Journal of Control Science and Engineering

Fault Detection, Isolation, and Prognosis for Complex System

Lead Guest Editor: Chunhui Zhao
Guest Editors: Furong Gao, Yuan Yao, Zhixing Cao, and Yongji Fu

Fault Detection, Isolation, and Prognosis for Complex System

Journal of Control Science and Engineering

Fault Detection, Isolation, and Prognosis for Complex System

Lead Guest Editor: Chunhui Zhao
Guest Editors: Furong Gao, Yuan Yao, Zhixing Cao, and Yongji Fu

Editorial Board

Contents

Hindawi
Journal of Control Science and Engineering
Volume 2018, Article ID 8783158, 2 pages
https://doi.org/10.1155/2018/8783158

Editorial

Fault Detection, Isolation, and Prognosis for Complex System

Chunhui Zhao,[1] **Furong Gao**,[2] **Yuan Yao**,[3] **Zhixing Cao**,[4] **and Yongji Fu**[5]

[1]*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China*
[2]*Hong Kong University of Science and Technology, Kowloon, Hong Kong*
[3]*National Tsing Hua University, Hsinchu, Taiwan*
[4]*Harvard University, Cambridge, MA, USA*
[5]*Hill-Rom, Batesville, IN, USA*

Correspondence should be addressed to Chunhui Zhao; chhzhao@zju.edu.cn

Received 12 December 2017; Accepted 12 December 2017; Published 23 January 2018

A complex system can be thought of as multiple interdependent working subsystems. With the increasing level of complexity, it may present more frequent unstable operation statuses. Once system faults have occurred, they can cause unrecoverable losses and unacceptable environmental pollution and so forth. It thus demands more effective and efficient techniques to monitor operation status, detect the occurrence and propagation of faults, and enable suitable decision making, before such anomalies result in great damage. As one of the most active research areas over the last few decades, fault diagnosis including detection, isolation, and prognosis has been of importance and necessary for improving the economy and safety of a complex system, ranging from industrial processes, such as steel production, papermaking, car manufacturing, and mineral processing, to biological processes and so forth.

In the era of big data of process industries, new challenge emerges for fault diagnosis with amount of data grown exponentially. In particular, there are many uncertainties in the system which shows the complexity of characteristics, including multimode and dynamics, multilevel and multiscale, nonlinearities, and strong coupling effects amongst the variables. This special issue aims to bring together state-of-the-art research contributions on fault diagnosis methods and their different applications, as well as future trends. Potential topics of this special issue include fault diagnosis problems for batch processes, fault prognosis for industrial processes, fault self-recovery/self-healing control, product quality monitoring and prediction, process performance assessment, incipient fault detection and diagnosis, fault classification and discrimination, data-driven approaches and knowledge-based approaches, and intelligence-based supervisory control. We received a total of 23 submissions, and after two rounds of rigorous review, 9 papers were accepted.

In the paper "Tooth Fracture Detection in Spiral Bevel Gears System by Harmonic Response Based on Finite Element Method," Y. Chen et al. establish a three-dimensional model and finite element model of the Gleason spiral bevel gear pair. The model considers the effect of tooth root fracture on the system due to fatigue. Finite element method is used to compute the mesh generation, set the boundary condition, and carry out the dynamic load. The harmonic response spectrums of the base under tooth fracture are calculated and the influences of main parameters on monitoring failure are investigated as well.

In the paper "Modeling of Complex Life Cycle Prediction Based on Cell Division," F. Zhang et al. propose to establish a life prediction model by studying the regularity of the cell life in the natural biological evolution. It has moderate complexity which can effectively simulate the state of cell division and does not need a large amount of data in comparison with the conventional life prediction methods. By controlling the concentration of some key factors in cell division, the effect on the normal life of a cell can be expressed intuitively.

In the paper "Degradation Data-Driven Remaining Useful Life Estimation in the Absence of Prior Degradation Knowledge," Y. Yu et al. propose a fusion algorithm to combine the excellent modeling ability of Bayesian updating

method for the multilevel data and the prominent estimation ability of ECM algorithm for incomplete data. Residual life distributions and posterior distributions are first calculated through the Bayesian updating method based on random initial a priori distributions. Then the a priori distributions are revised and improved for future predictions by the ECM algorithm.

In the paper "Parameter Selection Method for Support Vector Regression Based on Adaptive Fusion of the Mixed Kernel Function," H. Wang and D. Xu use a 5th-degree cubature Kalman filter to estimate the parameters. In this way, they realize the adaptive selection of mixed kernel function weighted coefficients, the kernel parameters, and the regression parameters. Compared with a single kernel function, unscented Kalman filter (UKF) support vector regression algorithms, and genetic algorithms, the decision regression function obtained by the proposed method has better generalization ability and higher prediction accuracy.

In the paper "Time-Varying Fault Diagnosis for Asynchronous Multisensor Systems Based on Augmented IMM and Strong Tracking Filtering," Y. Hu et al. propose a fault detection, isolation, and estimation approach based on Interactive Multimodel (IMM) fusion filtering and Strong Tracking Filtering (STF) for asynchronous multisensors dynamic systems. Time-varying fault is considered and a candidate fault model is built by augmenting the unknown fault amplitude directly into the system state for each kind of possible fault mode. Asynchronous IMM fusion filtering is performed to the multiple model sets to detect and isolate the fault based on model probabilities to avoid the dilemma of traditional IMM-based approaches.

In the paper "A Bayesian Approach to Control Loop Performance Diagnosis Incorporating Background Knowledge of Response Information," S. Zhou and Y. Wang incorporate background knowledge into Bayesian inference to isolate the problem source degrading the control loop performance. In an effort to reduce dependence on the amount of historical data available, they consider a general kind of background knowledge, known as response information, and translate it to constraints on the underlying probability distributions. In this way, the dimensionality of the observation space is reduced and thus the diagnosis can be more reliable.

In the paper "WOS-ELM-Based Double Redundancy Fault Diagnosis and Reconstruction for Aeroengine Sensor," Z. Zhao et al. propose a double redundancy diagnosis approach based on Weighted Online Sequential Extreme Learning Machine (WOS-ELM) in order to diagnose sensor fault of aeroengine more quickly and accurately. It assigns different weights to old and new data and implements weighted dealing with the input data to get more precise training models. The WOS-ELM-based double redundancy fault diagnosis and reconstruction approach contains two diagnosis models which can real-timely detect the hard fault and soft fault much earlier.

In the paper "Development of Fault Identification System for Electric Servo Actuators of Multilink Manipulators Using Logic-Dynamic Approach," V. Filaretov et al. present synthesis method of faults identification systems based on logic-dynamic approach for electric servo actuators of multilink manipulators which are described by nonlinear equations with significantly changing coefficients. An advantage of this approach is that it allows studying systems with nonsmooth nonlinearities by linear methods only.

In the paper "Fault Diagnosis of Nonlinear Uncertain Systems with Triangular Form," Q. Ding et al. propose a fault diagnosis approach for a class of nonlinear uncertain systems with triangular form based on the extended state observer (ESO) of the active disturbance rejection controller and linearization of dynamic compensation. This method is reported to be of utility and simple in construction and parameter tuning.

Acknowledgments

The guest editorial team would like to thank authors of all the papers submitted to this special issue. Given the space limitations, a number of high quality contributions could not be accommodated. The editors also wish to thank the anonymous reviewers, some of whom helped with multiple review assignments. We hope that you will enjoy reading this special issue devoted to this exciting and fast-evolving field as much as we have done.

Chunhui Zhao
Furong Gao
Yuan Yao
Zhixing Cao
Yongji Fu

Hindawi
Journal of Control Science and Engineering
Volume 2018, Article ID 5205698, 8 pages
https://doi.org/10.1155/2018/5205698

Research Article

Time-Varying Fault Diagnosis for Asynchronous Multisensor Systems Based on Augmented IMM and Strong Tracking Filtering

Yanyan Hu,[1,2] **Xiaoling Xue**,[1] **Zengwang Jin**,[1] **and Kaixiang Peng**[1]

[1]*School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[2]*Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China*

Correspondence should be addressed to Kaixiang Peng; kaixiang@ustb.edu.cn

Received 29 June 2017; Accepted 12 November 2017; Published 2 January 2018

Academic Editor: Chunhui Zhao

A fault detection, isolation, and estimation approach is proposed in this paper based on Interactive Multimodel (IMM) fusion filtering and Strong Tracking Filtering (STF) for asynchronous multisensors dynamic systems. Time-varying fault is considered and a candidate fault model is built by augmenting the unknown fault amplitude directly into the system state for each kind of possible fault mode. By doing this, the dilemma of predetermining the fault extent as model design parameters in traditional IMM-based approaches is avoided. After that, the time-varying fault amplitude is estimated based on STF using its strong ability to track abrupt changes and robustness against model uncertainties. Through fusing information from multiple sensors, the performance of fault detection, isolation, and estimation is approved. Finally, a numerical simulation is performed to demonstrate the feasibility and effectiveness of the proposed method.

1. Introduction

In recent years, modern engineering systems have become huge in investment, large in scale, and more and more sophisticated in structure. As a result, faults in these complex systems may lead to enormous losses. Consequently, fault detection and diagnosis (FDD) has attracted more and more attentions as an effective method to reduce the accident risk and enhance the security of systems [1, 2]. The main task of FDD is to perform detection, isolation, and identification of faults in modern systems, that is, to determine whether faults happen, locate the faults, and estimate fault amplitudes [3, 4]. System with faults consists of state evolution in continuous time and parameter or structure changes in discrete time and thus is a typical hybrid system. As the most cost-effective adaptive approach for state estimation of hybrid dynamic systems, Interactive Multimodel (IMM) has been successfully used for FDD of modern engineering systems [5, 6]. In [7], Zhang and Li proposed an integrated framework for FDD of sensor and actuator failures based on IMM. The IMM filtering was performed to model set consisting of the normal model and fault models corresponding to total and partial sensor and/or actuator faults, where extents of partial faults were taken as model parameters and should be predetermined. Ru and Li in [8] proposed an fault diagnosis algorithm by combining IMM and Maximum Likelihood Estimation (MLE), where the fault was detected and located by IMM, and then the fault amplitude was determined by MLE. Zhao et al. in [9] proposed an improved IMM-based FDD, which concentrated on dealing with the mismodeled transition probabilities. A modification operator was used to partition the posterior mode probabilities heuristically and automatically to make the residual error under the true modes approach the white Gaussian process. For aircraft actuator faults, fuzzy logic was utilized in [10] to tune the transition probabilities to make the fault detection process smooth and improve the diagnosis accuracy. Moreover, a novel sensor fault detection, isolation, and identification approach was introduced in [11] for gas turbine engines by combining the proposed multiple hybrid Kalman filters with

a modified generalized likelihood ratio method. In addition, IMM-based fault diagnosis techniques have been applied to unmanned aerial vehicles [12], satellites attitude control systems [13], and active fault-tolerant control [14].

On the other hand, with the rapid development of sensor techniques, the number and type of sensors used for system monitoring in modern engineering systems are increased greatly [15, 16]. Consequently, how to effectively integrate the information from multiple sensors to reduce system uncertainties and improve the FDD accuracy is becoming an important research issue [17, 18]. A multisensor fusion and fault detection approach for air traffic surveillance was introduced in [19] based on hybrid estimation. The bias fault and large deviation fault of multilateration and automatic dependence surveillance-broadcast were considered. The proposed method ran two IMM filters in parallel, each for one sensor, and the sensor fault was detected based on the residuals generated by individual filters. The work in [19] is subjected to synchronous sensors. However, in practice, sensors may have different sampling rates, initial sample times, and communication delays, which results in asynchronous measurements [20, 21]. Although asynchronous multisensors have widespread applications, studies aiming at the FDD for asynchronous multisensor systems are relatively much scarce. An actuator fault diagnosis strategy for dynamic systems with multiple asynchronous sensors was presented in [22], where the multiplicative fault factor was considered. An IMM fusion filter was adopted to detect and isolate the fault, while, after that, an augmented Kalman filter was used to estimate the unknown fault factor. Nevertheless, the work in [22] is still limited to time-invariant or slowly time-varying fault because of the sensitivity of Kalman filter to model mismatch.

The aim of this paper is to study the time-varying fault detection, isolation, and estimation problem of stochastic dynamic systems with multiple asynchronous sensors. In existing IMM-based FDD approaches, the unknown fault extent presents as model parameters in candidate fault model and needs to be predetermined in the process of model set design. Meanwhile, since fault extent actually takes value from a continuous interval, several fault models with distinct fault extents for a given kind of fault need to be included in the model set in order to have satisfactory coverage of all possible fault conditions. Different from existing approaches, the proposed FDD strategy in this paper regards the fault amplitude as unknown state variables and augments it directly into the system state to build the candidate fault model. By doing this, for each kind of fault, only one fault model is needed and the dilemma of predetermining the fault extent as model parameters in the model set design process is avoided. Then the asynchronous IMM fusion filtering is performed to the model set consisting of normal model and augmented fault models, and the fault is detected and isolated simultaneously based on the posterior model probabilities. Finally, STF is utilized to jointly estimate the system state and time-varying fault amplitude by fusing all asynchronous measurements from sensors and making use of its strong robustness to model uncertainties.

The rest of this paper is organized as follows. A description of time-varying FDD problem for stochastic dynamic systems with asynchronous sensors is presented in Section 2. The process of model set design is discussed in Section 3. Section 4 illustrates the fault detection and diagnosis based on asynchronous IMM fusion filtering. Section 5 presents the estimation algorithm using STF to estimate the amplitude of the fault. In Section 6, simulation examples are provided to show the effectiveness and feasibility of our method. Finally, conclusions are drawn in Section 7.

2. Problem Formulation

Consider the following continuous-time linear dynamic system:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + G(t)w(t) + F(t)e_m f(t), \tag{1}$$

where $x(t) \in R^{d_x}$ denotes the d_x dimension system state, $u(t) \in R^{d_u}$ denotes the d_u dimension actuator input, and $w(t)$ is the system process noise. $f(t)$ is the scalar fault amplitude signal with e_m denoting the fault direction, where e_m is the mth column of unit matrix I_{d_f} and $m \in \{1, 2, \ldots, d_f\}$. $A(t), B(t), G(t), F(t)$ are correspondingly coefficient matrices with appropriate dimensions. Suppose there are N asynchronous sensors with distinct sample rates and initial sampling times observing system (1).

Let N_k denote all the measurements of the N asynchronous sensors in a time interval $(t_{k-1}, t_k]$. For the fusion time k, we order all of the N_k measurements according to the chronological order to get the measure series $\{y_k^i\}_{i=1}^{N_k}$, where $y_k^i \in R^{d_y^i}$. Let t_k^i be the sampling time of y_k^i; then we have $t_k^i \le t_k^{i+1}$, where the equality holds when y_k^i and y_k^{i+1} are measured at the same time instant. The observation equation is given by

$$y_k^i = H_k^i x\left(t_k^i\right) + v\left(t_k^i\right), \tag{2}$$

where v_k^i is the zero mean Gaussian white noise with the covariance matrix $E\{v_k^i (v_k^j)^T\} = R_k^i \delta_{ij}$, and we assume that v_k^i is uncorrelated with the process noise $w(t)$.

In this paper, time-varying fault $f(t)$ is considered. Actually, in practical applications, the development of fault is relatively slow at the beginning stage and after a certain period of time will become more rapid. Consequently, for fault detection at the initial stage after fault occurs, we assume that it is slowly time-varying and thus it can be regarded as constant in a given fusion interval. Based on above assumption, an augmented IMM is used to simultaneously detect and locate the fault. Then, after the fault is isolated, the STF is performed to jointly estimate the system state and the time-varying fault amplitude $f(t)$, using its strong ability to track abrupt changes and robustness against model uncertainties.

3. Model Set Designs of the Augmented IMM

As we said above, for fault detection and isolation, an augmented IMM is used in this paper. The model set design of the augmented IMM is introduced in this section.

When the mth fault occurs, the system equation is given by (1). At the beginning stage, we assume that, in a small time interval $(t_2, t_1]$, $f(t)$ is time-invariant, where the time interval $(t_2, t_1]$ should be no longer than the fusion interval $T = t_k - t_{k-1}$. Then, from (1), we have the corresponding discrete-time state transition equation

$$x(t_2) = \Phi(t_2, t_1)\, x(t_1) + u(t_2, t_1) + \xi^m(t_2, t_1)\, f(t_2) + w(t_2, t_1), \tag{3}$$

where $\Phi(t_2, t_1)$ is the state transition matrix from t_1 to t_2, and

$$\begin{aligned} u(t_2, t_1) &= \int_{t_1}^{t_2} \Phi(t_2, \tau)\, B(\tau)\, u(\tau)\, d(\tau) \\ \xi^m(t_2, t_1) &= \int_{t_1}^{t_2} \Phi(t_2, \tau)\, F(\tau)\, e_m d(\tau) \\ w(t_2, t_1) &= \int_{t_1}^{t_2} \Phi(t_2, \tau)\, G(\tau)\, w(\tau)\, d(\tau) \\ \Omega(t_2, t_1) &= E\left\{w(t_2, t_1)\, w^T(t_2, t_1)\right\} \\ &= \int_{t_1}^{t_2} \Phi(t_2, \tau)\, G(\tau)\, Q(\tau)\, G^T(\tau)\, \Phi^T(t_2, \tau)\, d(\tau). \end{aligned} \tag{4}$$

Similarly, for fusion interval $(t_{k-1}, t_k]$, we have

$$x(t_k) = \Phi(t_k, t_{k-1})\, x(t_{k-1}) + u(t_k, t_{k-1}) + w(t_k, t_{k-1}) + \xi^m(t_k, t_{k-1})\, f(t_k). \tag{5}$$

We define

$$\overline{x}_k = \begin{bmatrix} x(t_k) \\ f(t_k) \end{bmatrix} \tag{6}$$

$$\overline{\Phi}^m_{k-1} = \begin{bmatrix} \Phi(t_k, t_{k-1}) & \xi^m(t_k, t_{k-1}) \\ 0 & 1 \end{bmatrix} \tag{7}$$

$$\overline{u}^x_{k-1} = \begin{bmatrix} u(t_k, t_{k-1}) \\ 0 \end{bmatrix} \tag{8}$$

$$\overline{w}_{k-1} = \begin{bmatrix} w(t_k, t_{k-1}) \\ 0 \end{bmatrix}. \tag{9}$$

Then after extending the fault amplitude $f(t_k)$ to the state vector as (6), we have

$$\overline{x}_k = \overline{\Phi}^m_{k-1}\overline{x}_{k-1} + \overline{u}^x_{k-1} + \overline{w}_{k-1}, \tag{10}$$

$$\overline{\Omega}_{k-1} = E\left\{\overline{w}_{k-1}\overline{w}^T_{k-1}\right\} = \operatorname{diag}\left\{\Omega(t_k, t_{k-1}), 0\right\}, \tag{11}$$

where (10) is the fault model corresponding to the mth failure. Let $m = 0$ denote the normal model without fault, while let $m = 1, 2, \ldots, d_f$ be the d_f number of fault models. The IMM model set is composed of the above $d_f + 1$ models.

4. Fault Detection and Isolation Using Augmented IMM

A complete cycle of the IMM-based-FDI scheme is discussed below.

4.1. Input Mixing. Given $Y_k := \{y^i_k\}^{N_k}_{i=1}$, $Y^k := \{Y_l\}^k_{l=1}$, then Y^k is the cumulative measurement set of all the asynchronous sensors until the fusion time t_k. Define

$$\begin{aligned} \widehat{x}^m_k &= E\left\{\overline{x}_k \mid \widetilde{m}(t_k) = m, Y^k\right\} \\ P^m_k &= \operatorname{Cov}\left\{\overline{x}_k \mid \widetilde{m}(t_k) = m, Y^k\right\} \\ \mu^m_k &= \operatorname{Prob}\left(\widetilde{m}(t_k) = m \mid Y^k\right), \end{aligned} \tag{12}$$

where $\widetilde{m}(t_k) = m$ denotes that the mth model is valid at t_k. The initial value of the mth basic filter is

$$\begin{aligned} \breve{x}^m_{k-1} &= \sum_{i=0}^{d_f} \widehat{x}^i_{k-1}\breve{\mu}^{i|m}_{k-1} \\ \breve{P}^m_{k-1} &= \sum_{i=0}^{d_f}\left[P^i_{k-1} + \left(\breve{x}^m_{k-1} - \widehat{x}^i_{k-1}\right)\left(\breve{x}^m_{k-1} - \widehat{x}^i_{k-1}\right)^T\right]\breve{\mu}^{i|m}_{k-1} \\ \breve{\mu}^{i|m}_{k-1} &= \frac{C^m_k}{\sum_{i=0}^{d_f} C^m_k}, \end{aligned} \tag{13}$$

where $m = 0, 1, \ldots, d_f$, $\lambda^{i,m} = \operatorname{Prob}(\widetilde{m}(t_k) = m \mid \widetilde{m}(t_{k-1}) = i)$ is the transition probability from model i at t_{k-1} to model m at t_k, and $C^m_k = \lambda^{i,m}\mu^i_{k-1}$.

