obtained by self-organizing mapping, and a hidden Markov model is used as the evaluation model to assess the bearing performance degradation trend. These performance evaluation models established using neural networks are able to predict the performance state of the mechanism, but a large amount of experimental data is required to establish convolutional neural networks and machine learning models, harmonic reducers have a long service life, and it takes more than ten years to collect hundreds of data sets to establish performance evaluation datasets, and the model training process is complex and prone to overfitting.

In brief, our significant technical contributions are the following:

(1) A test platform for the accelerated life of harmonic reducer is built, and multiple sets of multi-dimensional data parameters such as input speed, output speed, input torque, output torque, temperature and acceleration in X, Y and Z directions are collected during the operation.

(2) The designed MSET method can realize the comprehensive condition monitoring of the harmonic reducer, and the fault warning signal can be sent out 18 minutes in advance by setting the fault threshold.

(3) The dimensionality reduction of the characteristic parameters is realized by the LargeVis technology, and the performance degradation trend information of the harmonic reducer is predicted by combining with the LSTM method.

2

II. Accelerated Life Test of Harmonic Reducer

In this paper, a test platform is set up to conduct accelerated life experiments and collect multidimensional parameters that can characterize the operation state of the harmonic reducer during its full life cycle. The test platform consists mainly of drive frequency conversion motor, harmonic reducer, torque sensor, magnetic powder brake, accelerometer, as shown in Figure 2. Among them, the performance parameters of the drive motor are shown in Table 1, with a rated power of 1.1kW and a rated torque of

3.5 N·m. It is connected with the harmonic reducer through a coupling to drive the harmonic reducer. The performance parameters of the measured harmonic reducer are shown in Table 2. The torque sensor selects a range of 5 N·m and 200 N·m respectively to measure the torque of the input and output terminals. The TR-3 acquisition instrument matching the sensor is selected, and the RS232 interface is adopted to communicate with the computer. The model of the magnetic powder brake is FZ200J/Y, which can provide a rated torque of 200 N·m, all of which meet the test requirements.
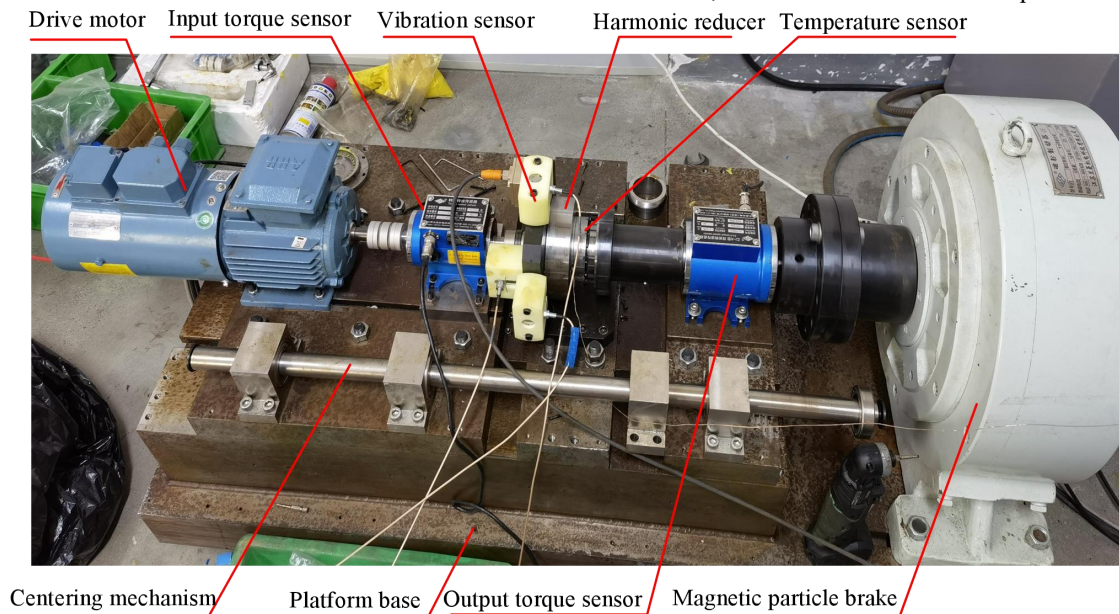

FIGURE 2. Accelerated life test device

TABLE 1. Drive motor performance parameters

Model	Rated power kW	Rated torque N·m	Rated speed r/min	Maximum allowable torque N·m	Moment of inertia kg·m²
QABP 80M2B	1.1	3.5	2845	2.8	0.00107

TABLE 2. Harmonic reducer performance parameters

Performance	Value
Reduction ratio	51
Rated torque (N·m)	32
Maximum load torque (N·m)	69
Maximum average load torque (N·m)	42
Maximum torque (N·m)	121
Maximum input speed (r/min)	6000
Average input speed (r/min)	3500
Backlash (Arcsec)	20
Design life (h)	10000

In order to obtain the full life cycle data of the harmonic reducer, several groups of accelerated life tests are designed and carried out in this paper, and the results are shown in Table 3.

3

IEEE *Access*

Multidisciplinary ┊ Rapid Review ┊ Open Access Journal

TABLE 3. Accelerated life test of harmonic reducer

Serial number	Load	Sampling time (s)	Running time (h)	Failure form
1	3 times	30	21.3	flexible wheel fracture
2	3 times	30	89	flexible wheel fracture
3	3 times	30	9.9	flexible bearing failure
4	3 times	30	69.3	flexible wheel fracture
5	3 times	30	54.4	flexible wheel fracture
6	4 times	1	3.1	flexible bearing failure
7	4 times	1	9.8	flexible wheel fracture
8	4 times	1	5.7	flexible wheel fracture
9	4 times	1	1.1	flexible wheel fracture
10	4 times	1	7.7	flexible wheel fracture
11	4 times	1	18.6	flexible wheel fracture
12	4 times	1	7.0	flexible wheel fracture
13	4 times	1	8.5	flexible wheel fracture
14	4 times	0.05	17.3	flexible wheel fracture

According to 14 groups of accelerated life tests, most of the failure forms of harmonic reducer are flexible wheel fracture, and a few are flexible bearing damage. The operating time ranges from several hours to dozens of hours, with strong uncertainty. If the traditional method based on physical model is used, there will be a large error. The MSET method based on data drive proposed in this paper is used for condition monitoring and fault warning to inform the appropriate time for preventive maintenance of equipment.