4.2. Model-Conditioned Fusion Filtering. From the state transition equation (3), we have

$$x(t_k) = \Phi\left(t_k, t^i_k\right)x\left(t^i_k\right) + u\left(t_k, t^i_k\right) + \xi^m\left(t_k, t^i_k\right)f\left(t^i_k\right) + w\left(t_k, t^i_k\right). \tag{14}$$

Substituting (14) into measurement equation (2), we have

$$\begin{aligned} y^i_k &= H^i_k\Phi^{-1}\left(t_k, t^i_k\right)\left[x(t_k) - u\left(t_k, t^i_k\right)\right. \\ &\quad \left. - \xi^m\left(t_k, t^i_k\right)f(t_k) - w\left(t_k, t^i_k\right)\right] + v^i_k = \overline{H}^{m,i}_k\overline{x}^m_k \\ &\quad + u^i_k + \eta^i_k, \end{aligned} \tag{15}$$

where

$$\begin{aligned} \overline{H}^{m,i}_k &= \Pi^i_k\left[I - \xi^m\left(t_k, t^i_k\right)\right] \\ \eta^i_k &= v^i_k - \Pi^i_k w\left(t_k, t^i_k\right) \\ u^i_k &= -\Pi^i_k u\left(t_k, t^i_k\right) \\ \Pi^i_k &= H^i_k\Phi^{-1}\left(t_k, t^i_k\right). \end{aligned} \tag{16}$$

Then we define

$$\overline{y}_k = \left[\left(y_k^1\right)^T, \left(y_k^2\right)^T, \ldots, \left(y_k^{N_k}\right)^T\right]^T. \tag{17}$$

Consequently, in the case of mth fault, y_k can be regarded as an equivalent measurement of the augmented system state $\overline{x}_k$ at time t_k, and the equivalent measurement equation is

$$\overline{y}_k = \overline{H}_k^m \overline{x}_k + \overline{u}_k + \overline{\eta}_k, \tag{18}$$

where

$$\overline{H}_k^m = \left[\left(\overline{H}_k^{m,1}\right)^T, \left(\overline{H}_k^{m,2}\right)^T, \ldots, \left(\overline{H}_k^{m,N_k}\right)^T\right]^T. \tag{19}$$

$\overline{u}_k$ and $\overline{\eta}_k$ have the same form as $\overline{H}_k^m$ except for replacing $\overline{H}_k^{m,i}$ with u_k^i and η_k^i, respectively.

Equivalent measurement noise $\overline{\eta}_k$ is a Gaussian white noise with zero mean and covariance matrix $\overline{R}_k = E\{\overline{\eta}_k \overline{\eta}_k^T\} = (\overline{R}_k^{i,j})$, and we further have

$$\overline{R}_k^{i,j} = E\left\{\eta_k^i \left(\eta_k^j\right)^T\right\} = R_k^i \delta_{i,j} + \Pi_k^i \Omega\left(t_k, t_k^i\right)\left(\Pi_k^i\right)^T. \tag{20}$$

At the same time, the equivalent measurement noise $\overline{\eta}_k$ and the process noise $w(t_k, t_{k-1})$ are relevant on account of the common process noise

$$\begin{aligned}\Psi_k = E\left\{w\left(t_k, t_{k-1}\right)\left(\overline{\eta}_k\right)^T\right\} = -\Big[&\Omega\left(t_k, t_k^1\right) \\ &\cdot \left(\Pi_k^1\right)^T, \Omega\left(t_k, t_k^2\right)\left(\Pi_k^2\right)^T, \ldots, \Omega\left(t_k, t_k^{N_k}\right)\left(\Pi_k^{N_k}\right)^T\Big].\end{aligned} \tag{21}$$

Now for the system composed of the augmented state equation (10) and the equivalent measurement equation (18), the updating equation of the $\hat{x}_k^m$ and error covariance P_k^m can be obtained as

$$\begin{aligned}\hat{x}_{k|k-1}^m &= E\left\{\overline{x}_k \mid \widetilde{m}\left(t_k\right) = m, Y^{k-1}\right\} \\ &= \overline{\Phi}_{k-1}^m \breve{x}_{k-1}^m + \overline{u}_{k-1}^x\end{aligned} \tag{22}$$

$$\begin{aligned}P_{k|k-1}^m &= \mathrm{Cov}\left\{\overline{x}_k \mid \widetilde{m}\left(t_k\right) = m, Y^k\right\} \\ &= \overline{\Phi}_{k-1}^m \breve{P}_{k-1}^m \left(\overline{\Phi}_{k-1}^m\right)^T + \overline{\Omega}_{k-1}\end{aligned} \tag{23}$$

$$\hat{x}_k^m = \hat{x}_{k|k-1}^m + \Theta_k^m \left(\Xi_k^m\right)^{-1} \tilde{y}_{k|k-1}^m \tag{24}$$

$$P_k^m = P_{k|k-1}^m - \Theta_k^m \left(\Xi_k^m\right)^{-1} \left(\Theta_k^m\right)^T, \tag{25}$$

where

$$\begin{aligned}\overline{\Psi}_k &= \left[\Psi_k^T, 0_{\sum_{i=1}^{N_k} d_y^i \times 1}\right]^T \\ \Theta_k^m &= P_{k|k-1}^m \left(\overline{H}_k^m\right)^T + \overline{\Psi}_k \\ \Xi_k^m &= \overline{H}_k^m P_{k|k-1}^m \left(\overline{H}_k^m\right)^T + \overline{R}_k + \overline{H}_k^m \overline{\Psi}_k + \left(\overline{H}_k^m \overline{\Psi}_k\right)^T \\ \tilde{y}_{k|k-1}^m &= \overline{y}_k - \overline{H}_k^m \hat{x}_{k|k-1}^m - \overline{u}_k.\end{aligned} \tag{26}$$

4.3. Model Probability Update. The model probability of the mth fault should be

$$\begin{aligned}\mu_k^m &= \mathrm{Prob}\left(\widetilde{m}\left(t_k\right) = m \mid Y^k\right) \\ &= \frac{\mathrm{Prob}\left(\widetilde{m}\left(t_k\right) = m \mid Y^{k-1}\right) p\left(\overline{y}_k \mid \widetilde{m}\left(t_k\right) = m, Y^{k-1}\right)}{\sum_{i=0}^d \mathrm{Prob}\left(\widetilde{m}\left(t_k\right) = i \mid Y^{k-1}\right) p\left(\overline{y}_k \mid \widetilde{m}\left(t_k\right) = i, Y^{k-1}\right)},\end{aligned} \tag{27}$$

where

$$\begin{aligned}&\mathrm{Prob}\left(\widetilde{m}\left(t_k\right) = m \mid Y^{k-1}\right) = \sum_{i=0}^{d_f} C_k^m \\ &p\left\{\overline{y}_k \mid \widetilde{m}\left(t_k\right) = m, Y^{k-1}\right\} = (2\pi)^{-(1/2)\sum_{i=1}^{N_k} d_y^i} \left|\Xi_k^m\right|^{-1/2} \\ &\quad \cdot \exp\left\{-\frac{1}{2}\left(\tilde{y}_{k|k-1}^m\right)^T \left(\Xi_k^m\right)^{-1} \tilde{y}_{k|k-1}^m\right\}.\end{aligned} \tag{28}$$

4.4. Output Combination. Finally from the total probability formula, we have

$$\begin{aligned}\hat{x}_k &= E\left\{\overline{x}_k \mid Y^k\right\} = \sum_{m=0}^{d_f} \hat{x}_k^m \mu_k^m \\ P_k &= E\left\{\left(\hat{x}_k - \overline{x}_k\right)\left(\hat{x}_k - \overline{x}_k\right)^T\right\} \\ &= \sum_{m=0}^{d_f} \left[P_k^m + \left(\hat{x}_k - \hat{x}_k^m\right)\left(\hat{x}_k - \hat{x}_k^m\right)^T\right] \mu_k^m.\end{aligned} \tag{29}$$

4.5. Fault Detection and Isolation. Once $\mu_k^i = \max\{\mu_k^m\}_{m=1}^{d_f} > \mu_T$, and for $j = 1, 2, \ldots, L$, we also have $\mu_{k-j}^i > \mu_T$; then the conclusion can be drawn that there is the ith failure in the system, where μ_T is the predetermined detection threshold and L can be regarded as the length of the time window.

Remark 1. The first four steps above constitute the proposed fusion IMM algorithm. The augmented state is estimated under each possible current model through fusing asynchronous measurements from N sensors by Section 4.2, with each filter reinitialed by Section 4.1. The mixed input to each filter is a combination of the previous model-conditioned estimates with mixing probabilities. By this input interaction step, IMM achieves the best compromise between complexity and performance.

5. Fault Amplitude Estimation Based on STF

Once the fault is successfully detected and isolated, the STF could be used to track the development of fault amplitude using its strong ability to track abrupt changes and strong robustness to model mismatch [23, 24]. The augmented estimate $\hat{x}_k^m$ and its estimation error covariance matrix P_k^m at t_k can be obtained based on STF as (22)–(25) with $\breve{x}_{k-1}^m$ in (22)

replaced by $\hat{x}_{k-1}^{m}$ and the one step predicted error covariance $P_{k|k-1}^{m}$ given by (23) replaced by

$$P_{k|k-1}^{m} = \operatorname{diag}\left\{\varphi_k(1), \varphi_k(2), \ldots, \varphi_k(d_x+1)\right\} \times \overline{\Phi}_{k-1}^{m} P_{k-1}^{m} \left(\overline{\Phi}_{k-1}^{m}\right)^{T} + \overline{\Omega}_{k-1}, \tag{30}$$

where for $i = 1, 2, \ldots, d_x + 1$

$$\varphi_k(i) = \begin{cases} \partial_k(i)\,\varpi_k, & \partial_k(i)\,\varpi_k > 1 \\ 1, & \partial_k(i)\,\varpi_k \le 1 \end{cases}$$

$$\varpi_k = \frac{\vartheta_k - \beta \operatorname{tr}\left(\overline{R}_k + \overline{H}_k^{m} \overline{\Omega}_{k-1} \left(\overline{H}_k^{m}\right)^{T}\right)}{\sum_{i=1}^{d_x+1} \partial_k(i)\, W_k(i,i)}$$

$$W_k = \overline{\Phi}_{k-1}^{m} \breve{P}_{k-1}^{m} \left(\overline{\Phi}_{k-1}^{m}\right)^{T} \left(\overline{H}_k^{m}\right)^{T} \left(\overline{H}_k^{m}\right) \tag{31}$$

$$\vartheta_k = \begin{cases} \left(\tilde{y}_1^{m}\right)^{T} \tilde{y}_1^{m}, & k = 1 \\ \dfrac{\rho\vartheta_{k-1} + \left(\tilde{y}_{k|k-1}^{m}\right)^{T} \tilde{y}_{k|k-1}^{m}}{1+\rho}, & k \ge 2, \end{cases}$$

where $0.95 \le \rho < 1$ is the forgetting factor and $\beta \ge 1$ and $\partial_k(i) \ge 1, i = 1, 2, \ldots, d_x + 1$ are the predetermined filter parameters.

6. Simulation Results

In this section, simulation results are provided to verify the proposed algorithm. Consider the dynamic system described by (1) with

$$A(t) = \begin{bmatrix} 0 & 1 & 0 & -0.08 \\ 0 & 0 & 0 & -0.2 \\ 0 & 0.08 & 0 & 1 \\ 0 & 0.2 & 0 & 0 \end{bmatrix} \tag{32}$$

$$B(t) = G(t) = F(t) = \begin{bmatrix} 0.5 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 1 \end{bmatrix}^{T},$$

where the control input is set as $u(t) = \begin{bmatrix} 1 & 1 \end{bmatrix}^{T}$ and the noise variance $Q(t) = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$. When the system works normally, $f(t) = 0$. Suppose the system is observed by three sensors with observation matrices

$$H_{(1)} = H_{(2)} = H_{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{33}$$

and measurement noise covariance matrices

$$R_{(1)} = R_{(2)} = R_{(3)} = \begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}. \tag{34}$$

The initial sampling instants of the three sensors are $t_{(1)} = 0.1$ s, $t_{(2)} = 0.2$ s, and $t_{(3)} = 0.3$ s, respectively. The sample periods of Sensor 1, Sensor 2, and Sensor 3 are $T_{(1)} = T_{(2)} = 0.6$ s and $T_{(3)} = 0.8$ s. The fusion period of the fusion center is $T = 1$ s. Apparently, these three sensors are asynchronous.

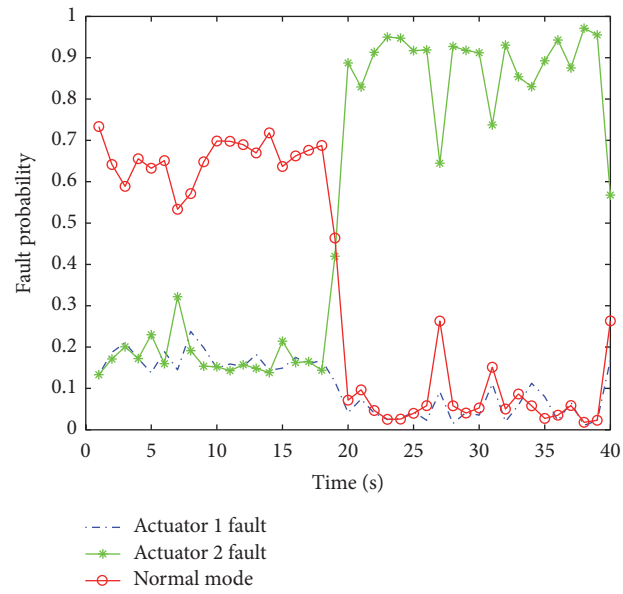


Figure 1: Model probability.

There exist two actuators in this system. Here we assume that Actuator 2 has a time-varying fault which occurs at $\tau_f = 16.5$ s with amplitude $f(t) = 1.5 + 0.1e^{0.04(t-\tau_f)}$. The proposed method is used to diagnose the fault in the system. To perform multimodel fusion filtering, three augmented models are constructed corresponding to three possible situations: normal mode, Actuator 1 fault, and Actuator 2 fault. Figure 1 gives the model probability evolutions during the simulation time in one run. From Figure 1 we can see that the probability of the normal model is dominant before the fault happens. After $t = 16.5$ s when the fault of Actuator 2 occurs, the probability of normal model has a sudden drop, while that of Actuator 2 fault increases rapidly. Therefore, the fault can be correctly detected and located. After that, the augmented IMM is switched to STF to track the variation of the fault amplitude. In the simulation, the value of detection time window L is set to 3, which means we need at least 3 s to conform the fault, which results in the detection delay in Figure 1. Figures 2 and 3 show the root mean-squared errors (RMSE) of the state estimation and the fault amplitude estimation over 500 Monte Carlo runs, respectively. It can be seen from Figures 2 and 3 that both the fused state estimates and the fused fault amplitude estimate have the smallest RMSEs compared to single sensors, which illustrate the effectiveness of the proposed fusion algorithm. Meanwhile, Sensor 3 has worse estimation performance than Sensor 1 and Sensor 2 since it has a lower sampling rate.

7. Conclusion

In this paper, we have presented a new approach for detecting, isolating, and estimating time-varying fault based on IMM and STF for asynchronous multisensor systems. An

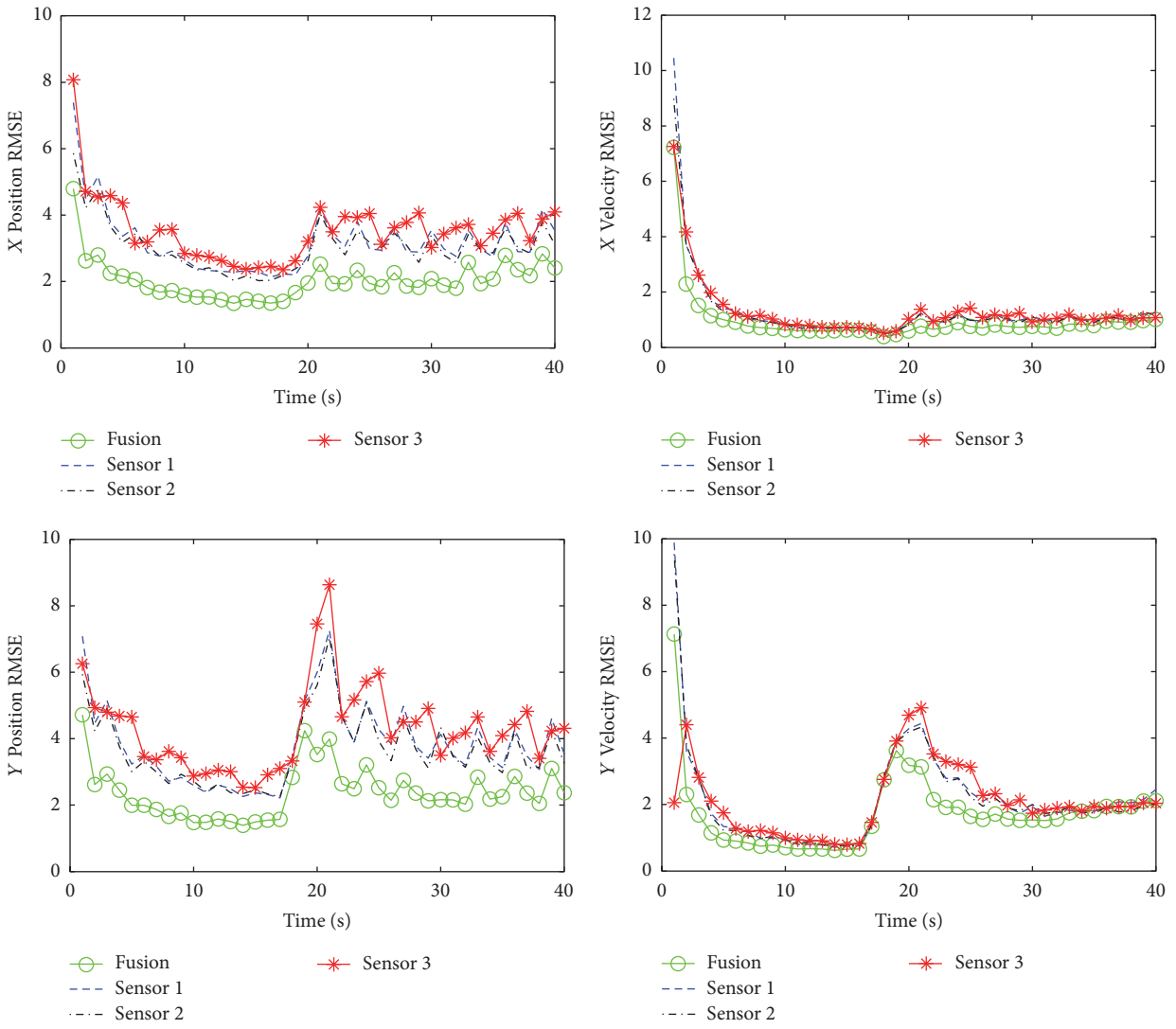


Figure 2: RMSE curves of state estimation.

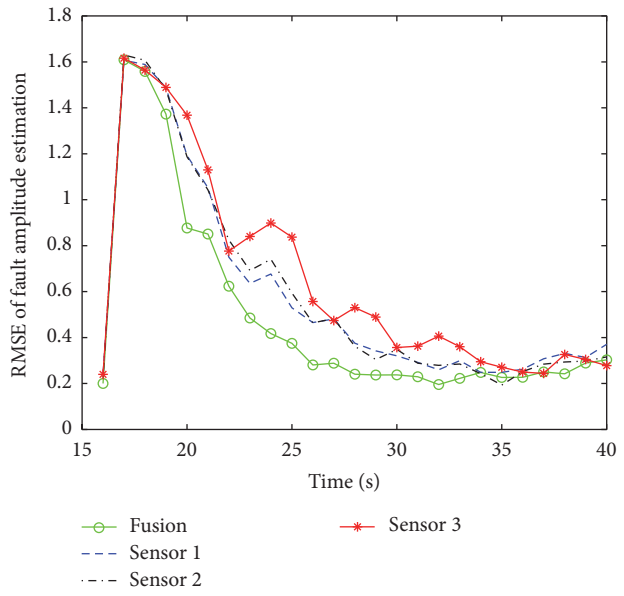


Figure 3: RMSE curves of fault amplitude.

augmented IMM has been performed by augmenting the fault amplitude directly into the state vector, and the model probability generated by the augmented IMM has been used to detect and isolate the fault, which is superior to other model-based fault detection methods in that it has a clear detection threshold, while the fault estimation has been achieved based on STF which has good tracking performance for the time-varying fault, even when abrupt changes exist. There are no constraints on the number of the sensors or the initial sample times and sampling rates of multiple sensors. The simulation results demonstrate that the proposed algorithm can detect fault quickly and estimate it accurately. Further works can focus on the fault diagnosis problems on networked systems and systems with state constraints.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61773055, in part by the Fundamental Research Funds for the Central Universities of China with Grant FRF-TP-16-029A3, and in part by the Beijing Key Laboratory of Knowledge Engineering for Materials Science under Grant FRF-BD-16-010A.

References

[1] D. Zhao, D. Shen, and Y. Q. Wang, "Fault diagnosis and compensation for two-dimensional discrete time systems with sensor faults and time-varying delays," *International Journal of Robust and Nonlinear Control*, 2017.

[2] N. Pous, D. Gingras, and D. Gruyer, "Intelligent Vehicle Embedded Sensors Fault Detection and Isolation Using Analytical Redundancy and Nonlinear Transformations," *Journal of Control Science and Engineering*, vol. 2017, Article ID 1763934, 2017.

[3] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010.

[4] Y. Wang, D. Zhao, Y. Li, and S. X. Ding, "Unbiased Minimum Variance Fault and State Estimation for Linear Discrete Time-Varying Two-Dimensional Systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 5463–5469, 2017.

[5] H. A. P. Blom and Y. Bar-Shalom, "Interacting multiple model algorithm for systems with Markovian switching coefficients.," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.

[6] S. A. Gadsden, K. McCullough, and S. R. Habibi, "Fault detection and diagnosis of an electrohydrostatic actuator using a novel interacting multiple model approach," in *Proceedings of the 2011 American Control Conference, ACC*, pp. 1396–1401, USA, July 2011.

[7] Y. Zhang and X. R. Li, "Detection and diagnosis of sensor and actuator failures using IMM estimator," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1293–1313, 1998.

[8] J. Ru and X. R. Li, "Interacting Multiple Model Algorithm with Maximum Likelihood Estimation for FDI," in *Proceedings of the IEEE International Symposium on Intelligent Control*, pp. 661–666, USA, October 2003.

[9] S. Zhao, B. Huang, and F. Liu, "Fault Detection and Diagnosis of Multiple-Model Systems with Mismodeled Transition Probabilities," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 5063–5071, 2015.

[10] S. Kim, J. Choi, and Y. Kim, "Fault detection and diagnosis of aircraft actuators using fuzzy-tuning IMM filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 940–952, 2008.

[11] B. Pourbabaee, N. Meskin, and K. Khorasani, "Sensor Fault Detection, Isolation, and Identification Using Multiple-Model-Based Hybrid Kalman Filter for Gas Turbine Engines," in *IEEE Transactions on Control Systems Technology*, vol. 24, pp. 1184–1200, 1184-1200, 24(4, 2016.

[12] L. Cork and R. Walker, "Sensor fault detection for UAVs using a nonlinear dynamic model and the IMM-UKF algorithm," in *Proceedings of the 2007 Information, Decision and Control, IDC*, pp. 230–235, Australia, February 2007.

[13] K. K. N. Tudoroiu, "Fault detection and diagnosis for satellite's attitude control system (ACS) using an interactive multiple model (IMM) approach," in *Proceedings of the 2005 IEEE Conference on Control Applications, 2005. CCA 2005.*, pp. 1287–1292, Toronto, Canada.

[14] I. Rapoport and Y. Oshman, "Fault-tolerant particle filtering by using interacting multiple model-based rao-blackwellization," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 6, pp. 1171–1177, 2005.

[15] L. Hou and N. W. Bergmann, "Novel industrial wireless sensor networks for machine condition monitoring and fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 10, pp. 2787–2798, 2012.

[16] L. Sela Perelman, W. Abbas, X. Koutsoukos, and S. Amin, "Sensor placement for fault location identification in water networks: A minimum test cover approach," *Automatica*, vol. 72, pp. 166–176, 2016.

[17] Al. Joelle Hage, E. Maan, and D. Pomorski, "Novel Classifier Fusion Approaches for Fault Diagnosis in Automotive Systems," in *IEEE Transactions on Instrumentation and Measurement*, vol. 58, pp. 602–611, 602-611, 58(3, 2009.

[18] A. Yunusa-Kaltungo, J. K. Sinha, and K. Elbhbah, "An improved data fusion technique for faults diagnosis in rotating machines," *Measurement*, vol. 58, pp. 27–32, 2014.

[19] W. Liu, J. Wei, M. Liang, Y. Cao, and I. Hwang, "Multi-sensor fusion and fault detection using hybrid estimation for air traffic surveillance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 4, pp. 2323–2339, 2013.

[20] T. Damarla, L. M. Kaplan, and G. T. Whipps, "Sniper localization using acoustic asynchronous sensors," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1469–1478, 2010.

[21] I. Jang, D. Pyeon, H. Yoon, and D. Kim, "Channel-quality-aware multihop broadcast for asynchronous multi-channel wireless sensor networks," *Wireless Networks*, vol. 22, no. 7, pp. 2143–2158, 2016.

[22] Y. Hu and D. Zhou, "Actuator fault diagnosis for dynamic systems with multiple asynchronous sensors using the IMM approach," in *Proceedings of the 3rd International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2011*, pp. 318–321, China, January 2011.

[23] D.-J. Jwo and S.-H. Wang, "Adaptive fuzzy strong tracking extended Kalman filtering for GPS navigation," *IEEE Sensors Journal*, vol. 7, no. 5, pp. 778–789, 2007.

[24] M. Narasimhappa, S. L. Sabat, and J. Nayak, "Adaptive sampling strong tracking scaled unscented Kalman filter for denoising the fibre optic gyroscope drift signal," *IET Science, Measurement & Technology*, vol. 9, no. 3, pp. 241–249, 2015.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 4375690, 11 pages
https://doi.org/10.1155/2017/4375690

Research Article

Degradation Data-Driven Remaining Useful Life Estimation in the Absence of Prior Degradation Knowledge

Yong Yu,[1,2] Changhua Hu,[1] Xiaosheng Si,[1] and Jianxun Zhang[1]

[1]*Department of Automation, Xi'an Institute of High-Technology, Xi'an, Shaanxi 710025, China*
[2]*Institute No. 25, The Second Academy of China Aerospace Science and Industry Corporation, Beijing 100854, China*

Correspondence should be addressed to Changhua Hu; hch_reu@sina.com

Received 6 August 2017; Accepted 23 October 2017; Published 3 December 2017

Academic Editor: Chunhui Zhao

Recent developments in prognostic and health management have been targeted at utilizing the observed degradation signals to estimate residual life distributions. Current degradation models mainly focus on a population of "identical" devices or an individual device with population information, not a single component in the absence of prior degradation knowledge. However, the fast development of science and technology provides us with many kinds of new systems, and we just have the real-time monitoring information to analyze the reliability for them. The fusion algorithm presented herein addresses this challenge by combining the excellent modeling ability of Bayesian updating method for the multilevel data and the prominent estimation ability of ECM algorithm for incomplete data. Residual life distributions and posterior distributions are first calculated through the Bayesian updating method based on random initial a priori distributions. Then the a priori distributions are revised and improved for future predictions by the ECM algorithm. Once a new signal is observed, we can reuse the fusion algorithm to improve the accuracy of residual life distributions. The applicability of this fusion algorithm is validated by a set of simulation experiments.

1. Introduction

Modern engineering systems are overwhelmingly complex because of increasing requirements on their functionalities and qualities. These systems often have a high standard of system reliability because a single failure can lead to catastrophic consequences with profound impacts, extreme costs, and potential safety hazards. It will take an exceedingly long time for a system to fail, so prognostics for systems have become extremely difficult, even if the actual operating conditions are severe and rigorous. Therefore, effective methods that can predict failure progression and evaluate the reliability of the system have long been sought. When there are sufficient monitoring data and efficient computational capability, prognostics for components based on observed degradation data is promising and effective [1–3].

The general path model is a typical method utilizing the observed degradation signals of these degraded systems. Lu and Meeker [4] introduced the model to the degradation literature in 1993, for the first time. In their model, the fixed-effects parameters affect the populations' characteristics and the random-effects parameters describe an individual unit's characteristics. Once the parameters are known, the residual life is deterministic. Therefore, the core work is to estimate the unknown fixed and random parameters. Lu and Meeker used a two-stage method to estimate the unknown parameters. Lu et al. [5] extended the degradation model and suggested likelihood-based estimation methods. However, these are not suitable for all types of degradation data. Su et al. [6] considered random sample sizes and random repeated measurement times for each product unit. They discussed the advantages and disadvantages of two-stage least-squares (LS) estimation, maximum modified likelihood (MML) estimation, and maximum likelihood estimation (MLE). They showed that the LS estimators are not consistent in the case of random sample sizes. However, MLE can provide a consistent estimator and has smaller biases and variances compared to the LS and MML estimates. In the further study, Weaver et al. [7] also used MLE to estimate the unknown parameters. They extended the research and examined effects of sample size on the estimation precision. Under the mixed-effect path model, Wu and Shao [8] built the asymptotic properties of the

(weighted) least-squares estimators. They used these properties to calculate approximate confidence intervals and point estimates for percentiles of the failure-time distribution. They used the weighted least-squares estimators to predict the resistor of metal film and the metal fatigue crack length.

However, the above papers only focus on the estimation of the unknown parameters about population devices and need a fair amount of samples in the test. In order to solve this problem, Robinson and Crowder [9] described a fully Bayesian approach which allows a small sample size. They used a variety of simple prior distributions and observed that this aspect has little effect on the posterior distributions of these data, showing that the information in the degradation data dominates. So the Bayesian approach is more suitable for the parameters estimation of an individual device with population information compared with two-stage LS estimation and MLE. Gebraeel et al. [10] developed two different exponential degradation signal models. One model assumes that the error fluctuations follow an iid random error process; however, the second model considers that the error terms follow a Brownian motion process. In their paper, they used the Bayesian updating methods to combine the distribution information of the parameters across the population and the monitored degradation data from the individual device. The Bayesian updating methods can update the stochastic parameters of degradation models, every time a new degradation signal comes. Gebraeel [11] extended the Bayesian updating procedure by assuming that the stochastic parameters in the exponential degradation models follow a bivariate normal distribution. Chakraborty et al. [12] further extended the updating procedure and investigated the difference of the life time distributions when the stochastic parameters do not follow the normal distribution. They also built methods for calculating Remaining Useful Life when the stochastic parameters of the exponential model follow more general distributions. Chen and Tsui [13] adopted a piecewise log-linear degradation model and assumed the change time of the two different phases is random. This assumption explicitly accounts for the characteristics of different degradation phases. When new observations were available, they updated the a posteriori information of the model stochastic parameters including regression coefficients and the variance of the error term by using Bayesian methods. They also suggested a new method which took the correlations into consideration, among degradation predictions, to compute the RUL distribution with better accuracy. Their approach can be naturally extended to more general degradation models.

The above Bayesian approaches can be applied to predicting the RUL of an individual device with population information. However, the fast development of science and technology provides us with kinds of newly made systems, and we just have the observed monitoring degradation information to analyze the reliability of them. Traditional ways to predict system failures often fail because the domain knowledge and expert experience are limited and historical data is nonexistent.

Therefore, effective methods that can predict failures of these newly made systems in the absence of prior degradation knowledge have long been sought. Considering the excellent modeling ability of Bayesian updating method for the multilevel data and the prominent estimation ability of ECM algorithm for incomplete data, the goal of our paper is to develop a collaboration method between Bayesian updating method and ECM algorithm to estimate the Remaining Useful Life (RUL) of the newly made system just with real-time sensing data. To verify the applicability of this fusion algorithm, a set of simulation experiments are conducted.

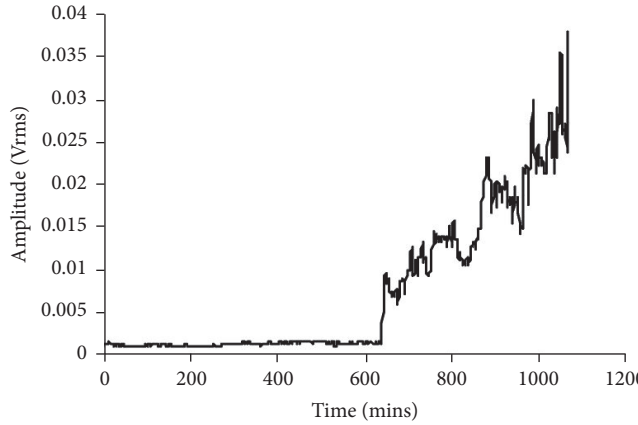


Figure 1: Vibration-based degradation signal.

The remainder of this paper is organized as follows. Section 2 develops a separate Bayesian updating method and calculates the RUL distributions for the exponential degradation model with a Brownian error term, under the assumption that the prior distributions are known. Section 3 explains the procedure of the fusion process in detail and estimates the prior information for the Bayesian updating method by our fusion algorithm. Section 4 illustrates the validity of the collaboration method by a set of simulation experiments. The paper concludes with some discussions and guidance for the estimation of RUL of the single component without prior degradation knowledge in Section 5.

2. Bayesian Updating Method and Residual Life Distribution

In order to develop our collaboration method, we first introduce the Bayesian updating method and its estimation result of the Remaining Useful Life in this section. We will adopt the exponential degradation signal model developed by Gebraeel et al. [10, 11], Kaiser and Gebraeel [14], and Elwany and Gebraeel [15] and assume that the error fluctuation of the degradation model follows a Brownian motion process. As our objective is to compute the RUL of a single system without prior distribution knowledge, we believe that the adopted exponential degradation model and the error fluctuation are adequate for the given degradation path in Figure 1. For a further discussion of model selection and evaluation, see Li et al. [16].

Under the above assumption, we could use the Bayesian updating procedure to compute the unknown random parameters of the exponential degradation model. Once

we have got the calculated posterior distributions of these random parameters in the exponential degradation model, we can derive the residual life distribution of the component. However, in our paper we only have the real-time monitoring degradation signal for the newly made component, and prior distributions are nonexistent. In order to utilize the Bayesian updating procedure, we first assume we have got the accurate and informative a priori information. The method of estimating these prior distributions will be illustrated and detailed in Section 3.

2.1. The Degradation Signal Model. First, we review the general definition of the Brownian motion process.

Definition 1. A standard Brownian motion process, $W(t)$, $t \geq 0$, possesses the following properties:

1. If $t_0 < t_1 < \cdots < t_n$, then $W(t_0), W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$ are mutually independent.
2. If $s, t \geq 0$, then

$$P\left(W\left(s+t\right) - W\left(s\right) \in A\right) = \int_A \left(2\pi t\right)^{-1/2} e^{-x^2/2t} dx. \quad (1)$$

3. With probability one $t \to W(t)$ is continuous.

The first part of the definition describes the fact that the process $W(t)$ has independent increments. The second part means that the increment $W(s+t) - W(s)$ follows a normal distribution with mean zero and variance t. The third part describes the fact that $W(t)$, $t \geq 0$, almost certainly has continuous paths. For an in-depth discussion about the Brownian motion process and its properties, see Durrett [17].

Then, we let $X(t)$ denote the real-time monitoring signal as a continuous stochastic process, with respect to time t. We define the functional form of $X(t)$ as

$$\begin{aligned} X\left(t\right) &= \varphi + \theta \exp\left(\beta t + \varepsilon\left(t\right) - \frac{\sigma^2 t}{2}\right) \\ &= \varphi + \theta \exp\left(\beta t\right) \exp\left(\varepsilon\left(t\right) - \frac{\sigma^2 t}{2}\right), \end{aligned} \quad (2)$$

where φ is the fixed intercept and represents the initial degradation, and β is a normal random variable such that the mean of β is μ_1 and the variance of β is σ_1^2. θ is a lognormal random variable with mean μ_0 and variance σ_0^2, and $\varepsilon(t) = \sigma W(t)$ is a Brownian motion with mean zero and variance $\sigma^2 t$. Under the assumption that θ, β, and $\varepsilon(t)$ are mutually independent, it is obvious to obtain that $E[\exp(\varepsilon(t) - (\sigma^2 t/2))] = 1$, and thus $E(S(t) \mid \theta, \beta) = \varphi + \theta \exp(\beta t)$.

Furthermore, we find that it is easy to calculate with the logarithmic degradation data. Thus, we define $S(t)$ as follows:

$$S\left(t\right) = \ln\left(X\left(t\right) - \varphi\right) = \ln\theta + \beta t + \varepsilon\left(t\right) - \frac{\sigma^2 t}{2}. \quad (3)$$

By defining $\theta' = \ln\theta$, $\beta' = \beta - \sigma^2/2$, we can further simplify $S(t)$ as follows:

$$S\left(t\right) = \theta' + \beta' t + \varepsilon\left(t\right). \quad (4)$$

2.2. Bayesian Updating Process of Stochastic Parameters. We let $S_k = S(t_k) - S(t_{k-1})$, $k = 2, 3, \ldots$ denote the difference value of the monitored logged degradation signal at times t_k and t_{k-1}, with $S_1 = S(t_1)$.

Then, suppose we have obtained the logged difference value $S_{1:k} = \{S_1, S_2, \ldots, S_k\}$ at times $t_1, \ldots, t_k$. And the error increments, $\varepsilon(t_i) - \varepsilon(t_{i-1})$, $i = 2, \ldots, k$, are independent normal random variables. If the stochastic parameters, θ' and β', are given, we can define the conditional joint density function of $S_{1:k}$ as

$$\begin{aligned} p\left(S_{1:k} \mid \theta', \beta'\right) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^k \\ &\cdot \exp\left(\begin{array}{c} -\dfrac{\left(S_1 - \theta' - \beta' t_1\right)^2}{2\sigma^2 t_1} \\ -\displaystyle\sum_{i=2}^{k}\left(\dfrac{\left(S_i - \beta'\left(t_i - t_{i-1}\right)\right)^2}{2\sigma^2\left(t_i - t_{i-1}\right)}\right) \end{array}\right). \end{aligned} \quad (5)$$

Generally, however, θ' and β' will be unknown. Based on the former assumption, we suppose we have got accurate and informative priors. And we let $\pi_0(\theta')$ and $\pi_1(\beta')$ denote the prior distributions on θ' and β', respectively, where $\pi_0(\theta') = N(\mu_0, \sigma_0^2)$ and $\pi_1(\beta') = N(\mu_1', \sigma_1^2)$, $\mu_1' = \mu_1 - \sigma^2/2$. Then, given the logged difference data, $S_{1:k}$, obtained at times $t_1, \ldots, t_k$, the posterior joint distribution of (θ', β') can be expressed as follows.