In addition, the lifespans of samples No. 6 and No. 9 are 3.1 hours and 1.1 hours, respectively, which is relatively short. Combine with the analysis of the test site and consultation with the manufacturer of the harmonic reducer, the two groups of samples are found to be short in running time due to manufacturing defects. Therefore, there is no reference value. At the beginning of the test process, the harmonic reducer is in an accelerating state, and the data after the harmonic reducer runs smoothly is intercepted for analysis. In addition, sensor errors during the test may cause a large difference between individual data and the overall sample, which should be discarded in the subsequent model calculation.

When the harmonic reducer is running, the temperature in the body rises due to friction heating. Especially in the later period of use, the vibration intensifies and the input and output torque cannot be maintained. Therefore, input speed, output speed, input torque, output torque, temperature and acceleration in X, Y, and Z directions are collected in the test to analyze the performance of the harmonic reducer.

III. Harmonic Reducer Performance Prediction Algorithm

Harmonic reducer performance prediction algorithm mainly includes condition monitoring and degradation trend prediction. The realization principle diagram is shown in Figure 3. The condition monitoring adopts the MSET method, and compares the residual error between the memory matrix constructed by the health state and the actual observation value, thereby realizing the failure warning of the harmonic reducer. Degradation trend prediction is to employ LargeVis feature dimensionality reduction and LSTM network to predict the degradation trend of harmonic reducer.
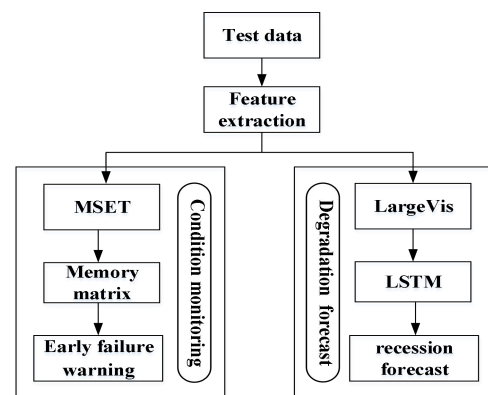

FIGURE 3. Process of condition monitoring and degradation trend prediction

A. HARMONIC REDUCER CONDITION MONITORING METHOD

1) MSET method

MSET [21] is a non-parametric statistical modeling method based on the weighted average of historical data, which is suitable for failure warning of mechanical equipment with higher test cost and greater risk factor. For industrial robot equipment, there is a certain risk in its fault test, and it is difficult to obtain the fault feature database. Hence, the MSET technology is utilized to estimate the true state of the observation vector by learning the mechanical knowledge contained in the historical monitoring data, and then a fault warning model is established.

Firstly, the historical memory matrix D is constructed by utilizing the monitoring data under in normal operation state, and then the MSET model is established. Assuming that n variables are collected by the sensor , the observation vector at a certain sampling time t_m can be expressed as:

$$X(t_m) = \left[x_1(t_m), x_2(t_m), x_3(t_m), \cdots, x_n(t_m)\right]^{\mathrm{T}} \quad (1)$$

Where $x_n(t_m)$ is the observed value of variable x_n at sampling time t_m. Then the historical memory matrix D can be constructed as:

$$D = \left[X(t_1), X(t_2), \cdots X(t_m)\right]$$

$$= \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_m) \\ x_2(t_1) & x_2(t_2) & \cdots & x_2(t_m) \\ \vdots & \vdots & & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_m) \end{bmatrix}_{n \times m} \quad (2)$$

The memory matrix D is the basis of the MSET model, and the construction of the memory matrix is the process of learning and remembering the normal operating characteristics of the harmonic reducer. Its columns represent the values of all observed variables at a certain sampling moment, and its rows represent the values of an observed variable at different sampling moments.

For a new observation vector X_{obs}, it can be compared with the historical observation vector stored in the memory matrix D, and then the prediction vector X_{est} can be estimated. X_{est} can be expressed as the product of D and the weight vector w :

$$X_{est} = D \bullet w = w_1 X(t_1) + w_2 X(t_2) + \cdots + w_m X(t_m) \quad (3)$$

Where X_{est} is the linear combination of the m historical observation vectors in D. The weight w represents the similarity measure between the vectors in X_{est} and D, and the weight w can be calculated by minimizing the residuals. The residual formula is as follows:

$$\varepsilon = X_{obs} \text{-} X_{est} \quad (4)$$

The optimal weight vector w is selected to minimize the residual sum of squares between the newly measured vector X_{obs} and the predicted value X_{est}. Then the residual sum of squares can be given by:

$$S(w) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^{\mathrm{T}} \varepsilon = (X_{obs} - X_{est})^{\mathrm{T}} \bullet (X_{obs} - X_{est})$$

$$= (X_{obs} - Dw)^{\mathrm{T}} \bullet (X_{obs} - Dw)$$

$$= \sum_{i=1}^{n} \left(X_{obs}(i) - \sum_{j=i}^{m} w_j D_{ij}\right)^2 \quad (5)$$

Take the partial derivatives of $S(w)$ with respect to w_1, $w_2, \cdots, w_m$ and make them equal to 0. Then the weight w can be written as follows:

$$w = (D^{\mathrm{T}} \otimes D)^{-1} \bullet (D^{\mathrm{T}} \otimes X_{obs}) \quad (6)$$

Where the nonlinear operator $\otimes$ chooses to use Euclidean distance operation.

$$\otimes(X, Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (7)$$

When the harmonic reducer works under normal conditions, the model input value X_{obs} is located in the normal working space of the memory matrix D, and is close to the historical observation vector in D. Once a failure occurs, X_{obs} deviates from the observation vector in the memory matrix, the accuracy of the output value decreases, and the residual will increase accordingly. Hence, the residual between X_{est} and X_{obs} can reflect the abnormal state and fault information. By means of monitoring changes in residuals and setting reasonable thresholds, fault alarm can be achieved.

2) Feature selection

In view of the construction of the memory matrix D, different strategies are employed to construct the rows and columns.