Theorem 2. *Given the obtained logged degradation signal, $S_{1:k}$, the conditional posterior joint distribution of (θ', β') is $\theta', \beta' \mid S_{1:k} \sim N(\mu_{\theta',k}, \sigma_{\theta',k}^2, \mu_{\beta',k}, \sigma_{\beta',k}^2, \rho_k)$, where*

$$\begin{aligned} &\mu_{\theta',k} \\ &= \frac{\left(S_1\sigma_0^2 + \mu_0\sigma^2 t_1\right)\left(\sigma_1^2 t_k + \sigma^2\right) - \sigma_0^2 t_1\left(\sigma_1^2\sum_{i=1}^{k} S_i + \mu_1'\sigma^2\right)}{\left(\sigma_0^2 + \sigma^2 t_1\right)\left(\sigma_1^2 t_k + \sigma^2\right) - \sigma_0^2\sigma_1^2 t_1}, \\ &\mu_{\beta',k} \\ &= \frac{\left(\sigma_1^2\sum_{i=1}^{k} S_i + \mu_1'\sigma^2\right)\left(\sigma^2 t_1 + \sigma_0^2\right) - \sigma_1^2\left(S_1\sigma_0^2 + \mu_0\sigma^2 t_1\right)}{\left(\sigma_0^2 + \sigma^2 t_1\right)\left(\sigma_1^2 t_k + \sigma^2\right) - \sigma_0^2\sigma_1^2 t_1}, \\ &\sigma_{\theta',k}^2 = \frac{\sigma^2\sigma_0^2 t_1\left(\sigma_1^2 t_k + \sigma^2\right)}{\left(\sigma_0^2 + \sigma^2 t_1\right)\left(\sigma_1^2 t_k + \sigma^2\right) - \sigma_0^2\sigma_1^2 t_1}, \\ &\sigma_{\beta',k}^2 = \frac{\sigma^2\sigma_1^2\left(\sigma_0^2 + \sigma^2 t_1\right)}{\left(\sigma_0^2 + \sigma^2 t_1\right)\left(\sigma_1^2 t_k + \sigma^2\right) - \sigma_0^2\sigma_1^2 t_1}, \\ &\rho_k = \frac{-\sigma_0\sigma_1\sqrt{t_1}}{\sqrt{\left(\sigma_0^2 + \sigma^2 t_1\right)\left(\sigma_1^2 t_k + \sigma^2\right)}}. \end{aligned} \quad (6)$$

Proof. Given the prior distributions of θ' and β', $\pi_0(\theta') = N(\mu_0, \sigma_0^2)$ and $\pi_1(\beta') = N(\mu_1', \sigma_1^2)$, we can express the conditional posterior joint distribution of (θ', β'), $p(\theta', \beta' \mid S_{1:k})$, as follows:

$$p\left(\theta', \beta' \mid S_{1:k}\right) \propto p\left(S_{1:k} \mid \theta', \beta'\right) \pi_0\left(\theta'\right) \pi_1\left(\beta'\right)$$

$$\propto \frac{1}{2\pi\sigma_{\theta',k}\sigma_{\beta',k}\sqrt{1-\rho_k^2}} \cdot \exp\left[-\frac{1}{2\left(1-\rho_k^2\right)}\right. \tag{7}$$

$$\left.\cdot \begin{pmatrix} \dfrac{\left(\theta' - \mu_{\theta',k}\right)^2}{\sigma_{\theta',k}^2} \\ -2\rho_k \dfrac{\left(\theta' - \mu_{\theta',k}\right)\left(\beta' - \mu_{\beta',k}\right)}{\sigma_{\theta',k}\sigma_{\beta',k}} \\ + \dfrac{\left(\beta' - \mu_{\beta',k}\right)^2}{\sigma_{\beta',k}^2} \end{pmatrix}\right].$$

As our focus is on the fusion algorithm, for a further detailed calculation of the posterior distribution of (θ', β'), see Gebraeel et al. [10]. □

2.3. Estimation of the Residual Life Distribution. Every time a new degradation signal comes, we can compute a new posterior distribution of (θ', β'). As the objective of our paper is to estimate the distribution of the RUL of the monitored system, we suppose that the system's failure occurs when the observed degradation signal reaches the failure threshold, w, and thus we need to estimate the time until the degradation signal reaches w. In our paper, we assume that the threshold value is a constant value.

The objective of prognostics is to compute the distribution of the failure time until the degradation signal reaches the threshold w. To achieve this goal, we first calculate the posterior distribution of (θ', β'). Then, we let the random variable $S(t + t_k)$ denote the logged degradation signal value obtained at time $t + t_k$, $t > 0$, given $S_{1:k}$ obtained at times $t_1, \ldots, t_k$. Under the above assumption, the distribution of $S(t + t_k)$ given $S_{1:k}$ can be expressed as follows.

Theorem 3. *Given the observed difference value of the logged degradation data, $S_{1:k}$, the distribution of $S(t + t_k)$ is $S(t + t_k) \mid S_{1:k} \sim N(\widetilde{\mu}(t + t_k), \widetilde{\sigma}^2(t + t_k))$, where*

$$\widetilde{\mu}\left(t + t_k\right) \triangleq \sum_{i=1}^{k} S_i + \mu_{\beta'} t = S\left(t_k\right) + \mu_{\beta',k} t,$$
$$\widetilde{\sigma}^2\left(t + t_k\right) \triangleq \sigma_{\beta',k}^2 t^2 + \sigma^2 t. \tag{8}$$

Proof. First, note that $S(t) = \theta' + \beta' t + \varepsilon(t)$ and we can express $S(t + t_k) = S(t_k) + \beta' t + \varepsilon(t + t_k) - \varepsilon(t_k)$, where $S(t_k) = \sum_{i=1}^{k} S_i$. Therefore, given $S_{1:k}$, $S(t + t_k)$ follows a normal distribution with mean $\widetilde{\mu}(t + t_k) = S(t_k) + E[\beta']t = S(t_k) + \mu_{\beta',k} t$ and variance $\widetilde{\sigma}^2(t + t_k) = t^2 V[\beta'] + V[\varepsilon(t + t_k) - \varepsilon(t_k)] = \sigma_{\beta',k}^2 t^2 + \sigma^2 t$. □

Then, we let T denote the RUL of the monitored system, and we know that T meets $S(T + t_k) = \ln w$. Given $S_{1:k}$, the conditional cumulative distribution function (CDF) of T, $F_{T|S_{1:k}}(t) = p\{T \le t \mid S_{1:k}\}$, can be expressed as

$$\begin{aligned} P\left(T \le t \mid S_{1:k}\right) &= 1 - P\left(S\left(t + t_k\right) \le \ln w \mid S_{1:k}\right) \\ &= 1 - P\left(Z < \frac{\ln w - \widetilde{\mu}\left(t + t_k\right)}{\sqrt{\widetilde{\sigma}^2\left(t + t_k\right)}}\right) \\ &= P\left(Z \ge \frac{\ln w - \widetilde{\mu}\left(t + t_k\right)}{\sqrt{\widetilde{\sigma}^2\left(t + t_k\right)}}\right) \\ &= \Phi\left(g\left(t\right)\right), \end{aligned} \tag{9}$$

where Z represents a standard normal variable, $\Phi(\bullet)$ denotes the CDF of the standard normal variable, and $g(t) = (\widetilde{\mu}(t + t_k) - \ln w)/\sqrt{\widetilde{\sigma}^2(t + t_k)}$.

We compute the residual life distribution of the system at time t_k, under the condition that the observed degradation signal does not reach the threshold w; that is, $\sum_{i=1}^{k} S_i = S(t_k) < \ln w$. Thus, we get

$$\lim_{t \to 0} g\left(t\right) = \lim_{t \to 0} \frac{\sum_{i=1}^{k} S_i + \mu_{\beta',k} t - \ln w}{\sqrt{\sigma_{\beta',k}^2 t^2 + \sigma^2 t}} = -\infty. \tag{10}$$

Therefore, we get $\lim_{t \to 0} F_{T|S_{1:k}}(t) = 0$, which means that the domain of the RUL, T, is $(0, \infty)$. We can express the conditional probability distribution function (PDF) of T, given $S_{1:k}$, as

$$f_{T|S_{1:k}, T \ge 0}\left(t\right) = \phi\left(g\left(t\right)\right) g'\left(t\right), \tag{11}$$

where $\phi(\cdot)$ denotes the PDF of the standard normal variable.

Given the conditional PDF of T, we can write the expectation of RUL, at time t_k, as

$$E_{T|S_{1:k}, T \ge 0}\left(t\right) = \int_0^{\infty} f_{T|S_{1:k}, T \ge 0}\left(t\right) dt. \tag{12}$$

In order to simplify the integral of expectation of RUL, we use the failure equation $S(T + t_k) = \ln w$ to compute the RUL, at time t_k, approximately. And according to Theorem 3, given the difference value of logged degradation signal, $S_{1:k}$, the distribution of $S(t + t_k)$ follows $S(t + t_k) \mid S_{1:k} \sim N(\widetilde{\mu}(t + t_k), \widetilde{\sigma}^2(t + t_k))$, and the mean of this distribution, $\widetilde{\mu}(t + t_k)$, closely approximates $S(t + t_k)$. Thus, we can write $\widetilde{\mu}(T + t_k) = \ln w$ and express the RUL, at time t_k, as

$$\mathrm{RUL}_k = \frac{\ln w - S\left(t_k\right)}{\mu_{\beta',k}}, \tag{13}$$

where RUL_k is closely approximated to the expectation of RUL.

In other literature, the prior distribution parameters, σ^2, μ_0, σ_0^2, μ_1', σ_1^2, are computed from historical monitoring data or derived from domain knowledge and expert experience. However, in this section we only have the real-time condition degradation information and do not know the accurate prior distribution parameters. What is more, the priors are fixed in the whole Bayesian updating procedure in the previous articles, and if the prior distribution parameters are inaccurate, the posterior distribution for these unknown parameters would have great errors. So our paper will solve these problems in the next section.

3. Estimating the Prior Information by Fusion Algorithm

In Section 2, we have estimated the RUL of the single component by Bayesian updating procedure under the assumption that a priori distributions were known. However, we only have real-time observations and the prior distributions needed in Bayesian updating process are nonexistent. So our paper will develop a collaboration algorithm between Bayesian updating and ECM algorithm to estimate these a priori parameters.

3.1. The ECM Algorithm. We let $\boldsymbol{\Theta} = [\sigma^2, \mu_0, \sigma_0^2, \mu_1', \sigma_1^2]$ denote the unknown prior distribution parameters. Given the difference value of logged degradation signal $S_{1:k}$, we can express the log-likelihood function of $S_{1:k}$ as

$$l_k(\boldsymbol{\Theta}) = \ln\left[p\left(S_{1:k} \mid \boldsymbol{\Theta}\right)\right] = \sum_{j=2}^{k} \ln\left[p\left(S_j \mid S_{1:j-1}, \boldsymbol{\Theta}\right)\right], \quad (14)$$

where $p(S_{1:k} \mid \boldsymbol{\Theta})$ is the joint PDF of $S_{1:k}$. Thus, we can write the MLE of $\boldsymbol{\Theta}$ at time t_k as

$$\widehat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} l_k(\boldsymbol{\Theta}), \quad (15)$$

where $\widehat{\boldsymbol{\Theta}}$ is the variable value of $\boldsymbol{\Theta}$ corresponding to maximum of $l_k(\boldsymbol{\Theta})$. Our goal is to estimate an appropriate prior distribution parameter $\widehat{\boldsymbol{\Theta}}$. However, in formula (4) θ' and β' are stochastic and unobserved, so the calculation of formula (15) is hard to complete.

In order to avoid the above problem, we propose a collaboration algorithm between Bayesian updating and ECM algorithm and use the fusion algorithm to estimate prior distributions based on degradation signals. We let $\boldsymbol{\Theta}_k = [\sigma_k^2, \mu_{0,k}, \mu_{1,k}', \sigma_{0,k}^2, \sigma_{1,k}^2]$ denote the estimated parameters based on degradation signals, $S_{1:k}$, and $\widehat{\boldsymbol{\Theta}}_k^{(i)} = [\widehat{\sigma}_k^{2(i)}, \widehat{\mu}_{0,k}^{(i)}, \widehat{\mu}_{1,k}'^{(i)}, \widehat{\sigma}_{0,k}^{2(i)}, \widehat{\sigma}_{1,k}^{2(i)}]$ denote the estimated variable value of $\boldsymbol{\Theta}_k$ of the ith iteration result in the ECM algorithm. We treat θ' and β' as the hidden variables, because θ' and β' are stochastic and unobserved in the whole degradation process. Our objective is to calculate the variable value of $\boldsymbol{\Theta}_k$, so the E-step of ECM algorithm can be written as

$$l\left(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i)}\right) = E_{\theta',\beta' \mid S_{1:k}, \widehat{\boldsymbol{\Theta}}_k^{(i)}}\left\{\ln p\left(S_{1:k}, \theta', \beta' \mid \boldsymbol{\Theta}_k\right)\right\}, \quad (16)$$

where $E_{\theta',\beta' \mid S_{1:k}, \widehat{\boldsymbol{\Theta}}_k^{(i)}}$ is the conditional expectation of hidden variables, θ' and β', given $\widehat{\boldsymbol{\Theta}}_k^{i}$ and $S_{1:k}$.

Next we express the CM-step of ECM algorithm as

$$\begin{aligned}
\widehat{\boldsymbol{\Theta}}_k^{(i+1/S)} &= \arg\max_{\boldsymbol{\Theta}} l\left(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i)}\right. \\
&\quad \left. = \left[\widehat{\sigma}_k^{2(i)}, \widehat{\mu}_{0,k}^{(i)}, \widehat{\mu}_{1,k}'^{(i)}, \widehat{\sigma}_{0,k}^{2(i)}, \widehat{\sigma}_{1,k}^{2(i)}\right]\right), \\
\widehat{\boldsymbol{\Theta}}_k^{(i+2/S)} &= \arg\max_{\boldsymbol{\Theta}} l\left(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i+1/S)}\right. \\
&\quad \left. = \left[\widehat{\sigma}_k^{2(i+1)}, \widehat{\mu}_{0,k}^{(i)}, \widehat{\mu}_{1,k}'^{(i)}, \widehat{\sigma}_{0,k}^{2(i)}, \widehat{\sigma}_{1,k}^{2(i)}\right]\right), \\
\widehat{\boldsymbol{\Theta}}_k^{(i+3/S)} &= \arg\max_{\boldsymbol{\Theta}} l\left(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i+2/S)}\right. \\
&\quad \left. = \left[\widehat{\sigma}_k^{2(i+1)}, \widehat{\mu}_{0,k}^{(i+1)}, \widehat{\mu}_{1,k}'^{(i)}, \widehat{\sigma}_{0,k}^{2(i)}, \widehat{\sigma}_{1,k}^{2(i)}\right]\right), \\
\widehat{\boldsymbol{\Theta}}_k^{(i+4/S)} &= \arg\max_{\boldsymbol{\Theta}} l\left(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i+3/S)}\right. \\
&\quad \left. = \left[\widehat{\sigma}_k^{2(i+1)}, \widehat{\mu}_{0,k}^{(i+1)}, \widehat{\mu}_{1,k}'^{(i+1)}, \widehat{\sigma}_{0,k}^{2(i)}, \widehat{\sigma}_{1,k}^{2(i)}\right]\right) \\
&\quad \vdots \\
\widehat{\boldsymbol{\Theta}}_k^{(i+1)} &= \arg\max_{\boldsymbol{\Theta}} l\left(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i+(S-1)/S)}\right. \\
&\quad \left. = \left[\widehat{\sigma}_k^{2(i+1)}, \widehat{\mu}_{0,k}^{(i+1)}, \widehat{\mu}_{1,k}'^{(i+1)}, \widehat{\sigma}_{0,k}^{2(i+1)}, \widehat{\sigma}_{1,k}^{2(i)}\right]\right),
\end{aligned} \quad (17)$$

where S ($S = 5$) is the dimension of vector $\boldsymbol{\Theta}$, and $\widehat{\boldsymbol{\Theta}}_k^{(i+1)}$ is the maximum value of $\boldsymbol{\Theta}$ based on $\widehat{\boldsymbol{\Theta}}_k^{(i+(S-1)/S)}$ from the derivative equation $\partial l(\boldsymbol{\Theta}_k \mid \widehat{\boldsymbol{\Theta}}_k^{(i+(S-1)/S)})/\partial\widehat{\sigma}_{1,k}^{2(i)} = 0$. As our focus is on the algorithm fusion process, for the theory of ECM algorithm and its convergence analysis, see Meng and Rubin [18], Van et al. [19], and Liu and Rubin [20].

3.2. The Collaboration between Bayesian Updating and ECM Algorithm. After we have reviewed the ECM algorithm, we begin our fusion process. Figure 2 shows the procedure of the fusion process.

First, we use random initial a priori distributions to start the fusion algorithm, when we collect the degradation signal $S(t_1)$ at time t_1. Then, we get the posterior distribution of (θ', β'), $p(\theta', \beta' \mid S_{1:k})$, and a residual life distribution by Bayesian updating method. Of course, the results are inaccurate.

Then, we use the posterior distribution of (θ', β'), $p(\theta', \beta' \mid S_{1:k})$, to substitute the distribution, given $\widehat{\boldsymbol{\Theta}}_k^{(i)}$, of hidden variables in the E-step. So, the E-step of the ECM algorithm can be rewritten as

$$\begin{aligned}
&l\left(\boldsymbol{\Theta}_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)\right) \\
&\quad = E_{\theta',\beta' \mid p(\theta',\beta' \mid S_{1:k})}\left\{\ln p\left(S_{1:k}, \theta', \beta' \mid \boldsymbol{\Theta}_k\right)\right\}.
\end{aligned} \quad (18)$$

Because of the different E-step, we get a rewritten CM-step which is different from the one in the ECM algorithm. In

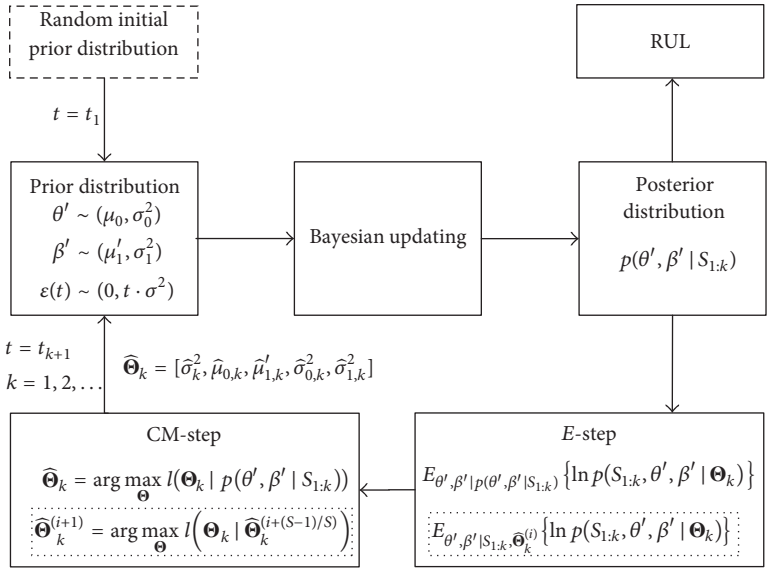


Figure 2: The collaboration algorithm between Bayesian updating and ECM algorithm.

order to get a more accurate estimated value, $\widehat{\Theta}_k$, in ECM algorithm we need multiple iterations. However, we can get the optimal solution of Θ_k only through S ($S = 5$) steps calculation, in our fusion algorithm. This will be proved in Theorem 4. In our fusion algorithm the rewritten CM-step can be expressed as

$$\begin{aligned}
\widehat{\Theta}_k^{(1/5)} &= \arg\max_{\Theta} l\left(\Theta_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)\right. \\
&\quad = \left.\left[\sigma_k^2, \mu_{0,k}, \mu_{1,k}', \sigma_{0,k}^2, \sigma_{1,k}^2\right]\right), \\
\widehat{\Theta}_k^{(2/5)} &= \arg\max_{\Theta} l\left(\Theta_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)^{(1/5)}\right. \\
&\quad = \left.\left[\widehat{\sigma}_k^2, \mu_{0,k}, \mu_{1,k}', \sigma_{0,k}^2, \sigma_{1,k}^2\right]\right), \\
\widehat{\Theta}_k^{(3/5)} &= \arg\max_{\Theta} l\left(\Theta_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)^{(2/5)}\right. \\
&\quad = \left.\left[\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \mu_{1,k}', \sigma_{0,k}^2, \sigma_{1,k}^2\right]\right), \\
\widehat{\Theta}_k^{(4/5)} &= \arg\max_{\Theta} l\left(\Theta_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)^{(3/5)}\right. \\
&\quad = \left.\left[\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}_{1,k}', \sigma_{0,k}^2, \sigma_{1,k}^2\right]\right), \\
\widehat{\Theta}_k &= \arg\max_{\Theta} l\left(\Theta_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)^{(4/5)}\right. \\
&\quad = \left.\left[\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}_{1,k}', \widehat{\sigma}_{0,k}^2, \sigma_{1,k}^2\right]\right).
\end{aligned} \tag{19}$$

In this five-step calculation of formula (19), we can get the optimal estimations $\sigma_k^2 = \widehat{\sigma}_k^2$, $\mu_{0,k} = \widehat{\mu}_{0,k}$, $\mu_{1,k}' = \widehat{\mu}_{1,k}'$, $\sigma_{0,k}^2 = \widehat{\sigma}_{0,k}^2$, $\sigma_{1,k}^2 = \widehat{\sigma}_{1,k}^2$, respectively. The results of the rewritten CM-step are the a priori distributions of the next Bayesian updating procedure. Each time we collect a new degradation signal $S(t_k)$ at time t_k, we can recalculate the a priori distributions and the residual life distribution.

3.3. Estimation of Prior Information. After we have finished the fusion process, we begin to calculate the optimal variable value of Θ_k by our fusion algorithm. The log-likelihood function of complete data can be expressed as

$$\begin{aligned}
&\ln p\left(S_{1:k}, \theta', \beta' \mid \Theta_k\right) \\
&\quad = \ln p\left(S_{1:k} \mid \theta', \beta', \Theta_k\right) + \ln p\left(\theta', \beta' \mid \Theta_k\right) \\
&\quad = -\frac{k+2}{2} \ln 2\pi - \frac{k}{2} \ln \sigma_k^2 - \frac{\left(S_1 - \theta' - \beta' t_1\right)^2}{2\sigma_k^2 t_1} \\
&\qquad - \sum_{i=2}^{k} \frac{\left(S_i - \beta'\left(t_i - t_{i-1}\right)\right)^2}{2\sigma_k^2\left(t_i - t_{i-1}\right)} - \frac{1}{2} \ln \sigma_{0,k}^2 - \frac{1}{2} \ln \sigma_{1,k}^2 \\
&\qquad - \frac{\left(\theta' - \mu_{0,k}\right)^2}{2\sigma_{0,k}^2} - \frac{\left(\beta' - \mu_{1,k}'\right)^2}{2\sigma_{1,k}^2}.
\end{aligned} \tag{20}$$

According to formulas (18) and (19), we can compute the rewritten E-step as follows:

$$\begin{aligned}
l\left(\Theta_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)\right) &= E_{\theta', \beta' \mid p(\theta', \beta' \mid S_{1:k})} \left\{\ln p\left(S_{1:k}, \theta', \beta' \mid \Theta_k\right)\right\} \\
&= -\frac{k+2}{2} \ln 2\pi - \frac{k}{2} \ln \sigma_k^2
\end{aligned}$$

$$
\begin{aligned}
&- \frac{S_1^2 - 2S_1\left(\mu_{\theta',k} + \mu_{\beta',k}t_1\right) + \mu_{\theta',k}^2 + \sigma_{\theta',k}^2 + 2t_1\left(\rho_k\sigma_{\theta',k}\sigma_{\beta',k} + \mu_{\theta',k}\mu_{\beta',k}\right) + t_1^2\left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)}{2\sigma_k^2 t_1} \\
&- \sum_{i=2}^{k} \frac{S_i^2 - 2S_i\mu_{\beta',k}\left(t_i - t_{i-1}\right) + \left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)\left(t_i - t_{i-1}\right)^2}{2\sigma_k^2\left(t_i - t_{i-1}\right)} - \frac{1}{2}\ln\sigma_{0,k}^2 - \frac{1}{2}\ln\sigma_{1,k}^2 \\
&- \frac{\mu_{\theta',k}^2 + \sigma_{\theta',k}^2 - 2\mu_{\theta',k}\mu_{0,k} + \mu_{0,k}^2}{2\sigma_{0,k}^2} - \frac{\mu_{\beta',k}^2 + \sigma_{\beta',k}^2 - 2\mu_{\beta',k}\mu_{1,k}' + \mu_{1,k}'^2}{2\sigma_{1,k}^2}.
\end{aligned} \tag{21}
$$

Then, based on the rewritten CM-step, we can get the optimal $\widehat{\sigma}_k^2$, $\widehat{\mu}_{0,k}$, $\widehat{\mu}_{1,k}'$, $\widehat{\sigma}_{0,k}^2$, $\widehat{\sigma}_{1,k}^2$ by the following steps, respectively.

Step 1. By calculating $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k}))/\partial\sigma_k^2 = 0$, we can get the optimal

$$
\widehat{\sigma}_k^2 = \frac{1}{k}\left[\begin{array}{c}
\dfrac{S_1^2 - 2S_1\left(\mu_{\theta',k} + \mu_{\beta',k}t_1\right) + \mu_{\theta',k}^2 + \sigma_{\theta',k}^2 + 2t_1\left(\rho_k\sigma_{\theta',k}\sigma_{\beta',k} + \mu_{\theta',k}\mu_{\beta',k}\right) + t_1^2\left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)}{t_1} \\
+\sum_{i=2}^{k} \dfrac{S_i^2 - 2S_i\mu_{\beta',k}\left(t_i - t_{i-1}\right) + \left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)\left(t_i - t_{i-1}\right)^2}{\left(t_i - t_{i-1}\right)}
\end{array}\right]. \tag{22}
$$

Step 2. By calculating $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k})^{(1/5)})/\partial\mu_{0,k} = 0$, we can get the optimal $\widehat{\mu}_{0,k} = \mu_{\theta',k}$.

Step 3. By calculating $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k})^{(2/5)})/\partial\mu_{1,k}' = 0$, we can get the optimal $\widehat{\mu}_{1,k}' = \mu_{\beta',k}$.

Step 4. By calculating $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k})^{(3/5)})/\partial\sigma_{0,k}^2 = 0$, we can get the optimal $\widehat{\sigma}_{0,k}^2 = \sigma_{\theta',k}^2$.

Step 5. By calculating $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k})^{(4/5)})/\partial\sigma_{1,k}^2 = 0$, we can get the optimal $\widehat{\sigma}_{1,k}^2 = \sigma_{\beta',k}^2$.

According to the above five steps, $\widehat{\boldsymbol{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}_{1,k}', \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$ can be expressed as

$$
\begin{aligned}
\widehat{\sigma}_k^2 &= \frac{1}{k}\left[\begin{array}{c}
\dfrac{S_1^2 - 2S_1\left(\mu_{\theta',k} + \mu_{\beta',k}t_1\right) + \mu_{\theta',k}^2 + \sigma_{\theta',k}^2 + 2t_1\left(\rho_k\sigma_{\theta',k}\sigma_{\beta',k} + \mu_{\theta',k}\mu_{\beta',k}\right) + t_1^2\left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)}{t_1} \\
+\sum_{i=2}^{k} \dfrac{S_i^2 - 2S_i\mu_{\beta',k}\left(t_i - t_{i-1}\right) + \left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)\left(t_i - t_{i-1}\right)^2}{\left(t_i - t_{i-1}\right)}
\end{array}\right], \\
\widehat{\mu}_{0,k} &= \mu_{\theta',k}, \\
\widehat{\sigma}_{0,k}^2 &= \sigma_{\theta',k}^2, \\
\widehat{\mu}_{1,k}' &= \mu_{\beta',k}, \\
\widehat{\sigma}_{1,k}^2 &= \sigma_{\beta',k}^2.
\end{aligned} \tag{23}
$$

We get the optimal $\widehat{\sigma}_k^2$, $\widehat{\mu}_{0,k}$, $\widehat{\mu}_{1,k}'$, $\widehat{\sigma}_{0,k}^2$, $\widehat{\sigma}_{1,k}^2$ in each step, respectively. However, it does not mean that $\widehat{\boldsymbol{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}_{1,k}', \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$ is the only maximum point of $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k}))/\partial\boldsymbol{\Theta}_k = 0$. So, we need a further discussion.

Theorem 4. *According to the result of formula (23), the estimated $\widehat{\boldsymbol{\Theta}}_k$ is the only maximum point of $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k}))/\partial \boldsymbol{\Theta}_k = 0$.*

Proof. First, from formula (21) we can get $\widehat{\boldsymbol{\Theta}}_k$ which is the only solution of $\partial l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k}))/\partial \boldsymbol{\Theta}_k = 0$. Next, we prove that $\widehat{\boldsymbol{\Theta}}_k$ is corresponding to maximum $l(\boldsymbol{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k}))$. We can perform the second derivative of log-likelihood function on vector $\boldsymbol{\Theta}_k$ as

$$\frac{\partial^2 l\left(\boldsymbol{\Theta}_k \mid p\left(\theta', \beta' \mid S_{1:k}\right)\right)}{\partial \boldsymbol{\Theta}_k \partial \boldsymbol{\Theta}_k^T} = \begin{bmatrix} \frac{k}{2\sigma_k^4} - \frac{\tau}{\sigma_k^6} & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{\sigma_{0,k}^2} & 0 & \frac{\mu_{0,k} - \mu_{\theta',k}}{\sigma_{0,k}^4} & 0 \\ 0 & 0 & -\frac{1}{\sigma_{1,k}^2} & 0 & \frac{\mu'_{1,k} - \mu_{\beta',k}}{\sigma_{1,k}^4} \\ 0 & \frac{\mu_{0,k} - \mu_{\theta',k}}{\sigma_{0,k}^4} & 0 & \frac{1}{2\sigma_{0,k}^4} - \frac{\psi_1}{\sigma_{0,k}^6} & 0 \\ 0 & 0 & \frac{\mu'_{1,k} - \mu_{\beta',k}}{\sigma_{1,k}^4} & 0 & \frac{1}{2\sigma_{1,k}^4} - \frac{\psi_2}{\sigma_{1,k}^6} \end{bmatrix}, \tag{24}$$

where

$$\begin{aligned} \tau &= \frac{S_1^2 - 2S_1\left(\mu_{\theta',k} + \mu_{\beta',k} t_1\right) + \mu_{\theta',k}^2 + \sigma_{\theta',k}^2 + 2t_1\left(\rho_k \sigma_{\theta',k}\sigma_{\beta',k} + \mu_{\theta',k}\mu_{\beta',k}\right) + t_1^2\left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)}{t_1} \\ &\quad + \sum_{i=2}^{k} \frac{S_i^2 - 2S_i\mu_{\beta',k}\left(t_i - t_{i-1}\right) + \left(\mu_{\beta',k}^2 + \sigma_{\beta',k}^2\right)\left(t_i - t_{i-1}\right)^2}{\left(t_i - t_{i-1}\right)}, \\ \psi_1 &= \mu_{\theta',k}^2 + \sigma_{\theta',k}^2 - 2\mu_{\theta',k}\mu_{0,k} + \mu_{0,k}^2, \\ \psi_2 &= \mu_{\beta',k}^2 + \sigma_{\beta',k}^2 - 2\mu_{\beta',k}\mu'_{1,k} + \mu'^2_{1,k}. \end{aligned} \tag{25}$$

Then we use the estimated variable value $\widehat{\boldsymbol{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}'_{1,k}, \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$ in (23), and we can find the order principal minor of the matrix in (24) as follows:

$$\begin{aligned} &\Delta_1\big|_{\boldsymbol{\Theta}_k = \widehat{\boldsymbol{\Theta}}_k} < 0, \\ &\Delta_2\big|_{\boldsymbol{\Theta}_k = \widehat{\boldsymbol{\Theta}}_k} > 0, \\ &\Delta_3\big|_{\boldsymbol{\Theta}_k = \widehat{\boldsymbol{\Theta}}_k} < 0, \\ &\Delta_4\big|_{\boldsymbol{\Theta}_k = \widehat{\boldsymbol{\Theta}}_k} > 0, \\ &\Delta_5\big|_{\boldsymbol{\Theta}_k = \widehat{\boldsymbol{\Theta}}_k} < 0. \end{aligned} \tag{26}$$

Thus, we can conclude that (24) is a negative-definite matrix when $\boldsymbol{\Theta}_k = \widehat{\boldsymbol{\Theta}}_k$, which means that $\widehat{\boldsymbol{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}'_{1,k}, \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$ is the only maximum point. □

This fusion procedure between Bayesian updating and ECM algorithm can be performed each time a new degradation signal is observed. That is to say, each time the degradation signal $S(t_k)$ is observed, we can recalculate $\widehat{\boldsymbol{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}'_{1,k}, \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$ and obtain new estimates of residual life for the newly made system. What is more, the initial values of priors in the Bayesian updating for the first time are unrestricted, and once we get a new $\widehat{\boldsymbol{\Theta}}_k$, it would be used in the next Bayesian updating as the priors. In Section 4, we will use simulation method to further evaluate the performance of our collaboration algorithm and illustrate that the unrestricted initial values of priors have little influence on the estimated vector $\widehat{\boldsymbol{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}'_{1,k}, \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$.

4. Simulation and Analysis

In this section, we will adopt the simulation method to further evaluate the performance of our collaboration algorithm. First, in order to represent a degradation process, we create a set of simulation data based on the exponential degradation model in Section 2. In the simulation, we assume that $\varphi = 0$, $\theta' \sim N(0.02, 2 \times 10^{-6})$, $\beta' \sim N(0.01, 1 \times 10^{-6})$, $\varepsilon(t) \sim N(0, t \cdot 4 \times 10^{-4})$. In order to observe enough degradation signals and

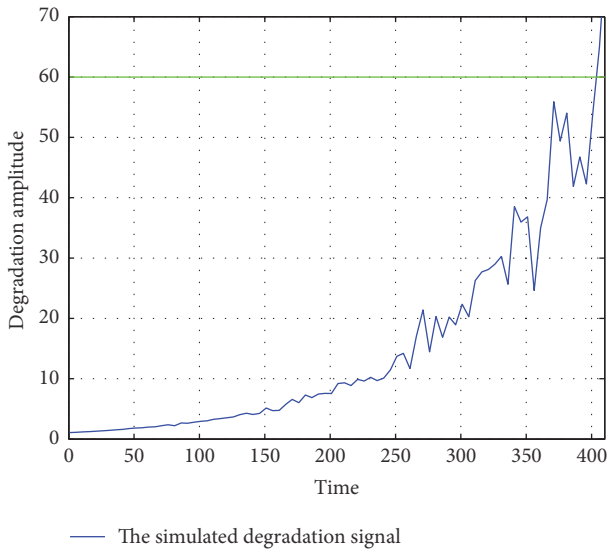


Figure 3: Simulation of degradation path.

Table 1: Different initial prior distributions.

Data	θ'	β'	$\varepsilon(t)$
(1)	$N(0.1, 2\times10^{-4})$	$N(0.05, 1\times10^{-4})$	$N(0, t\cdot 4\times10^{-3})$
(2)	$N(0.2, 2\times10^{-4})$	$N(0.1, 1\times10^{-4})$	$N(0, t\cdot 4\times10^{-3})$
(3)	$N(0.3, 2\times10^{-4})$	$N(0.15, 1\times10^{-4})$	$N(0, t\cdot 4\times10^{-3})$

obvious degradation trend, we let the threshold $w = 60$ and the sampling interval $(t_i - t_{i-1}) = 10$, $i = 2, 3, 4, \ldots$. Figure 3 shows the trajectory of the simulated degradation signal. We obtain 41 degradation samples, and the degradation reaches the standard threshold, $w = 60$, at time $t = 403.5$.

We know that $\theta' \sim N(\mu_0, \sigma_0^2)$ and $\beta' \sim N(\mu_1 - \sigma^2/2, \sigma_1^2)$, and μ_0 and $\mu_1 - \sigma^2/2$ dominate the degradation and RUL. We will use the estimated results of them to prove that these unrestricted initial priors have little influence on the estimated accuracy of the RUL. The different initial prior distributions are as shown in Table 1.

On the other hand, in order to prove that our collaboration algorithm can get a more accurate RUL than other separate Bayesian updating methods, we will compare our fusion algorithm with the method in Gebraeel et al. [10]. by using the 2nd set of initial priors in Table 1. Figure 4 represents the estimated results of μ_0. Figure 5 represents the estimated results of $\mu_1 - \sigma^2/2$.

From Figures 4 and 5, we know that the collaboration algorithm can estimate the mean of θ' and β' accurately even with inaccurate prior information. Although we have used a variety of inaccurate prior distributions, it can still be observed that this aspect has little effect on the estimation of the means of θ' and β', showing that the degradation measurements dominate and the collaboration algorithm is effective. We can know that the estimated means of θ' and β' gradually approximate to the simulated value, 0.02 and 0.01, at time about 50, in spite of different initial priors.

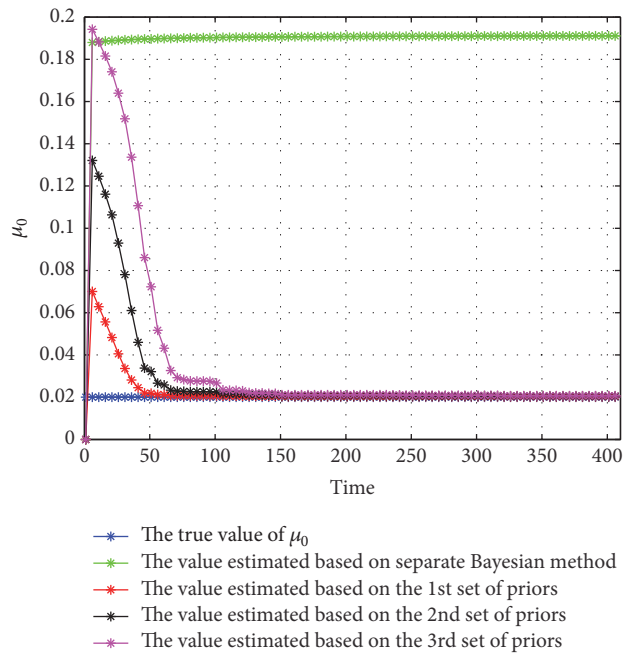


Figure 4: Estimated results of μ_0.

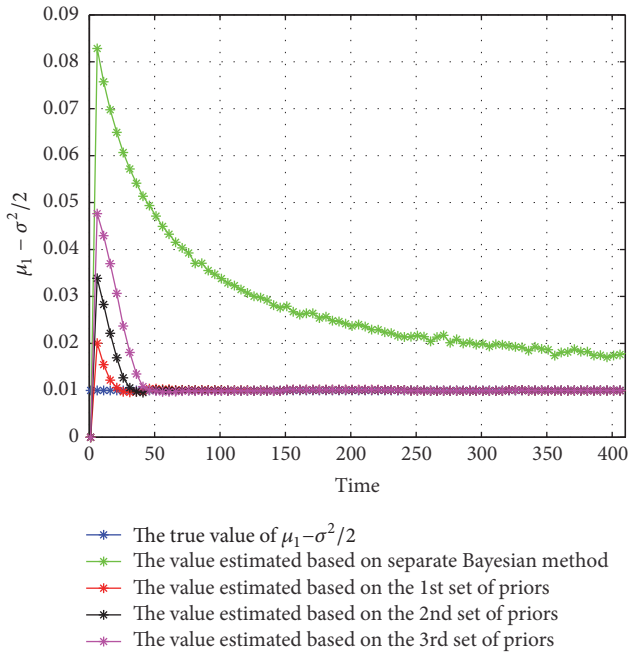


Figure 5: Estimated results of $\mu_1 - \sigma^2/2$.

What is more, our collaboration algorithm can get more accurate estimations compared with the separate Bayesian updating by using the 2nd set of initial priors in Table 1.

Given the estimated prior distributions we calculate the point estimations of RUL of the newly made component. Figure 6 shows the point estimations of RUL by our algorithm and separate Bayesian updating method.

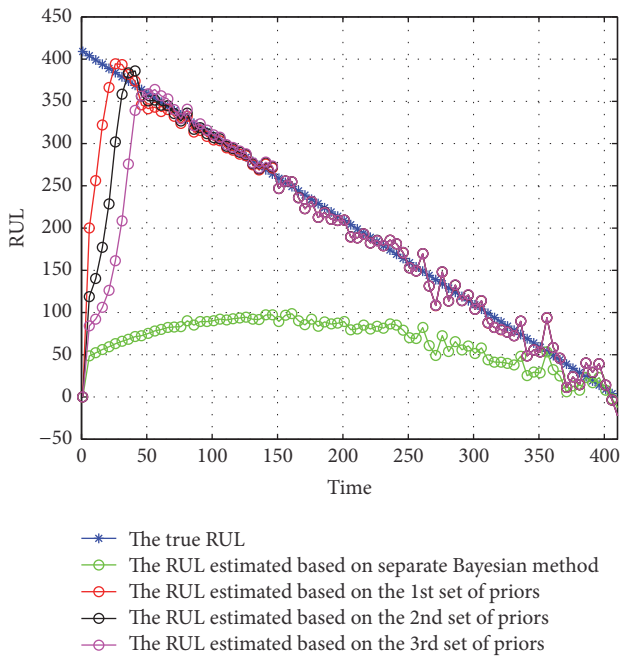


FIGURE 6: Point estimations of RUL by collaboration algorithm and separate Bayesian Updating method.