The purpose of feature selection is to transform the raw data into a more meaningful form, thereby simplifying the model and improving performance. The feature selection of the Pearson correlation coefficient method is employed to screen the rows of the memory matrix. Because it has a simple process, high computational efficiency, and it is suitable for large-scale data. The degree of linear correlation between two variables is calculated by the Pearson correlation coefficient, and the value range is generally [-1, 1]. The larger the absolute value, the higher the degree of linear correlation. Then the overall Pearson correlation coefficient formula between the two variables is as follows:

$$\rho_{x,y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{E[(x - \mu_x, y - \mu_y)]}{\sigma_x \sigma_y} \quad (8)$$

The formula for calculating the correlation coefficient between two samples can be expressed as:

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (9)$$

Due to the long sampling time and the huge amount of data, the memory matrix will be redundant, thereby reducing the computational efficiency. The commonly used equal interval sampling method is a typical one-time offline

5

sample selection method, which can effectively control the size of the memory matrix. However, considering the changes of actual machining parameters and operating conditions, the estimation accuracy of the MSET with a fixed memory matrix will gradually decrease and frequent maintenance is required. Accordingly, it is necessary to improve the flexibility of the fixed memory matrix. In this paper, a dynamic memory matrix construction method based on KNN algorithm is adopted to construct the columns of the matrix, realize online calculation of MSET, and observe the distance between the vector and each sample in the training set in real time. Taking variable x_i as an example, the method flow of adding observation vectors to the memory matrix is shown in Figure 4. Here, δ is a small positive number.
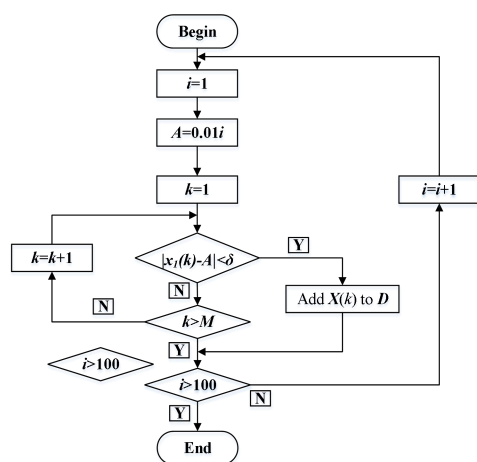

FIGURE 4. Flowchart of memory matrix construction

3) Sliding window method

Within the context of the current paper, the sliding window method is adopted to eliminate the uncertain factors and random interference of the harmonic reducer in operation, thereby improving the reliability of fault warning. By finding an appropriate sliding window width, the statistical characteristics of continuously changing residuals can be quickly captured.

Assuming that in a certain period of time, the sequence number of residuals estimated by the MSET method can be given by:

$$\varepsilon(X_{obs}, X_{est}) = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N] \qquad (10)$$

Setting the width of the sliding window to N, then the moving average of N consecutive residual values within the window can be calculated as:

$$\bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \qquad (11)$$

The threshold of fault warning ε_t is determined by the maximum average residual value in the sliding window. Under the condition of normal operation of the equipment, the maximum residual average value between the normal observation vector and the estimated vector of the MSET model is ε_N. The fault warning threshold can be designed as:

$$\varepsilon_T = k \varepsilon_N \qquad (12)$$

Where k is the alarm threshold coefficient, which is determined by the operator according to actual experience, and the value is generally greater than 1. When the residual value is greater than the residual threshold ε_T, the system sends out an alarm signal.
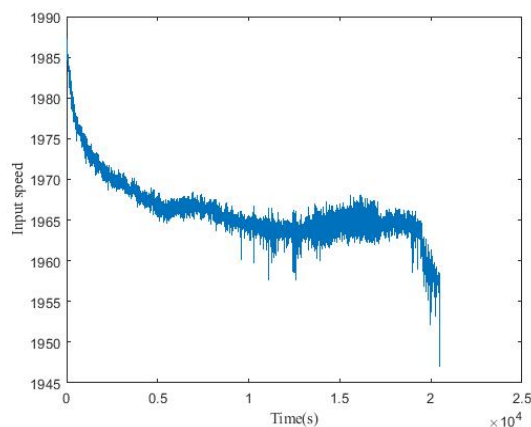
4) Experimental verification

Taking the sample with a lifespan of 5.7 hours in the 8th test as an example, the total number of samples is 20000 after deleting the data of abnormal and initial acceleration stage. The magnitude of the correlation between variables is calculated, and the Pearson feature correlation diagram is shown in Figure 5. After calculation, it can be known that the output speed and torque have little correlation with other variables, as shown in Table 4.

FIGURE 5. Pearson feature correlation diagram

TABLE 4. Correlation of each state variable

State variables	Correlation	State variables	Correlation
input speed	0.73	temperature	0.67
input torque	0.71	X- acceleration	0.76
output speed	0.56	Y- acceleration	0.77
output torque	0.60	Z- acceleration	0.75