From Figure 6, we can know that our collaboration algorithm can get a more accurate RUL comparing with the separate Bayesian method. And inaccurate a priori distributions have little effect on the estimation of the RUL. Furthermore, the point estimations of RUL by our collaboration algorithm can also reflect the fluctuation of degradation caused by Brownian motion error, and this can be known from Figure 3.

5. Conclusion

In this paper, we presented a collaboration algorithm that contains the characteristic of Bayesian updating and ECM algorithm. The difficulties of the fusion process mainly consist of two parts; the first is building the connection between ECM algorithm and Bayesian updating. It is not easy to find another substitute for the ith iteration distribution in the rewritten E-step, and the posterior distribution of the Bayesian updating is an optimal one. The second is proving the optimal estimating result about the prior information. We can get a maximum estimating result, $\widehat{\mathbf{\Theta}}_k = [\widehat{\sigma}_k^2, \widehat{\mu}_{0,k}, \widehat{\mu}'_{1,k}, \widehat{\sigma}_{0,k}^2, \widehat{\sigma}_{1,k}^2]$, by the rewritten CM-step. However, it does not mean that the estimating result is the only maximum point of $\partial l(\mathbf{\Theta}_k \mid p(\theta', \beta' \mid S_{1:k})/\partial \mathbf{\Theta}_k = 0$. In this paper, we use the second derivative of log-likelihood function on vector $\mathbf{\Theta}_k$ and the order principal minor of the matrix in (24) to prove it.

Our fusion algorithm can predict failures of newly made systems in the absence of prior degradation knowledge. Although our fusion algorithm started with random initial a priori distributions, the simulation experiments show that the inaccurate a priori distributions have little effect on the estimation of the RUL, and our fusion algorithm can get a better prediction than the separate Bayesian method. Nevertheless, there are still some issues that needed further investigation for the estimation of the RUL on the single component without prior degradation knowledge.

First, we assume that the exponential degradation signal model and the error fluctuation are adequate for the given degradation path. Actually, we can not guarantee which model is the best one by visual judgment. So we also need to focus on model selection and evaluate the goodness-of-fit of various degradation path models.

Second, we assume that the stochastic parameters of the exponential degradation model are normally distributed. However, the stochastic parameters may follow a Gamma distribution or other distributions. Therefore, we also should investigate the performance of our collaboration algorithm when the underlying normal distribution assumptions are not satisfied.

Third, our work started the RUL calculation in the degradation processes assuming that the exact point-in-time of the initial degradation is known. However, the functioning system would be stable within a period of time, and the point-in-time for the initial degradation is unknown and stochastic. By considering the distribution of the initial degradation point, we may be able to predict the RUL right after its installment.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study is financially supported by the National Natural Science Foundation of China (61573365, 61374120, and 61473094) and the National Science Fund for Distinguished Youth Scholars of China (61025014).

References

[1] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Applied Stochastic Models in Business and Industry*, vol. 31, no. 1, pp. 16–32, 2015.

[2] M. A. Freitas, M. L. G. De Toledo, E. A. Colosimo, and M. C. Pires, "Using degradation data to assess reliability: a case study on train wheel degradation," *Quality and Reliability Engineering International*, vol. 25, no. 5, pp. 607–629, 2009.

[3] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—a review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, no. 1, pp. 1–14, 2011.

[4] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, 1993.

[5] J.-C. Lu, J. Park, and Q. Yang, "Statistical inference of a time-to-failure distribution derived from linear degradation data," *Technometrics*, vol. 39, no. 4, pp. 391–400, 1997.

[6] C. Su, J.-C. Lu, D. Chen, and J. M. Hughes-Oliver, "A random coefficient degradation model with ramdom sample size," *Lifetime Data Analysis*, vol. 5, no. 2, pp. 173–183, 1999.

[7] B. P. Weaver, W. Q. Meeker, L. A. Escobar, and J. Wendelberger, "Methods for planning repeated measures degradation studies," *Technometrics*, vol. 55, no. 2, pp. 122–134, 2013.

[8] S.-J. Wu and J. Shao, "Reliability analysis using the least squares method in nonlinear mixed-effect degradation models," *Statistica Sinica*, vol. 9, no. 3, pp. 855–877, 1999.

[9] M. E. Robinson and M. J. Crowder, "Bayesian methods for a growth-curve degradation model with repeated measures," *Lifetime Data Analysis*, vol. 6, no. 4, pp. 357–374, 2000.

[10] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: a Bayesian approach," *IIE Transactions*, vol. 37, no. 6, pp. 543–557, 2005.

[11] N. Gebraeel, "Sensory-updated residual life distributions for components with exponential degradation patterns," *IEEE Transactions on Automation Science and Engineering*, vol. 3, no. 4, pp. 382–393, 2006.

[12] S. Chakraborty, N. Gebraeel, M. Lawley, and H. Wan, "Residual-life estimation for components with non-symmetric priors," *IIE Transactions*, vol. 41, no. 4, pp. 372–387, 2009.

[13] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: revisited," *IIE Transactions*, vol. 45, no. 9, pp. 939–952, 2013.

[14] K. A. Kaiser and N. Z. Gebraeel, "Predictive maintenance management using sensor-based degradation models," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 39, no. 4, pp. 840–849, 2009.

[15] A. Elwany and N. Gebraeel, "Real-time estimation of mean remaining life using sensor-based degradation models," *Journal of Manufacturing Science and Engineering*, vol. 131, no. 5, pp. 0510051–0510059, 2009.

[16] Z. Li, Y. Deng, and C. Mastrangelo, "Model selection for degradation-based Bayesian reliability analysis," *Journal of Manufacturing Systems*, vol. 37, pp. 72–82, 2015.

[17] R. Durrett, *Probability: Theory and Examples*, Cambridge University Press, Cambridge, UK, 2010.

[18] X.-L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[19] D. A. Van, X.-L. Meng, and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: Computing the asymptotic variance," *Statistica Sinica*, vol. 5, no. 1, pp. 55–75, 1995.

[20] C. Liu and D. B. Rubin, "The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence," *Biometrika*, vol. 81, no. 4, pp. 633–648, 1994.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 3169172, 8 pages
https://doi.org/10.1155/2017/3169172

Research Article

Tooth Fracture Detection in Spiral Bevel Gears System by Harmonic Response Based on Finite Element Method

Yuan Chen,[1,2] Rupeng Zhu,[1] Guanghu Jin,[1] and Yeping Xiong[2]

[1] *College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2] *Engineering and Environment, University of Southampton, Boldrewood Innovation Campus, Southampton SO16 7QF, UK*

Correspondence should be addressed to Rupeng Zhu; rpzhu@nuaa.edu.cn

Received 11 August 2017; Revised 11 October 2017; Accepted 19 October 2017; Published 3 December 2017

Academic Editor: Zhixing Cao

Spiral bevel gears occupy several advantages such as high contact ratio, strong carrying capacity, and smooth operation, which become one of the most widely used components in high-speed stage of the aeronautical transmission system. Its dynamic characteristics are addressed by many scholars. However, spiral bevel gears, especially tooth fracture occurrence and monitoring, are not to be investigated, according to the limited published issues. Therefore, this paper establishes a three-dimensional model and finite element model of the Gleason spiral bevel gear pair. The model considers the effect of tooth root fracture on the system due to fatigue. Finite element method is used to compute the mesh generation, set the boundary condition, and carry out the dynamic load. The harmonic response spectra of the base under tooth fracture are calculated and the influence of main parameters on monitoring failure is investigated as well. The results show that the change of torque affects insignificantly the determination of whether or not the system has tooth fracture. The intermediate frequency interval (200 Hz–1000 Hz) is the best interval to judge tooth fracture occurrence. The best fault test region is located in the working area where the system is going through meshing. The simulation calculation provides a theoretical reference for spiral bevel gear system test and fault diagnosis.

1. Introduction

Spiral bevel gears are one of the most important components of the aeronautical transmission system. Due to the harsh working environment, failure occurs sometimes, and the most common situation is resonance failure. The failure of the spiral bevel gear transmission system often leads to serious accidents; therefore, accurate detection, the positioning of the fault, and eliminating hidden danger have very important significance in improving the operating efficiency of the gear system.

Most studies were related to spiral bevel gear modeling and tooth contact analysis (TCA); Tsai and Chin [1] developed mathematical modeling of the tooth surface geometry for bevel gear pairs based on the basic gearing kinematics and involute geometry along with the tangent planes geometry. Litvin et al. [2–6] proposed an integrated computerized approach for spiral bevel gear drives and simulated mesh and contact stress analysis; they reduced the magnitude of transmission errors for reduction of noise and vibration and found severe contact stresses areas to increase the endurance of the gear drives; the proposed results were proved by the manufacturing and test of prototypes. Sheveleva et al. [7, 8] developed a modified computer program for tooth contact analysis (TCA) for the most general case and provided computer codes for contact path, bearing contact, and contact pressure distribution. Simon [9] presented computer aided tooth contact analysis in mismatched spiral bevel gears and discussed the influence of relative position errors of meshing pinion on tooth contact.

About spiral bevel gear fault detection, Zakrajsek et al. [10] applied a variety of gear fault detection techniques to investigate the growth and propagation of the fault; the relationship between the system running time and pinion damage was specified as well. Dempsey et al. [11] developed the diagnostic tool by collecting vibration and oil debris data from fatigue tests performed in Glenn spiral bevel gear fatigue rigs. Decker and Lewicki [12] discussed the

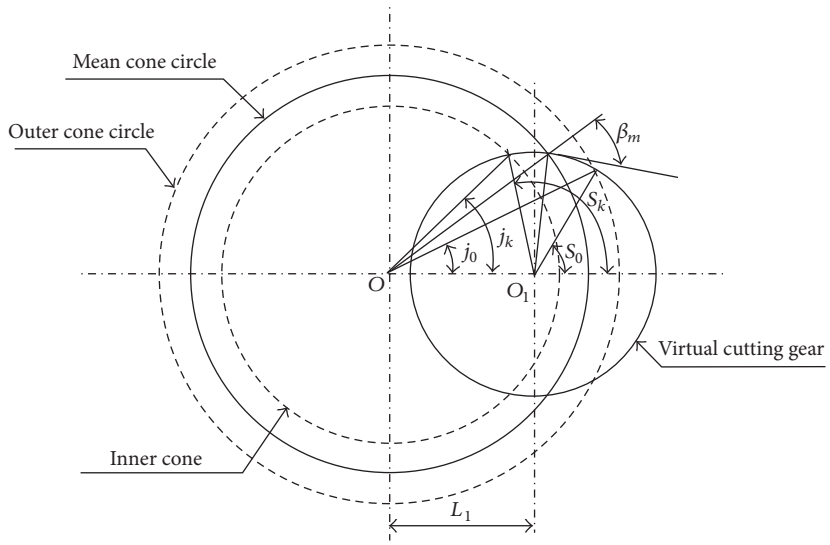


Figure 1: Pitch surface unfolded map.

damage detection effectiveness of the different metrics and a comparison of effects of different accelerometer locations by spiral bevel pinion running test. Ural et al. [13] predicted the crack shape and fatigue life for a spiral bevel pinion gear by using computational fracture mechanics, which was based on linear elastic fracture mechanics theories combined with finite element method. However, the detection in the frequency domain and the influence of excitation were not provided in these papers.

In the area of spiral gear dynamics analysis, Li and Hu [14] analyzed the axial–lateral–torsional coupled spiral bevel geared system theoretically and the dynamic behavior of the system was investigated by numerical method. Yinong et al. [15] studied the effect of the asymmetric mesh stiffness on the 8-DOF spiral bevel gear transmission system. Feng and Song [16] investigated the effects of the dynamic meshing force and dynamic transmission error.

Therefore, through system vibration signal to determine the occurrence of tooth fracture, thus eliminating the fault, is a meaningful work. In this paper, a pair of Gleason spiral bevel gears is mathematically modeled and a three-dimensional solid model is generated as well. Finite element method is applied to analyze the harmonic response of the system; deformation amplitude and phase of the base are calculated. Therefore some tooth fracture features and main parameters' influences are analyzed to assist the fault recognition.

2. System Modeling

2.1. Three-Dimensional Precise Solid Modeling. Typical spiral bevel gears applied in aeronautical transmission system are manufactured by Gleason face hobbing process [17]. The tooth profile of the spiral bevel gear is a spherical involute, and its tooth profile is circular, and the tooth surface is a complex three-dimensional surface. Therefore, the method of modeling the cylindrical gear tooth profile is not suitable for the spiral bevel gear. By adding the auxiliary involute to the tooth line equation, the accuracy of the tooth surface could be increased.

In the spherical coordinate system, the spherical involute equation on one side of outer space width is

$$\begin{aligned} r &= R \\ \theta &= \delta_f + t\left(\delta_a - \delta_f\right) \\ \varphi &= \frac{\arccos\left(\cos\theta/\cos\delta_b\right)}{\sin\delta_b} - \arccos\left(\frac{\tan\delta_b}{\tan\theta}\right), \end{aligned} \tag{1}$$

where δ_b is base angle; δ_f is root angle; δ_a is tip angle; R is sphere radius, that is, outer cone distance.

The spherical involute equation on the other side of outer space width is

$$\begin{aligned} r &= R \\ \theta &= \delta_f + t\left(\delta_a - \delta_f\right); \\ \varphi &= \frac{\arccos\left(\cos\theta/\cos\delta_b\right)}{\sin\delta_b} - \arccos\left(\frac{\tan\delta_b}{\tan\theta}\right) \\ &\quad - \left(\frac{360}{z} - \phi_b\right) \end{aligned} \tag{2}$$

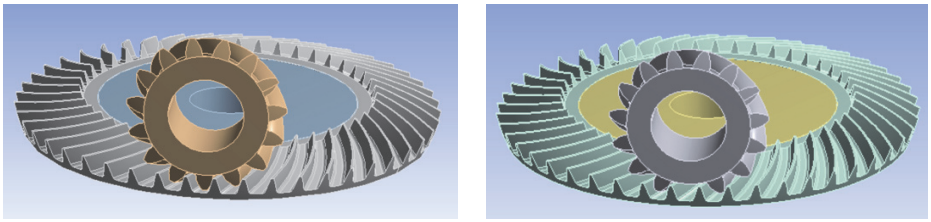
here, z is tooth number; φ_b is tooth thickness base angle.

When spiral bevel gears are in cutting processing, tooth pitch surface unfolded drawing is shown in Figure 1.

The distance between cutting tool circle O_1 and cone circle O is depicted as

$$L_1 = \sqrt{{R_m}^2 + {R_d}^2 - 2R_m R_d \sin\beta_m}, \tag{3}$$

where β_m is mean spiral angle; R_m is mean cone distance; R_d is the radius of cutting tool.



(a) Healthy model (b) Tooth fracture model

Figure 2: Spiral bevel gear assembly model.

In circle O_1, offset angle S_0 of outer cone circle and S_k of inner cone circle are

$$S_0 = 180 - \arccos\frac{L_1 + {R_d}^2 - R^2}{2L_1R_d}$$
$$S_k = 180 - \arccos\frac{L_1 + {R_d}^2 - (R-B)^2}{2L_1R_d}; \quad (4)$$

here, B is tooth thickness.

In circle O, offset angle j_0 of outer cone circle and j_k of inner cone circle are

$$j_0 = \arctan\frac{R_d \sin S_0}{L_1 + R_d \cos S_0}$$
$$j_k = \arctan\frac{R_d \sin S_k}{L_1 + R_d \cos S_k}. \quad (5)$$

According to the spherical geometrical relationship, spherical angle q corresponding to plane angle j could be

$$q_0 = \frac{j_0}{\sin\delta}$$
$$q_k = \frac{j_k}{\sin\delta}; \quad (6)$$

here, δ is reference cone angle.

Thus, $q_{0k} = q_k - q_0$, that is, the rotation angle between outer cone and inner cone in the spherical coordinate system.

The spherical involute equation on one side of inner space width could be derived by substituting r to (R-B) and adding the value of φ to q_{0k} in (1). Similarly, involute equation on the other side of inner space width could be deduced as well.

In order to improve the accuracy of the three-dimensional model, in this paper, a number of equally spaced auxiliary spherical involute lines are inserted in the direction of the tooth line. Substitute (R-B) to (R-0.1nB) in (1) and (4), and n is followed by $1, 2, 3 \ldots 9 \ldots$, then involute equation could be rederived.

Based on the above-mentioned involute equation, key points of the curve are established by "law command" in CATIA. Then "spline command" is applied to connect these key points for the tooth profile involute curve, and "multisection surface command" is used to generate the tooth surface. Moreover, tooth space is modeled by "split command," which is based on these tooth surfaces. Finally, with the Boolean subtraction calculation of the solids, the parametric modeling of the spiral bevel gear could be successfully achieved, which verifies the correctness of the design method [18].

Table 1: System parameters.

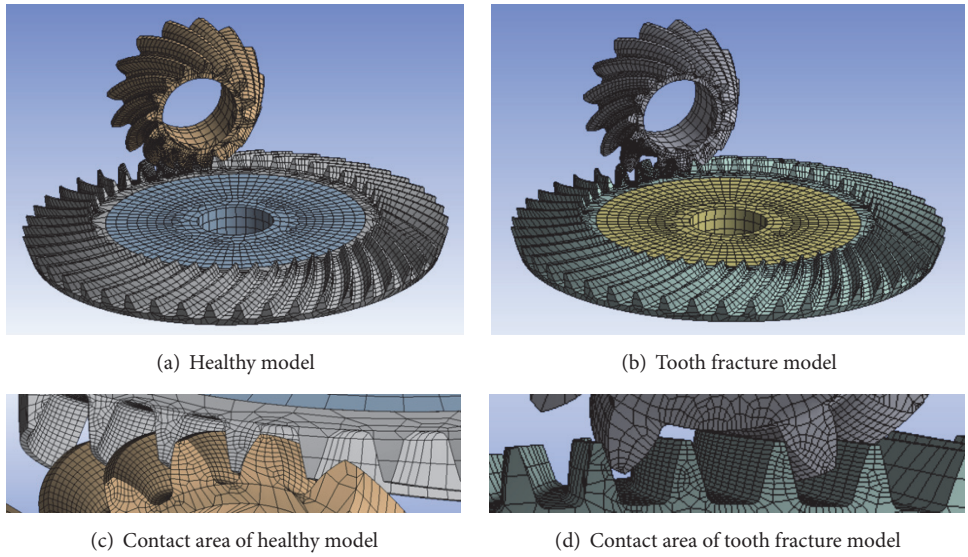
	Active gear	Driven gear
Modulus	6	
Pressure angle (°)	20	
Tooth number	15	46
Tooth width (m)	0.44	0.44
Shaft angle Σ(°)	90	
Mean spiral angle (°)	35	

In addition, due to long-term operation in the high-speed circumstance, tooth root is prone to fatigue fracture. This paper simulates the fractured tooth and completes the healthy and fracture spiral bevel gear assembly model, as shown in Figure 2; main parameters are listed in Table 1.

2.2. Finite Element Meshing and Pretreatment Process. The geometric model in CATIA is introduced into ANSYS workbench 18.1, so finite element analysis model is obtained. As friction contact analysis is a nonlinear problem, in order to save the computing resources and improve calculation accuracy, hexahedral meshing method is applied for calculation, and then the grid is refined in contact areas. The grid size of the noncontact region of this model is defined as 5 mm, and the contact area is defined as 1 mm, which ensures the rationality of the mesh and the maximum efficiency. At the same time, in order to ensure the convergence of the calculation results, the meshing model is firstly calculated, and then the mesh size of the contact area is gradually reduced. If the difference between calculation results is small after multiple attempts, the result is convergent and acceptable. Figure 3 shows the meshing model of spiral bevel gear pair; the results tend to converge when the number of nodes is 151443 and the number of elements is 47982 in healthy model; the number of nodes is 151749 and the number of elements is 47728 in tooth fracture model.

After mesh is complete, the model is processing multiple settings, that is, pretreatment process:

(1) Material definition: the model is set according to steel properties, the active and driven gears belong to



(a) Healthy model (b) Tooth fracture model

(c) Contact area of healthy model (d) Contact area of tooth fracture model

Figure 3: Spiral bevel gear meshing.

the same material, Young's modulus is 2.1×10^{11} Pa, Poisson's ratio is 0.3, and the density is 7850 kg/m^3.