(a) Input speed

6

(b) Input torque

(e) Y-acceleration
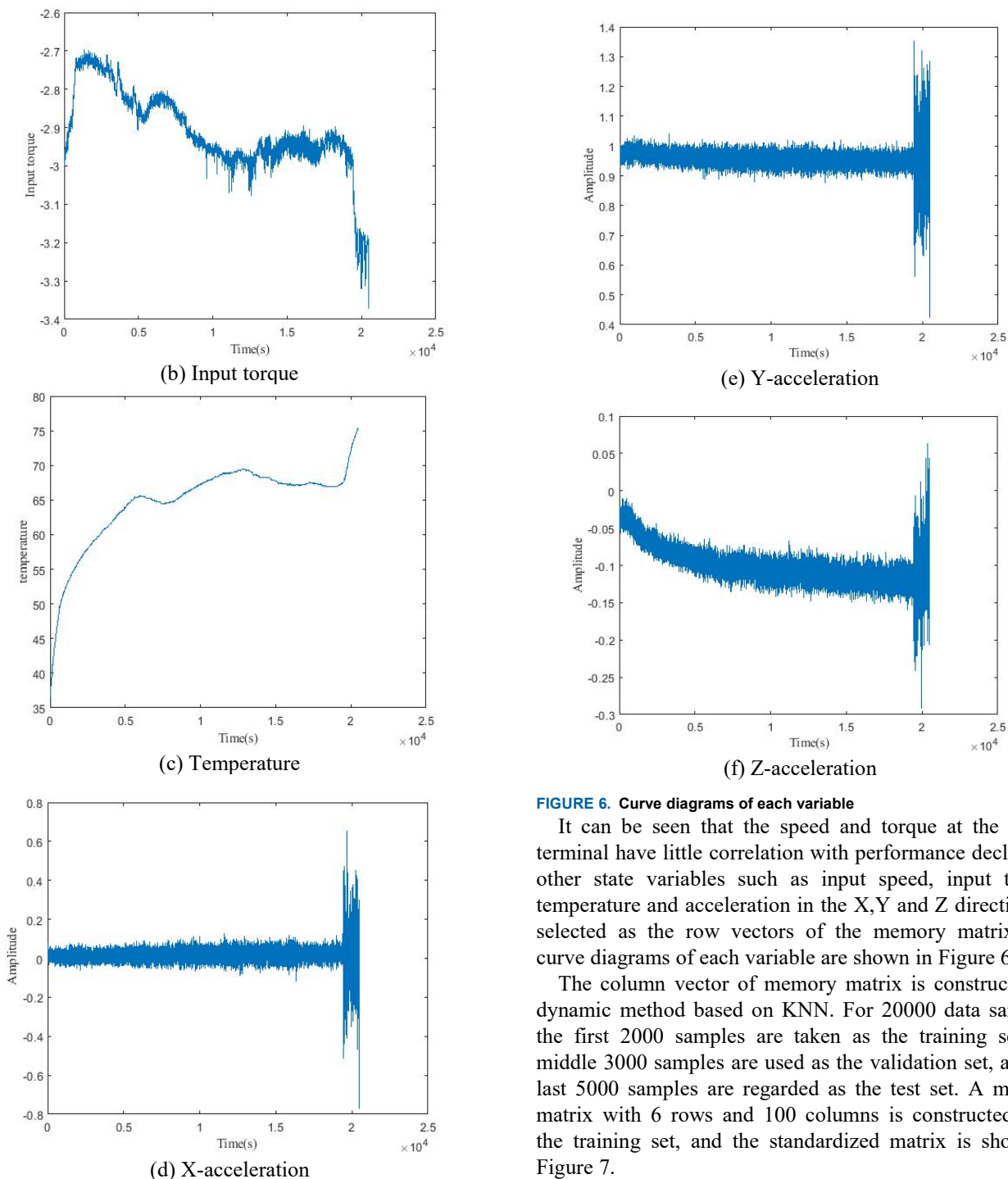
(c) Temperature

(f) Z-acceleration

FIGURE 6. Curve diagrams of each variable

It can be seen that the speed and torque at the output terminal have little correlation with performance decline, so other state variables such as input speed, input torque, temperature and acceleration in the X,Y and Z direction are selected as the row vectors of the memory matrix. The curve diagrams of each variable are shown in Figure 6.

The column vector of memory matrix is constructed by dynamic method based on KNN. For 20000 data samples, the first 2000 samples are taken as the training set, the middle 3000 samples are used as the validation set, and the last 5000 samples are regarded as the test set. A memory matrix with 6 rows and 100 columns is constructed from the training set, and the standardized matrix is shown in Figure 7.

(d) X-acceleration

$$\begin{bmatrix} 0.851585 & 0.855908 & 0.860231 & \cdots & 0.988473 & 0.987032 & 0.984150 & 0.984150 \\ 0.998986 & 0.998783 & 0.998833 & \cdots & 0.996805 & 0.996957 & 0.996145 & 0.995841 \\ 1.003311 & 1.016556 & 1.012509 & \cdots & 1.084253 & 1.079838 & 1.093451 & 1.091979 \\ 0.134831 & 0.303371 & 0.303371 & \cdots & 0.202247 & 0.662921 & 0.202247 & 0.123596 \\ 0.983135 & 0.958333 & 0.982143 & \cdots & 0.972222 & 0.960317 & 0.957341 & 0.953373 \\ 1.421875 & 1.500000 & 1.125000 & \cdots & 1.812500 & 1.734375 & 1.671875 & 1.593750 \end{bmatrix}_{6\times100}$$

FIGURE 7. Memory matrix

7

The failure threshold is determined through the validation set and the application of fault warning is carried out on the test set. When the residual exceeds the threshold, it directly indicates that the current state is abnormal, and then a fault alarm is issued. The residual distribution on the validation set is shown in Figure 8.
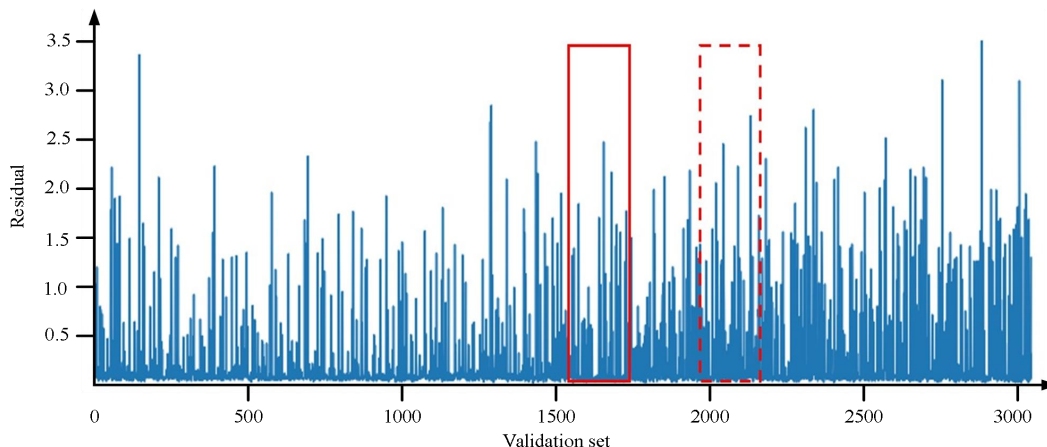

FIGURE 8. Sliding window analysis of residual distribution based on validation set

Using the sliding window method to calculate the maximum residual average value ε_N, the sliding window width is set to 200. Through calculation, the maximum residual average value is 0.4. Setting the alarm threshold coefficient k to 2.5, then the calculation method of the fault warning threshold is as follows:

$$\varepsilon_T = k\varepsilon_N = 2.5 \times 0.4 = 1 \qquad (13)$$

The 5000 sets of operating data before the failure occurred as the test set, and the residuals are compared with the estimated vector calculated by MSET. The residual results are shown in Figure 9. According to the fault warning threshold, calculate the average residual value in a sliding window with a width of 200. The first alarm is issued at the 3890th sampling point, that is, about 18 minutes before the fault occurs, and the residual value at the subsequent sampling points increases sharply. The MSET algorithm can be used to evaluate the residuals between prediction vector and observation vector, and an alarm can be issued in time, thereby winning precious time for troubleshooting. In the meanwhile, the proposed MSET method can uniformly describe the states of multiple variables without modeling each variable separately, which can quickly realize the fault warning and has higher application value.
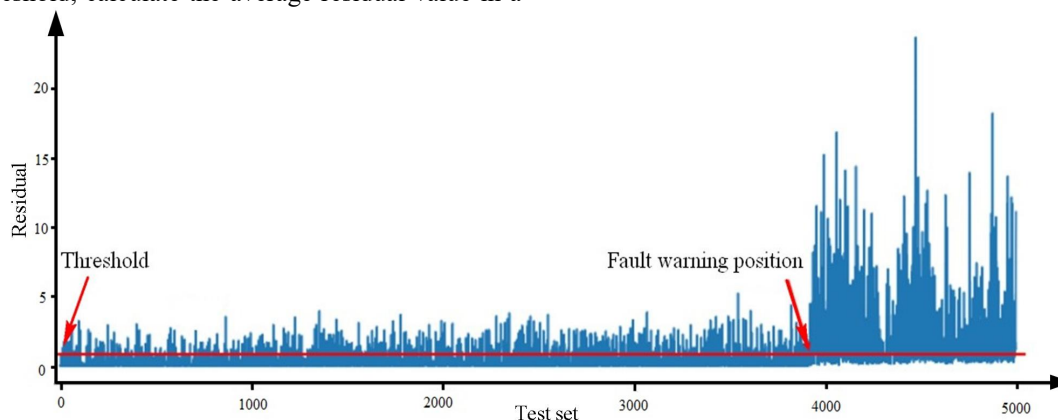

FIGURE 9. Fault warning based on residual distribution of test set

B. HARMONIC REDUCER DEGRADATION TREND PREDICTION

1) LargeVis dimensionality reduction method

The LargeVis technique is a data dimensionality reduction method for high-dimensional space. The state data of the harmonic reducer during the operation process contains abundant state information. The high-dimensional vector is mapped to the low-dimensional space through the dimensionality reduction method, and the overall distribution law of the observation data is visualized [22]. The basic process is shown in Figure 10.

(a)　　　　　(b)　　　　　(c)

8
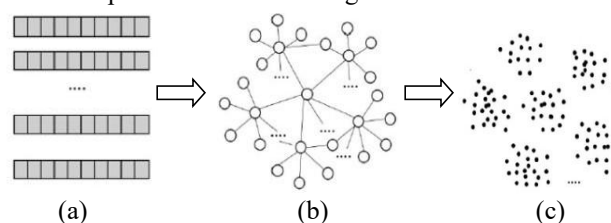
(a) High-dimensional feature vectors
(b) K-nearest neighbor graph
(c) Visualization of dimensionality reduction results

FIGURE 10. LargeVis dimensionality reduction process.

The conditional probability distribution of x_i to x_j in the higher dimensional space can be written as:

$$p_{j|i} = \frac{e^{-\|x_i - x_j\|^2 / 2\sigma_i^2}}{\sum_{k \neq i} e^{-\|x_i - x_j\|^2 / 2\sigma_i^2}} \qquad (14)$$

Where $p_{j|i}$ denotes the probability that the sample point x_j is a neighbor of x_i, and σ_i represents the variance of the Gaussian distribution with x_i as the center point.

The joint probability in high-dimensional space can be given by:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n} \qquad (15)$$

The conditional probability distribution of the low-dimensional space can be expressed as:

$$p(e_{ij} = 1) = f(\|y_i - y_j\|) \qquad (16)$$

Where y_i and y_j represent two points in low-dimensional space, which have a binary edge with weight $e_{ij} = 1$ in KNN graph, where $f(x)$ is:

$$f(x) = \frac{1}{1 + e^{x^2}} \qquad (17)$$

The objective function of the LargeVis algorithm can be written as:

$$O = \sum_{(i,j) \in E} p_{ij} \log p(e_{ij} = 1) + \sum_{(i,j) \in \bar{E}} \gamma \log(1 - p(e_{ij} = 1)) \qquad (18)$$

Where E is the set of positive samples, $\bar{E}$ is the weight uniformly set for the negative sample edge.

2) Distance evaluation method

The Mahalanobis distance is adopted to measure the covariance distance of the data, which is an effective method to calculate the similarity between two location sample sets. The dimension reduction feature of the harmonic reducer in the healthy state is selected as the baseline data, and the Mahalanobis distance between the sample data and the baseline data under different working conditions is compared. The health index is utilized to measure the health degree of the component. The smaller the distance between the sample point to be tested and the baseline, the better the health degree. In this case, the health index is defined as 1. Conversely, it represents the decline of health to a very poor state, and the health index is defined as 0.

$$H_l = e^{-k_0 \times d(x,y)} \qquad (19)$$

Where the parameter k_0 is adopted to adjust the density of the health value distribution corresponding to each state, so that the health value is evenly distributed in the interval $(0,1)$. $d(x, y)$ represents the Mahalanobis distance, which can be described as:

$$d(x, y) = \sqrt{(x - y)^{\mathrm{T}} \sum_{xy} (x - y)} \qquad (20)$$

3) LSTM prediction method

LSTM [23] is a special type of RNN network that can solve the problems of vanishing and exploding gradients of traditional RNN networks. LSTM adds a memory unit to the neural nodes of the hidden layer of RNN to record the historical information. And three gates of input, forget and output are added to control the use of historical information. For LSTM, its parameter settings are shown in Table 5, and the internal structure of neurons is shown in Figure 11.

TABLE 5. Parameter settings of LSTM

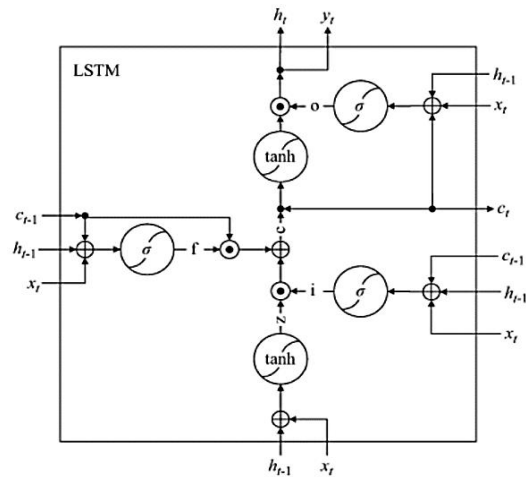| Parameters	Value
numFeatures	1
numResponses	1
numHiddenUnits	40
Optimizer	adam
MaxEpochs	1000
GradientThreshold	1
InitialLearnRate	0.01
LearnRateDropFactor	0.2

FIGURE 11. Internal structure of LSTM neuron

Observing the picture in Figure 11, where i, f, c, and o represent the input gate, forget gate, unit state, and output gate, respectively. σ and tanh represent the sigmoid and hyperbolic tangent activation functions, respectively. Then, at moment t, the calculation method of the input and output values of the LSTM network can be expressed by the following formula:

$$i_t = sigmoid(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i)$$
$$f_t = sigmoid(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f)$$
$$c_t = f_t c_{t-1} i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \qquad (21)$$
$$o_t = sigmoid(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_{t-1} + b_o)$$
$$h_t = o_t \tanh(c_t)$$

9

The Root Mean Square Error (RMSE) is the square root of the ratio of the deviation between the predicted value and the true value to the number of observations. Here, RMSE is used as the performance measurement index of the predicted results, and the formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{pi}-y_{ti})^2} \qquad (22)$$

Where y_{pi} is the predicted value and y_{ti} is the actual value.