(2) Contact definition: the calculation of the spiral bevel gear pair is surface-to-surface contact problem between the elastomers. It is of great importance to select contact surface and target surface; the inappropriate selection will lead to excessive penetration, affecting the accuracy of the solution. In general, when the convex and concave surfaces contact, the concave surface should be set as the target surface, so the spiral bevel gear outer tooth surface is defined as the target surface. In addition, the amount of penetration between the two contact surfaces depends on the normal contact stiffness. If the normal contact stiffness is too large, it will increase solution iteration number, which may lead to nonconvergence; if the normal contact stiffness is too small, the penetration between nodes could be too large, resulting in model instability. Based on the above analysis, this paper firstly uses the small normal contact stiffness coefficient and then gradually increases until analysis results deviation is so small. The results show that the best normal contact stiffness coefficient is 1.0 when the normal contact stiffness coefficient is set from 0.001, 0.01, 0.1, 1.0, and 1.1.

(3) Rotational pair definition: the inner surfaces of the gears are set as reference surfaces, and then the rotational centers of reference surfaces are created. Finally, a pair of rotational gears is defined.

(4) Solver definition: the general equation for harmonic response analysis is

$$[M]\left\{\ddot{X}\right\} + [C]\left\{\dot{X}\right\} + [K]\{X\} = \{F\}, \quad (7)$$

where $[M]$, $[C]$, and $[K]$ are the mass matrix, damping matrix, and stiffness matrix of the system. Matrix $[F]$ is external excitation and is equal to $F_0 \cos(\omega t)$.

The time-varying meshing stiffness, meshing line displacement, and dynamic meshing force can be regarded as periodic format and can be expanded in Fourier series under fundamental meshing frequency. In this paper, dynamic meshing force is simulated as the excitation load, the torque is loaded in the sinusoidal form, the sweep frequency is set from 0 to 2000 Hz, initial phase angle is 0°, and solution intervals are 100.

(5) Boundary conditions and load setting: degree of freedom (DOF) is released only in rotational direction, and the value of torque is 1200 N·m, which is gradually applied on the active gear.

3. Fault Diagnosis Analysis

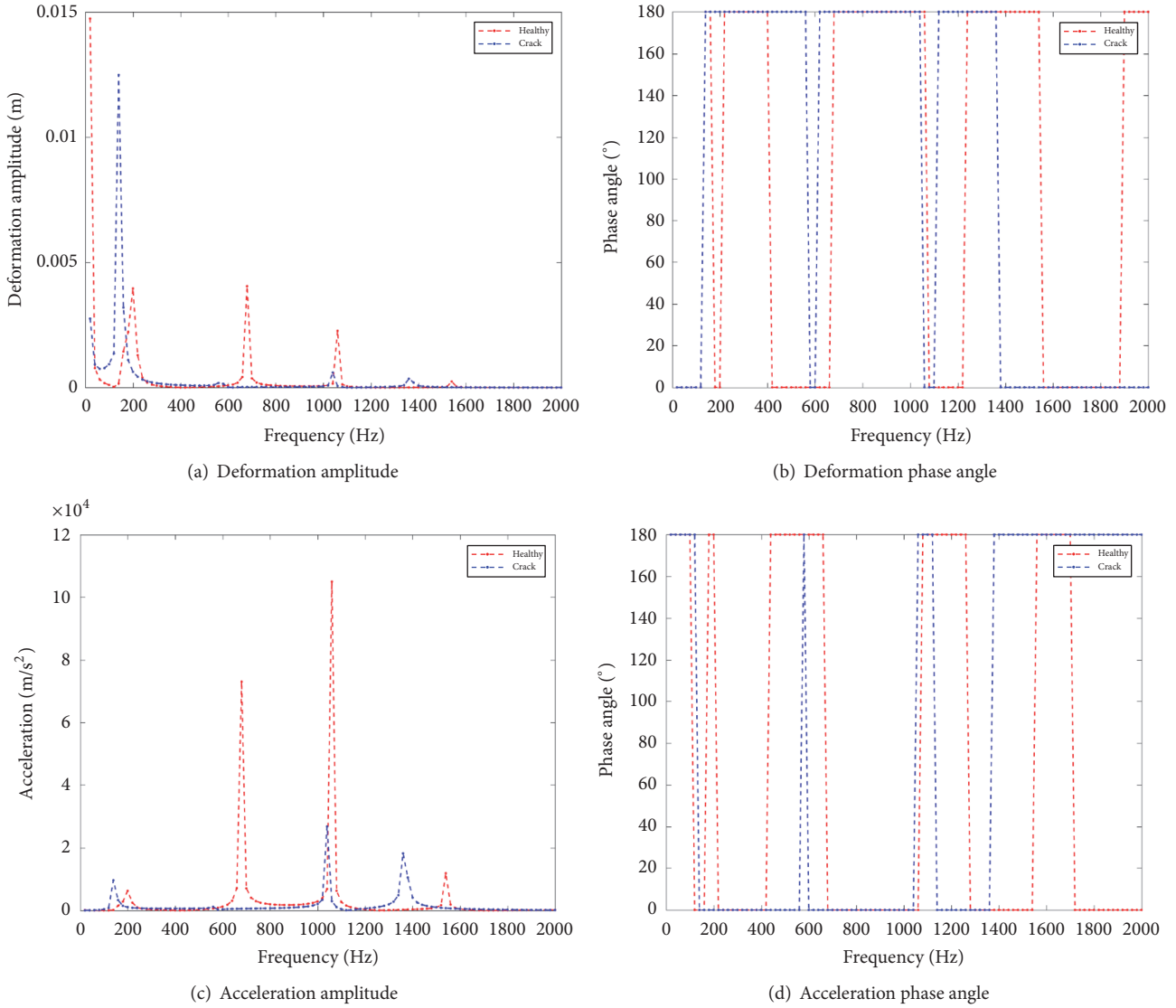
3.1. Tooth Fracture Detection. The system is solved by the augmented Lagrangian method in ANSYS and the first six natural frequencies of the system are obtained, as shown in Table 2. As illustrated in the table, the natural frequency of the tooth fracture model decreases in each order; the frequency deviation of the fourth and sixth order is relatively large.

The base of the driven gear is the optimum position for deformation vibration detection sensor and acceleration vibration detection sensor so that the base is also set in response to the output in finite element method. Figure 4 shows the calculated deformation and acceleration harmonic response maps.

Here, in order to separate ideal monitoring frequency, three frequency intervals are defined to describe the sweep frequency, that is, low frequency interval (below than 200 Hz), intermediate frequency interval (200 Hz to 1000 Hz), and high frequency interval (greater than 1000 Hz) [19].

Table 2: Natural frequency.

Order	Healthy tooth model frequency (Hz)	Fracture tooth model frequency (Hz)	Deviation (Hz)
(1)	18.4	12.0	6.4
(2)	168.0	140.8	27.2
(3)	206.5	147.3	59.2
(4)	678.1	571.6	106.5
(5)	1061.2	1042.4	18.8
(6)	1543	1367.3	175.7



(a) Deformation amplitude

(b) Deformation phase angle

(c) Acceleration amplitude

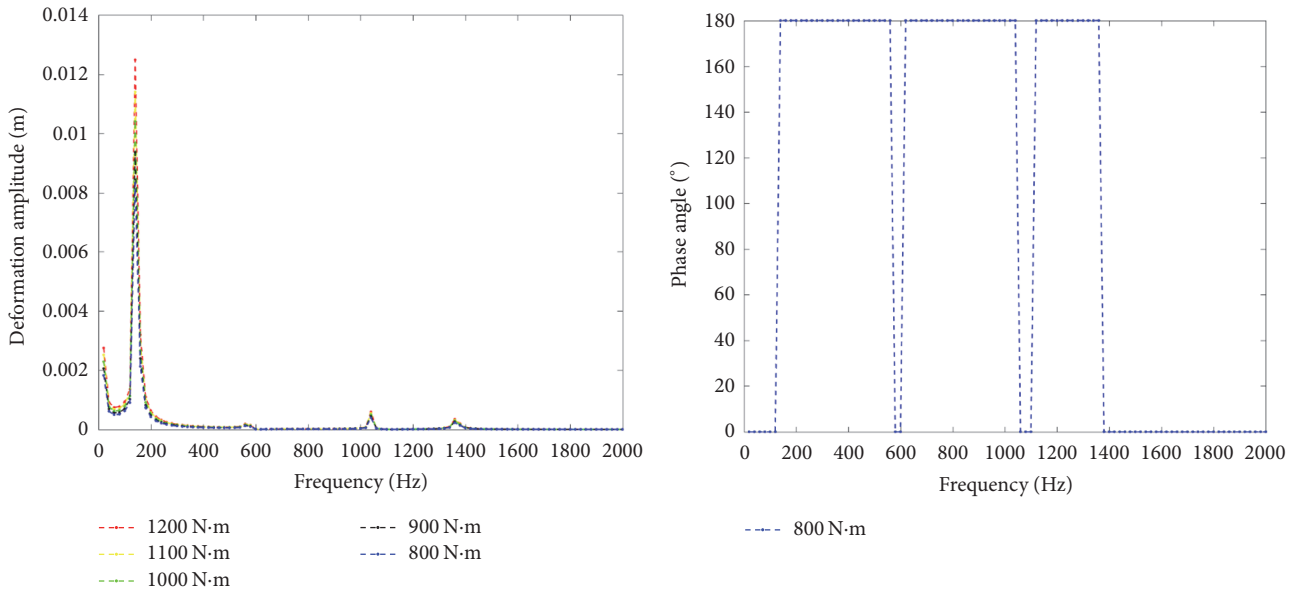
(d) Acceleration phase angle

Figure 4: Harmonic response spectrum of spiral bevel gear.

In low frequency interval, the tooth fracture model has a significant peak in the deformation amplitude spectrum, while the healthy tooth model has a valley in this interval. However, in the acceleration amplitude spectrum, the peaks of two models are equivalent.

In intermediate frequency interval, the tooth fracture model does not show an obvious peak in deformation and acceleration amplitude spectrum, while the peak value of healthy tooth model is very obvious. And there is no other interference signal in this interval and adjacent interval, so it can be effective in monitoring frequency interval for tooth fracture detection.

In high frequency interval, the acceleration amplitude spectrum is peaked at about 1050 Hz in fracture model, and



(a) Deformation amplitude

(b) Deformation phase angle

Figure 5: Influence of torque on harmonic response of spiral bevel gear.

the peak of healthy tooth model is four times that of the fracture model.

Compared with Figures 4(a) and 4(c), several peaks of the fracture are in the vicinity of the sidebands of the healthy, and for each corresponding peak, the fracture is less than the healthy; as is shown in Figure 4(a), the peaks also gradually reduce with the increase of the frequency, so high frequency interval cannot easily detect fracture tooth signal. However, acceleration signal is mainly concentrated in intermediate and high frequency intervals.

Compared with Figures 4(b) and 4(d), it can be seen that deformation phase spectrum and acceleration phase spectrum are corresponding to the contrary. And during the sweep process, the phase will be reversed between 0° and 180° in different frequencies. In low frequency interval, the healthy and the fracture coincide and separate from each other quickly, so it is difficult to judge whether the tooth is fractured. The phase mutation inversion is most pronounced at about 600 Hz in the intermediate interval, and it is depicted in both the deformation and acceleration phase spectra. In addition, phase reversal at 1400 Hz in high frequency interval is also relatively obvious.

In brief, low frequency, intermediate frequency, and high frequency interval sections all have tooth fracture characteristic signal; however, the intermediate frequency interval (200 Hz–1000 Hz) is the best interval for detection among others.

3.2. Influence of Torque on Harmonic Response of Spiral Bevel Gear. In order to investigate the effect of torque (dynamic meshing force) and other excitations on the harmonic response of the spiral bevel gear system, tooth fracture model is target object, and the torque of different sizes is loaded without changing the other pretreatment process, so excitation's influence on tooth fracture detection could be obtained. The harmonic response of each load is calculated, respectively, and the result is shown in Figure 5. It can be seen from Figure 5(a) that the torque magnitude has a great influence on the peak value in low frequency interval, and the peak value increases with the torque obviously. However, the torque magnitude has little impact on the intermediate frequency and high frequency intervals. As can be seen from Figure 5(b), the phase spectrum does not change with the increase of torque; that is, torque does not affect the harmonic response phase. This conclusion successfully verifies the experimental test in [12].

3.3. Influence of Measuring Point Position on Harmonic Response of Spiral Bevel Gear. In order to improve the accuracy of the test, the influence of the measuring point position on the harmonic response of the spiral bevel gear system is studied. The responses of the different measuring points are calculated without changing the other pretreatment, so the best monitoring area could be found. The position of the test points is shown in Figure 6, where point A is the measuring point near the working area. The harmonic response spectra of each measurement point are calculated, respectively, and the results are shown in Figure 7. It can be seen from Figure 7(a) that the peak of point C is more obvious in the low frequency interval, and the peak of point A is prominent in the intermediate frequency interval, whereas the vibration of each point is not visible at high frequency interval. According to Figure 7(b), each point has obvious phase mutation inversion at about 600 Hz (intermediate frequency interval) and 1400 Hz (high frequency interval). Based on the above analysis and the conclusion in Section 3.1,

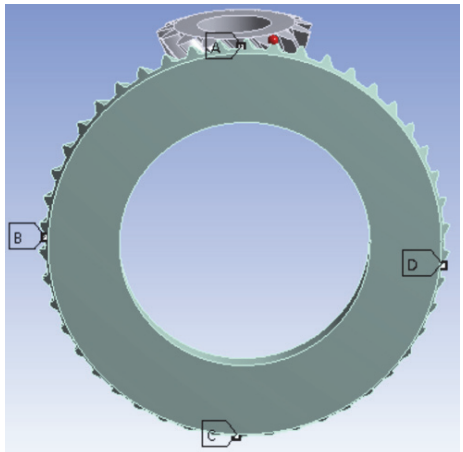
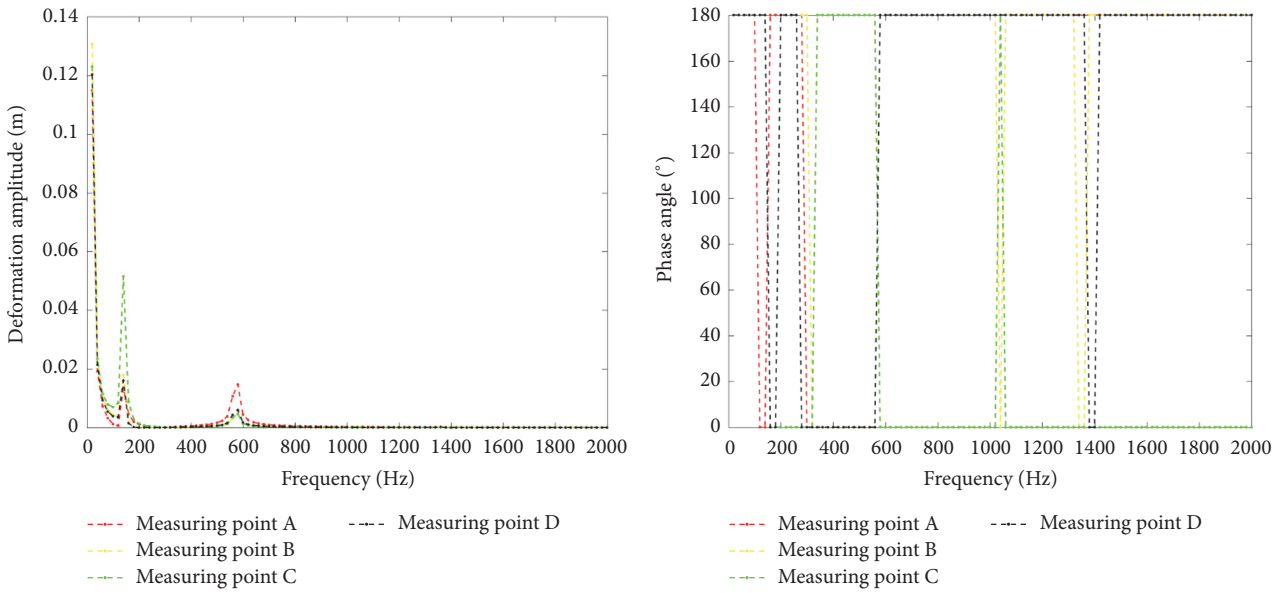


Figure 6: Measuring points position.



(a) Deformation amplitude

(b) Deformation phase angle

Figure 7: Harmonic responses of spiral bevel gear under different measuring points.

it can be seen that the measuring point near the working area is the best test point when the system is monitored under intermediate frequency interval.

4. Conclusion

Based on the meshing principle and gear cutting process, the tooth involute equation of the bevel gear is obtained, and the parametric spiral bevel gear three-dimensional model and the finite element model are established.

The analysis results draw the following conclusions:

(1) Compared with the healthy tooth model, natural frequency in each order decreases in tooth fracture model, and the natural frequency deviation of the fourth and sixth order is relatively large. Although the low frequency, intermediate frequency, and high frequency interval section have tooth fracture signals, the intermediate frequency interval (200 Hz–1000 Hz) is the best interval to determine whether the tooth is fractured.

(2) The effect of torque on the harmonic response of the spiral bevel gear is quite little, and the change of excitation does not affect the fractured tooth monitoring.

(3) The amplitude of the vibration signal near the working area is relatively strong, and it is the best measuring point area for detecting the fractured tooth signal.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work described in this paper was fully supported by the National Natural Science Foundation of PRC (Grant nos.

51375226 and 51475226); Postgraduate Research and Practice Innovation Program of Jiangsu Province; China Scholarship Council's support for joint research with Professor Yeping Xiong in University of Southampton.

References

[1] Y. C. Tsai and P. C. Chin, "Surface geometry of straight and spiral bevel gears," *Journal of Mechanisms, Transmissions, and Automation in Design*, vol. 109, no. 4, pp. 443–449, 1987.

[2] F. L. Litvin, A. Fuentes, Q. Fan, and R. F. Handschuh, "Computerized design, simulation of meshing, and contact and stress analysis of face-milled formate generated spiral bevel gears," *Mechanism and Machine Theory*, vol. 37, no. 5, pp. 441–459, 2002.

[3] F. L. Litvin, A. Fuentes, and K. Hayasaka, "Design, manufacture, stress analysis, and experimental tests of low-noise high endurance spiral bevel gears," *Mechanism and Machine Theory*, vol. 41, no. 1, pp. 83–118, 2006.

[4] A. Fuentes, F. L. Litvin, B. R. Mullins, R. Woods, and R. F. Handschuh, "Design and stress analysis of low-noise adjusted bearing contact spiral bevel gears," *Journal of Mechanical Design*, vol. 124, no. 3, pp. 524–532, 2002.

[5] I. Gonzalez-Perez, A. Fuentes, and K. Hayasaka, "Analytical determination of basic machine-tool settings for generation of spiral bevel gears from blank data," *Journal of Mechanical Design*, vol. 132, no. 10, Article ID 101002, 2010.

[6] J. Argyris, A. Fuentes, and F. L. Litvin, "Computerized integrated approach for design and stress analysis of spiral bevel gears," *Computer Methods Applied Mechanics and Engineering*, vol. 191, no. 11-12, pp. 1057–1095, 2002.

[7] G. I. Sheveleva, A. E. Volkov, and V. I. Medvedev, "Algorithms for analysis of meshing and contact of spiral bevel gears," *Mechanism and Machine Theory*, vol. 42, no. 2, pp. 198–215, 2007.

[8] F. L. Litvin, G. I. Sheveleva, D. Vecchiato, I. Gonzalez-Perez, and A. Fuentes, "Modified approach for tooth contact analysis of gear drives and automatic determination of guess values," *Computer Methods Applied Mechanics and Engineering*, vol. 194, no. 27-29, pp. 2927–2946, 2005.

[9] V. Simon, "Computer simulation of tooth contact analysis of mismatched spiral bevel gears," *Mechanism and Machine Theory*, vol. 42, no. 3, pp. 365–381, 2007.

[10] J. Zakrajsek, R. Handschuh, and H. Decker, "Application of fault detection techniques to spiral bevel gear fatigue data," 1994.

[11] P. J. Dempsey, R. F. Handschuh, and A. A. Afjeh, "Spiral bevel gear damage detection using decision fusion analysis," in *Proceedings of the 5th International Conference on Information Fusion*, pp. 94–99, July 2002.

[12] H. Decker and D. Lewicki, "Spiral bevel pinion crack detection in a helicopter gearbox," 2003.