4) Experimental verification

In order to verify the effectiveness of LSTM method, the 8th and 14th groups of harmonic reducer samples with lifespans of 5.7 hours and 17.3 hours are used as examples for analysis. According to the correlation degree calculation of state variables in the previous section, acceleration signals in the X, Y and Z directions with the strongest correlation are selected for research. Similarly, the total sample size is 20,000 and 1.2 million after deleting the data of abnormal and initial acceleration phases, respectively. The mean, peak-to-peak value (ppv), standard deviation (std), and kurtosis value (kv) of time domain features are extracted to form the feature vector X, then:

$$X = \left[X^1, X^2, X^3 \right]$$
$$= \left[X_{\text{mean}}^1, X_{\text{ppv}}^1, X_{\text{std}}^1, X_{\text{kuv}}^1, \cdots, X_{\text{mean}}^3, X_{\text{ppv}}^3, X_{\text{std}}^3, X_{\text{kuv}}^3 \right] \qquad (23)$$

For the two sets of experiments, features are extracted every minute, and high-dimensional vectors V_1 and V_2 are constructed respectively, as follows:

$$V_1 = \begin{bmatrix} X_{\text{mean}}^{[1]^1} & X_{\text{ppv}}^{[1]^1} & \cdots & X_{\text{kuv}}^{[1]^3} \\ X_{\text{mean}}^{[2]^1} & X_{\text{ppv}}^{[2]^1} & \cdots & X_{\text{kuv}}^{[2]^3} \\ \vdots & \vdots & \vdots & \vdots \\ X_{\text{mean}}^{[333]^1} & X_{\text{ppv}}^{[333]^1} & \cdots & X_{\text{kuv}}^{[333]^3} \end{bmatrix}_{333\times12} \qquad (24)$$

$$V_2 = \begin{bmatrix} X_{\text{mean}}^{[1]^1} & X_{\text{ppv}}^{[1]^1} & \cdots & X_{\text{kuv}}^{[1]^3} \\ X_{\text{mean}}^{[2]^1} & X_{\text{ppv}}^{[2]^1} & \cdots & X_{\text{kuv}}^{[2]^3} \\ \vdots & \vdots & \vdots & \vdots \\ X_{\text{mean}}^{[1000]^1} & X_{\text{ppv}}^{[1000]^1} & \cdots & X_{\text{kuv}}^{[1000]^3} \end{bmatrix}_{1000\times12} \qquad (25)$$

Utilizing LargeVis to reduce the dimension of the extracted features, and effectively calculate the similarity between different data sets by means of Mahalanobis distance. Since the data of the harmonic reducer is relatively stable in the first 10 minutes of operation, it is regarded as healthy data. The Mahalanobis distance is based on the data of the first 10 minutes, and the dimensionality reduction features in the health state are taken as the baseline data to calculate the distance between the sample data to be tested and the baseline data. Afterwards, the Mahalanobis distance is nonlinearly mapped, and the health index of each sample point is calculated according to the health index expression. Then

the change graph of the health values is shown in Figure 12. For the sake of a simple notation, (a) represents the 8th group of experiment, (b) displays the 14th group of experiment, the following experiments are consistent with the characterization method of this experiment, and will not be repeated one by one. Here, the value of k_0 is 1. It can be clearly seen from the figure that the overall trend of the health value is on a downward trend, and the results of the health assessment are consistent with the actual situation. It is directly proved that the combination of the LargeVis dimensionality reduction and the Mahalanobis distance method can better evaluate the performance degradation state of the harmonic reducer.
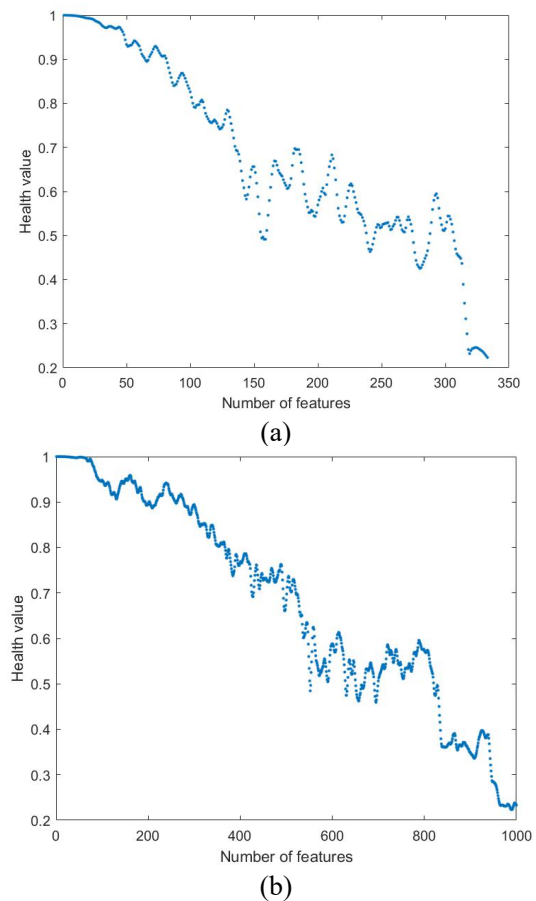

(a)

(b)

FIGURE 12. Health performance curve

LSTM is employed to predict the performance state of the harmonic reducer. Figure 13 demonstrates the comparison between the actual and the predicted degradation curves of the two groups of experiments. Where blue indicates the actual degradation curve, and red represents the predicted degradation curve. It is found that with regard to the performance degradation, the predicted results are similar to the actual value, and infinitely close to the actual value.
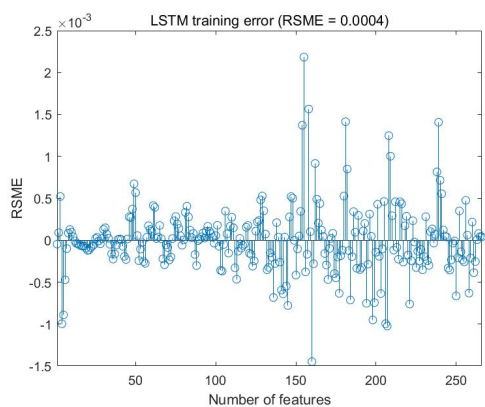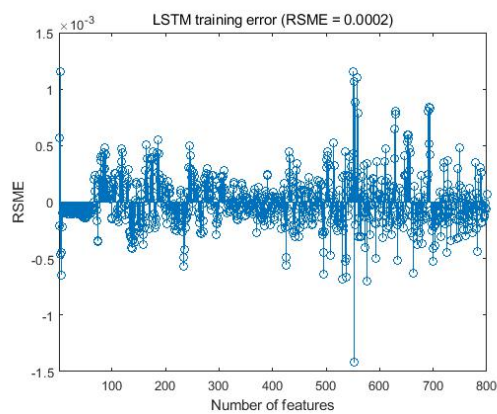
10
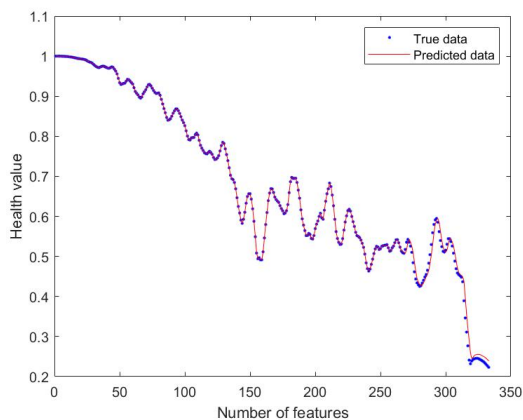
(a)
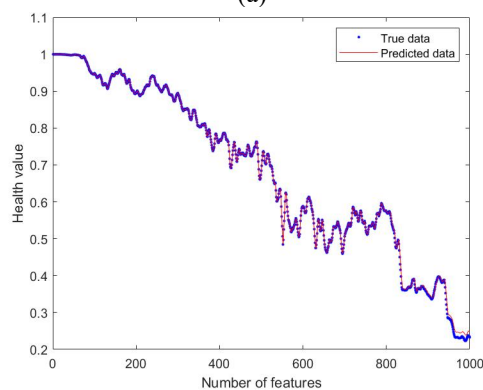
(b)

FIGURE 13. **LSTM-based performance degradation prediction graph**

(b)

FIGURE 14. **LSTM training error**

Here, we select the first 80% of the data in the full life cycle samples as the training set, and the last 20% as the test set. The training error and test error are shown in Figures 14 and 15, respectively. As can be obviously understood from the figures, the training errors of the two experiments are 0.0004 and 0.0002, respectively, and the test errors are 0.0188 and 0.0096, respectively. The test error is slightly higher than the training error. But in general, there is no doubt that the LSTM method for predictive analysis has less error and better effect. It directly indicates that the LSTM method has a greater prediction effect on the performance decline trend of the harmonic reducer.
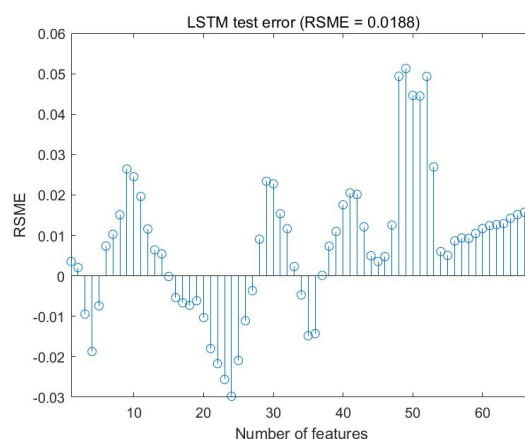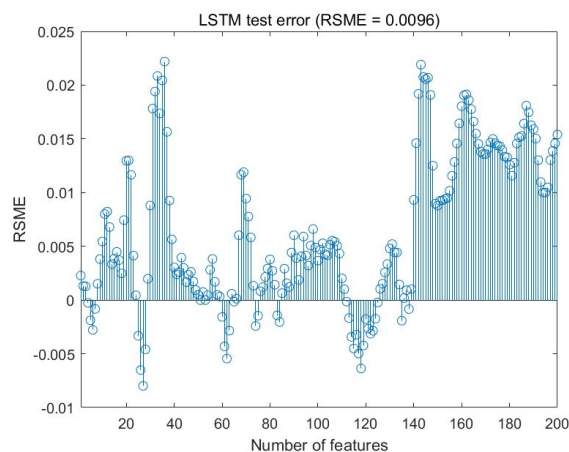

(a)

(b)

FIGURE 15. **LSTM test error**

(a)

IV. CONCLUSION

In this study, the performance of harmonic reducer is analyzed based on multi-variable condition monitoring method and performance degradation prediction method. The target includes the following contents: a) The condition monitoring of the harmonic reducer is carried out to realize the function of fault early alarm. b) Predict the

11

decline trend of harmonic reducer and provide technical support for preventive maintenance. We consider utilizing the MSET to realize the fault monitoring of the harmonic reducer, employing LargeVis to reduce the dimensionality of the multi-dimensional features, and combining the LSTM network to solve the prediction of the decline trend of the harmonic reducer. The accelerated life test verifies that the proposed method can not only give fault warning signals 18 minutes in advance in the sample with a life of 5.7 hours, but also has a strong ability to predict the degradation trend.

ACKNOWLEDGEMENT

REFERENCES

[1] J. Lee, H. A. Kao, and S. H. Yang, "Service innovation and smart analytics for Industry 4.0 and big data environment," *6th CIRP Conference on Industrial Product-Service Systems (IPSS)*, Univ Windsor, Windsor, CANADA, May 01-02 2014, vol. 16, in Procedia CIRP, 2014, pp. 3-8, doi: 10.1016/j.procir.2014.02.001. [Online]. Available: <Go to ISI>://WOS:000345448900001

[2] Q. Yang, X. Li, H. Cai, Y. M. Hsu, J. Lee, C. H. Yang, Z. L. Li, and M. Y. Lin,"Fault prognosis of industrial robots in dynamic working regimes: Find degradation in variations,"*Measurement*,vol. 173,P.108545,2021.

[3] Z. Zhi, L. Liu, D. Liu and C. Hu, "Fault detection of the harmonic reducer based on CNN-LSTM with a novel denoising algorithm," *IEEE Sensors Journal*, doi: 10.1109/JSEN.2021.3137992.

[4] Y. Li et al., "Performance margin modeling and reliability analysis for harmonic reducer considering multi-source uncertainties and wear," *IEEE Access*, vol. 8, pp. 171021-171033, 2020, doi: 10.1109/ACCESS.2020.3023543.

[5] J. L. Zhao, Y. X. Gu, S. Z. Yan and J. N. Wu, "Dynamic analysis of flexible space manipulators with harmonic reducer", *Key Eng. Mater.*, vol. 584, pp. 200-207, Sep. 2013.

[6] J. Zhao, J. Wu, S. Yan, J. Li and Y. Gu, "Dynamic modeling and motion precision analysis of spacecraft manipulator with harmonic drive considering the alternate thermal field in orbit," *Proc. Inst. Mech. Eng. G J. Aerosp. Eng.,* vol. 229, no. 1, pp. 135-148, Jan. 2015.

[7] B. Wei, J. X. Wang, G. W. Zhou, R. S. Yang, H. J. Zhou, and T. He, "Mixed lubrication analysis of modified cycloidal gear used in the RV reducer," *Proc. Inst. Mech. Eng. Part J.-J. Eng.* Article vol. 230, no. 2, pp. 121-134, Feb 2016, doi: 10.1177/1350650115593301.

[8] X. Zhang, G. D. Jiang, H. Zhang, X. L. Yun, and X. S. Mei, "Time-dependent reliability analysis of harmonic drive based on transient FEA and accelerated life test," *Eng. Comput.*, vol. 37, no. 7, pp. 2293-2317, 2020, doi: 10.1108/ec-10-2019.0466.

[9] H. M. Qian, Y. F. Li, and H. Z. Huang, "Time-variant reliability analysis for industrial robot RV reducer under multiple failure modes using Kriging model," *Reliab. Eng. Syst. Saf.*, vol. 199, p. 9, Jul 2020, Art no. 106936, doi: 10.1016/j.ress.2020.106936.

[10] Y. H. Liu, "Fault diagnosis based on SWPT and Hilbert transform," *International Conference on Advanced in Control Engineering and Information Science (CEIS)*, Dali, PEOPLES R CHINA, Aug 18-19 2011, vol. 15, in Procedia Engineering, 2011, doi: 10.1016/j.proeng.2011.08.726.

[11] J. J. H. Shiau and H. H. Lin, "Analyzing accelerated degradation data by nonparametric regression,", *IEEE Trans. Reliab.*, vol. 48, no. 2, pp. 149-158, Jun 1999. Available: <Go to ISI>://WOS:000082105800008.

[12] V. Crk and Iest, "Reliability assessment from degradation data," *Annual Reliability and Maintainability Symposium - 2000 Proceedings*, Proceedings: Annual Reliability and Maintainability Symposium. New York: IEEE, 2000, pp. 155-161.

[13] K. A. Doksum and A. Hoyland, "Models for variable-stress accelerated life testing experiments based on wiener-processes and the inverse gaussian distribution," *Technometrics*, vol. 34, no. 1, pp. 74-82, Feb 1992, doi: 10.2307/1269554.

[14] Y. Li et al., "Performance Margin Modeling and Reliability Analysis for Harmonic Reducer Considering Multi-Source Uncertainties and Wear," *Ieee Access*, vol. 8, pp. 171021-171033, 2020, doi: 10.1109/access.2020.3023543.

[15] W. Jun, D. Chao, X. Huan, and X. Yao, "Equipment Performance Degradation Modeling For Reliability Prediction Based on n-ARIMA model," in 3rd International Conference on Advances in Product Development and Reliability (PDR 2012), Wuhan, PEOPLES R CHINA, Jul 28-30 2012, vol. 544, in Advanced Materials Research, 2012, pp. 24, doi: 10.4028/www.scientific.net/AMR.544.24. [Online].

[16] J. Riascos-Ochoa, M. Sanchez-Silva, and G.-A. Klutke, "Modeling and reliability analysis of systems subject to multiple sources of degradation based on Levy processes," Probabilistic Engineering Mechanics, vol. 45, pp. 164-176, Jul 2016, doi: 10.1016/j.probengmech.2016.05.002.

[17] X. M. Ma, J. Yang, Y. Q. Zhang, "Life prediction reliability analysis of mine reducer based on particle swarm optimization,"*Coal Technology*, vol. 36, pp.237-239, Mar 2017.

[18] B. Wang, Y. G. Lei, N. P. Li, and N. B. Li, "A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings," *IEEE Trans. Reliab.*, vol. 69, no. 1, pp. 401-412, Mar 2020, doi: 10.1109/tr.2018.2882682.

[19] A. Prudhom, J. Antonino-Daviu, H. Razik, and V. Climente-Alarcon, "Time-frequency vibration analysis for the detection of motor damages caused by bearing currents," *Mechanical Systems and Signal Processing*, vol. 84, pp. 747-762, Feb 2017, doi: 10.1016/j.ymssp.2015.12.008.

[20] P. Wang, R. X. Gao, and W. A. Woyczynski, "Levy Process-Based Stochastic Modeling for Machine Performance Degradation Prognosis," *Ieee Transactions on Industrial Electronics*, vol. 68, no. 12, pp. 12760-12770, Dec 2021, doi: 10.1109/tie.2020.3047037.

[21] K. Lu, W. K. Wang, X. Meng, J. Li, and X. Pang, "A Multivariate State Estimation Technique and Multilayer Forward Neural Network Framework on Steam Turbine Valve Flow Curve Simulation," *IOP Conference Series Earth and Environmental Science*, vol. 526, p. 012153, 2020.

[22] J. Tang, J. Z. Liu, M. Zhang, Q. Z. Mei, and Acm, "Visualizing Large-scale and High-dimensional Data," *in 25th International Conference on World Wide Web (WWW)*, Montreal, CANADA, May 11-15 2016, NEW YORK: Assoc Computing Machinery, 2016, pp. 287-297, doi: 10.1145/2872427.2883041. [Online].

[23] X. Song et al., "Pedestrian Trajectory Prediction Based on Deep Convolutional LSTM Network," (in English), *IEEE Trans*. Intell. Transp. Syst., Article vol. 22, no. 6, pp. 3285-3302, Jun 2021, doi: 10.1109/tits.2020.2981118.