[13] A. Ural, G. Heber, P. A. Wawrzynek, A. R. Ingraffea, D. G. Lewicki, and J. B. C. Neto, "Three-dimensional, parallel, finite element simulation of fatigue crack growth in a spiral bevel pinion gear," *Engineering Fracture Mechanics*, vol. 72, no. 8, pp. 1148–1170, 2005.

[14] M. Li and H. Y. Hu, "Dynamic analysis of a spiral bevel-geared rotor-bearing system," *Journal of Sound and Vibration*, vol. 259, no. 3, pp. 605–624, 2003.

[15] L. Yinong, L. Guiyan, and Z. Ling, "Influence of asymmetric mesh stiffness on dynamics of spiral bevel gear transmission system," *Mathematical Problems in Engineering*, vol. 2010, Article ID 124148, 13 pages, 2010.

[16] Z. Feng and C. Song, "Effects of geometry design parameters on the static strength and dynamics for spiral bevel gear," *International Journal of Rotating Machinery*, vol. 2017, Article ID 6842938, 8 pages, 2017.

[17] Q. Fan, "Computerized modeling and simulation of spiral bevel and hypoid gears manufactured by Gleason face hobbing process," *Journal of Mechanical Design*, vol. 128, no. 6, pp. 1315–1327, 2006.

[18] K. Shi, J. Xia, and C. Wang, "Design of noncircular bevel gear with concave pitch curve," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 227, no. 3, pp. 542–553, 2012.

[19] W. Levine, *The Control Handbook*, CRC Press, Boca Raton, Fla, USA, 1996.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 7097561, 8 pages
https://doi.org/10.1155/2017/7097561

Research Article

Modeling of Complex Life Cycle Prediction Based on Cell Division

Fucheng Zhang, Yahui Wang, and De Zhang

Institute of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Correspondence should be addressed to Fucheng Zhang; 475521864@qq.com

Received 9 August 2017; Accepted 16 October 2017; Published 3 December 2017

Academic Editor: Chunhui Zhao

Effective fault diagnosis and reasonable life expectancy are of great significance and practical engineering value for the safety, reliability, and maintenance cost of equipment and working environment. At present, the life prediction methods of the equipment are equipment life prediction based on condition monitoring, combined forecasting model, and driven data. Most of them need to be based on a large amount of data to achieve the problem. For this issue, we propose learning from the mechanism of cell division in the organism. We have established a moderate complexity of life prediction model across studying the complex multifactor correlation life model. In this paper, we model the life prediction of cell division. Experiments show that our model can effectively simulate the state of cell division. Through the model of reference, we will use it for the equipment of the complex life prediction.

1. Introduction

Industrial systems and large-scale application equipment work in the long term in the high pressure fluctuations, corrosive, and other harsh environments. After a certain period of time, it will inevitably fail. When the fault is repairable, maintenance is usually done, but the maintenance of the defective parts will generally reduce the accuracy, performance, and reliability of the equipment. It ultimately leads to degradation of equipment functions and improves the overall cost of equipment spending. With the increase in the number of equipment repairs, the degree of functional degradation and maintenance costs gradually increased [1]. Therefore, when we make decisions on repairing industrial systems and large equipment, we need to balance the factors like prolonged service life, equipment downturn, increased cost, and other factors. After several repairs, the functional decline in the equipment and the total cost of the expenditure are greater than the extension of the life of the industrial equipment. At this point, it makes no sense to continue to repair. In order to avoid overrepair, we need to do a proper repair of the equipment and estimate a reasonable life, Ensuring the balance between safety and economy of industrial equipment and realizing the health management of gas industrial equipment in complex environment. It is necessary to establish a life forecast of industrial equipment related to multiparameters (including operating environment, service life, functional recession, and cost stacking) model. In order to establish an effective model for predicting equipment life, we learn from the biological perspective and abstract the life prediction model of the normal equipment from the biological life model.

Determining the life of industrial equipment involves many factors, and its use environment is very complex. In principle, it can be compared to the determination of biological life as shown in Table 1. In order to predict the life of industrial equipment, we need to draw on the life prediction method of the cell to obtain the mathematical model of the limited life of industrial equipment [2].

Figure 1 is a schematic diagram of the industrial equipment life estimation based on the limited division mechanism [3] of biological cells. Its main body has two parts: one is the industrial equipment system and the other is the biological system. We divided the factors that affect the life of the living body to internal factors and external factors. The same factors affecting the life of industrial equipment are also divided into internal factors and external factors. The establishment of a life-based estimation model based on the limited division mechanism of biological cells requires a clear understanding of the process structure of cell division (the

Table 1: Analogy of fault diagnosis in biology and engineering.

Medical science	Engineering
Disease diagnosis Through observation, laboratory tests, and medical equipment The conclusion is given by the doctor	*Fault diagnosis* Through sensors, signal processing and testing instrument The conclusion is given by diagnostic software
Multidisciplinary consultation The doctor diagnosed the disease in many ways	*Comprehensive diagnosis* Diagnose current faults with multiple diagnostic techniques
Disease prevention and health care Physical examination, disease prediction, health care system (health, subhealth, life expectancy, and safeguard measures: past, present, and future)	*Fault diagnosis and health management* Health monitoring, fault diagnosis, equipment life prediction (health, subhealth, failure, component life prediction, decision/maintenance recommendations, past, present, and future)

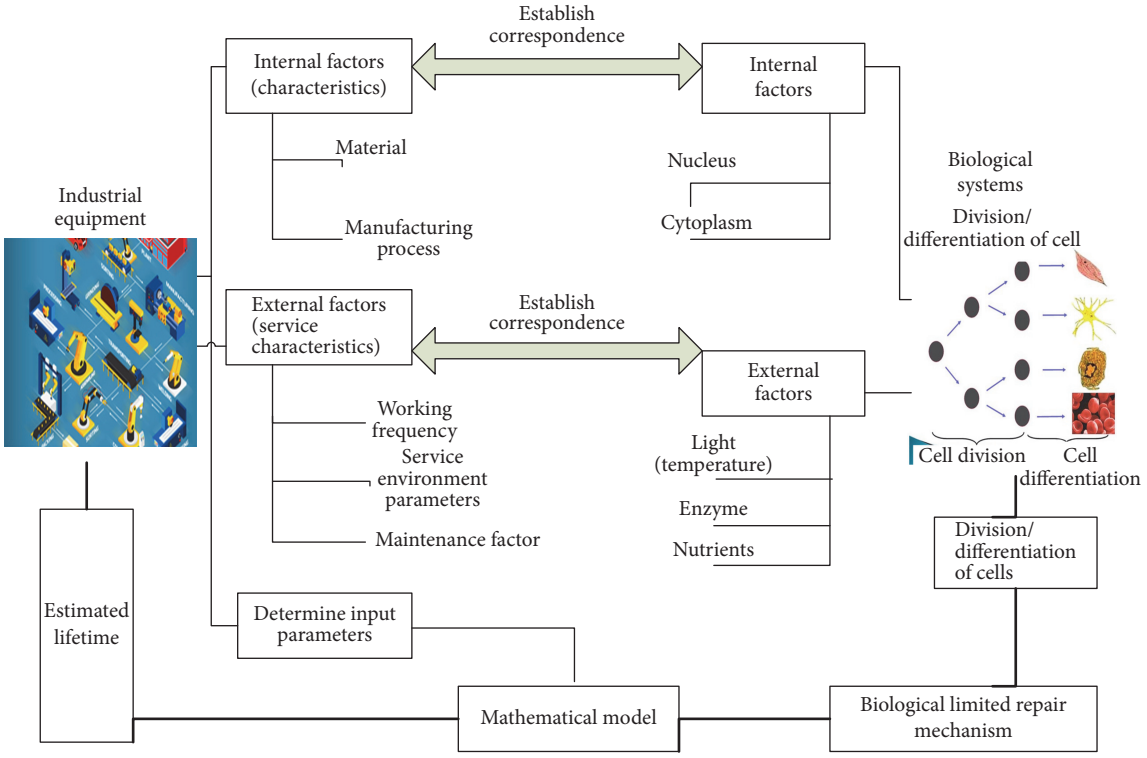


Figure 1: A schematic diagram of biological cell finite splitting mechanism and industrial equipment life estimation.

limited repair mechanism of the organism). Second, it is necessary to clarify which parameter variables are related (internal nuclear parameters, external light, nutrients, and other factors). A mathematical model for estimating the life expectancy of industrial equipment with moderate heterozygosity is abstracted from the limited repair mechanism of organisms. The relation between the parameter variables of an organism and the physical parameters of an industrial device determines the input of the mathematical model for the estimation of the lifetime of an industrial device and ultimately to achieve the purpose of estimating the life of industrial equipment and to achieve a reasonable, nonexcessive maintenance.

2. Complex System

The study of complex systems was the rise of the "Complex Systems," the album published by the University of Science on April 2, 1999. Its emergence is closely related to the study of complexity. The question about complexity originated in Austria, and in 1928 Bertalanffy raised the question of complexity for the first time in his essay on "biological organism systems."

A system is a combination of elements that interact more and more with each other. Any system has three aspects that need to be studied: entity, attribute, and activity. Because of the interaction between the entities that make up the system, the concept of "state" is usually described. The research system is to study the state change of the system, that is, the evolution of the system.

2.1. The Classification and Characteristics of the System

(1) According to the Size of the System. There are small system, large system, and giant system.

(2) According to the Degree of Association for Subsystems within the System. There are simple system and complex system.

(3) According to the System and the outside World Contact Situation. There are isolated system and open system.

The term "complex system" appeared on the "Complex Systems" published by the University of Science on April 2, 1999, and the magazine made a brief description of the "complex system": by understanding the components of a system (subsystems), the nature of the system cannot be fully explained, and such systems are called complex systems. The complex system is characterized by high order, being multiloop, being nonlinear, multitime scale, being hierarchical, openness, uncertainty, morbid structure, and so on [4].

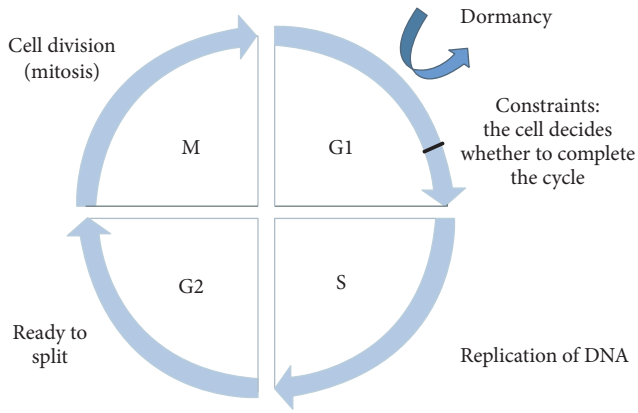


Figure 2: Cell division cycle.

3. Cell Division Systems in Organism

Medical experiments have amply demonstrated that the human body's fiber cells can only be divided into 50 to 60 generations; thus the human lifespan is limited [5]. The discovery of cell-limited division mechanism has opened a wide world for the discussion of "biological life" from the cell level. Postdoctoral Blackburn at Yale University found telomere structure of the end of the cell chromosome; further studies have also found telomerase; the role of telomerase is to take off the telomere to replenish up, so that, for telomere to maintain a certain length, the cell can continue to divide and to avoid aging and apoptosis. However, after the differentiation and maturation of various tissues in normal adults, the gene for producing telomerase was closed. It can be seen that in the process of human evolution, the mechanism of cell cycle regulation has evolved into a very ingenious system, which controls cell division properly and ensures the stability, rapidity, and accuracy of cell division.

The parameters that affect cell division are as follows: ① Cell volume: DNA, RNA, ribose, and proteins make up the quality of the cell; the cells must meet some of the necessary conditions before splitting, one of which is the duplication of DNA; ② the interaction force between the nuclei: during the period of normal metabolism, there is interaction between the nucleus and the cytoplasm (with the exchange of matter, energy, and information, internal and external factors); ③ regional area mass density having an effect on cell division; ④ the rate of cell division related to various environmental factors.

Cell division is the basis of cell differentiation, tissue renewal, and biological reproduction. As early as 19th century, the researchers analyzed the process of cell division, which was based on qualitative analysis, and quantitative analysis can help to study the biological reaction process of various factors within the cell division and analyze the details of cell division.

Figure 2 shows each stage of cell division.

Cell cycle refers to the normal continuous division of cells from the previous mitosis to the next time of the completion of the continuous dynamic process, but also multistage, multifactor participation in the precise and orderly regulation process, which can be divided into 5 periods, that is, G0 period (Gap0, resting period), G1 period (Gap1, prophase of DNA synthesis), S phase (DNA synthesis phase, DNA synthesis period), G2 period (anaphase of DNA synthesis), and M phase (mitosis, mitotic period). Two main limit points are found, that is, 1 G/S limit points (from G1 phase to S phase control cells can prevent the damage of base replication and repair mutation of chromosome) and G2/M limit point (cells in two control points can make the cell before the split phase repairing on the duplicated DNA damage). By limiting the point, even if the stimulus signal is removed, the cell will still begin to replicate DNA. There are many enzymes involved in the mammalian cell cycle. But one of the most important enzymes is a group of two complexes, which act as a cyclic protein-dependent kinase (CDKs) and a group of periodic proteins (cyclins), which act as a subunit of the subunits and are a kind of very significant silk/threonine protease [6, 7].

The main research content of this paper is to systematically analyze the cycle of cell division from the point of view of the system and analyze the influence of the elements in the cell on the cell division cycle from the simulation. The relationship between cell life cycle and internal components of cells was obtained. The relationship between cell cycle and internal components was obtained by simplifying the cell cycle by cell distribution.

4. Analysis and Modeling of Cell Division Systems

4.1. The Role of Various Cell Division Proteins and Their Influencing Factors

Gene Regulatory Protein E2F, pRB. Most animal cells over the initial checkpoint rely on regulatory protein-inducible G1/S gene expression. The most important ones of these regulatory factors is the E2F family, which regulates the expression of most genes in the G1 term (e, a). The precise regulation of the E2F activity ensures that the G1/S gene is expressed at the appropriate point in the cell cycle. However, the function of E2F is mainly controlled by PRB protein, and in stationary cells, pRB inhibited the expression of E2F target gene.

pRBp Protein. It provides a cell cycle brake device, but it is not essential to regulating the cell cycle process but mainly involved in cell cycle withdrawal and maintenance of cell resting state. When E2F-dependent protein expression is not inhibited by pRB, the cells exit the mitosis directly from the Start checkpoint.

CDK Inhibitory Protein. P27 is a protein that has negative effects on CDK kinase activity in cells. It can inhibit the binding of CDK kinase and play an important role in the negative regulation of cell cycle.

Cycle Protein D. Cyclin D mainly regulates the G1 phase of cells, and its major synthesis is activated by AP1 (activation of growth factor GF). The transcription factor E2F promotes it, the inhibitory factors pRB and pRBp inhibit its transcription, and cyclin D can activates the cycle protein-dependent enzyme Cdk4–6, and then the active complex in the cell cycle plays the primary role of pRB phosphorylation into pRBp.

Cyclic Protein E. Cyclic protein E mainly regulates G1/S phase, and its synthesis is mainly activated by E2F, and it is inhibited by pRB, which can be rapidly degraded by skp2. Cycle E can activate Cdk2 protein E. The main role of the synthetic complex cyclin E/Cdk2 is to phosphorylate pRBp to pRBpp, followed by phosphorylation of p27 to degrade, and when p27 is degraded, the cells cross the G1/S test point and enter the S period.

Cyclic Protein A. Cyclic protein A mainly regulates the cell S phase and crosses S/G2 checkpoint. Its synthesis is mainly affected by E2F activation, while being inhibited by pRBp with degradation of APC-PCdc20. Cyclin A activates Cdc2 to form a complex, cyclin A/Cdk2, which is inhibited by p27 binding in the G1 phase and then released in the S phase. The complex can phosphorylate E2F simultaneously, reducing the synthesis of cyclin D, cyclin E, and cyclin A and inactivating APC (late-promoting complex). APC (late-promoting complex) is a large E3-linked enzyme that mediates cell cycle-regulated proteins that rely on ubiquitin-hydrolyzed proteins to control events during replication and cell division, combining ubiquitin and substrate, and degrade by proteasome.

Cycle Protein B. Cycle protein B mainly controls cell internal checkpoints across the G2/M and M. It is mainly suppressed by pRBp. It can be degraded by APC^{Cdh1}. When Cdk1 is activated, complex cyclin B/Cdc1 nuclear protein phosphorylation fiber layer will trigger the nuclear membrane decomposition while phosphorylating APC, so that APC can be activated by Cdc20. APC^{Cdc20} has a biodegradable effect on cyclin A and cyclin B. It breaks the protein safety factor, releases the isolated enzyme, destroys the connection between the sister chromosomes, and promotes cell mitosis.

4.2. Internal Processes and Modeling of Cell Division Cycles

4.2.1. Cell Division Process. Growth factor, GF, promotes the expression of cyclin D and cyclin D with cdk4–6 (cyclin-dependent kinases) and phosphorylates pRB to pRBp, and with the decrease of PRB, E2F and cyclin E begin synthesizing. E2F synthesis accelerates the synthesis of cyclin E and cyclin A. Cyclic protein E can phosphorylate pRBp to pRBpp, while promoting the inhibition of phosphorylation of protein p27 and then release periodic protein A. With the decrease of p27, cyclin A increases and the cells enter S phase. Cyclin A promotes the phosphorylation of E2F and degrades Cdh1. At the same time, the binding of APC to Cdh1 is delayed and the APC is inactivated. A decrease in APC activity degrades cyclin E. During the release of cyclin B, the cell division cycle enters M phase. At the same time cyclin B phosphorylated APC, during degradation of protein safety factor, and then released the separation of enzymes, to promote cell division.

4.2.2. Cell Division System Modeling. Cell division modeling consists of 3 parts [8]: synthesis, degradation, and regulation. The differential equation for each protein is

$$\frac{d\text{Cyclin}}{dt} = V_{\text{synthesis}} - V_{\text{Degradation}} + \left(V_{\text{Activation}} - V_{\text{inactivation}}\right). \tag{1}$$

Synthesis Regulation. If n transcription factor (X_i) activates or suppresses the transcription of a gene and each gene is only combined with one transcription factors, the rate of periodic protein synthesis is

$$V_{\text{synthesis}} = \frac{K_0 + K_1 * K_{M1} X_1 + \cdots + K_n * K_{Mn} * X_n}{1 + K_{M1} * X_1 + \cdots + K_{Mn} * X_n} * \text{DT}, \tag{2}$$

where DT is the gene concentration, K_{Mi} is the ratio of the transcription factor to the gene binding and separation, and K_i is the transcription rate after the transcription factor binds to the gene.

Degradation Regulation. The degradation of protein A is regulated by the degradation factor B_i in n and each protein binds only to one transcription factor; then the degradation rate is

$$V_{\text{Degradation}} = K_0 * a + K_1 * BT_1 * \frac{n}{A + K_{M1}} + \cdots + K_n * BT_n * \frac{A}{A + K_{Mn}}, \tag{3}$$

where BT_i is the total concentration of degradation factor B_i, K_{Mi} is the ratio of degradation factor and protein separation to synthesis, and K_i is the degradation rate of degradation factor and protein synthesis.

Modification Regulation. Protein modification is an enzymatic reaction, which is mainly Rice kinetic:

$$V_{\text{Activation/inactivation}} = K * [E_0] * \frac{[S]}{K_{M+[S]}}. \tag{4}$$

E_0 is the total concentration of enzymes, S is the substrate concentration, and K_M is the Michaelis constant.

The process of synthesis modification of cyclin A in cells is as follows:

$$\frac{dA}{dt} = V_{\text{synthesis}} - \left(K_0 * A + K_1 * \text{Cdc20a} * \frac{A}{K_{M1} + A}\right). \tag{5}$$

$dA/dt = 0$; you can know when Cdc20a (it is cell cycle regulator) = 0 that we get the maximum $A = V_{\text{synthesis}}/K_0$. When Cdc20a is large enough, $A_{\min} = V_{\text{synthesis}}/(K_1 * \text{Cdc20a} - V_{\text{synthesis}}) * K_{M1}$, in direct proportion to K_{M1}.

Then the formula can be simplified as follows:

$$\frac{dA}{dt} = -K_0 * \frac{A^2 - \left(V_{\text{synthesis}} - K_1 * \text{Cdc20a} - K_0 * K_{M1}\right)/K_0 * A - \left(K_{M1} * V_{\text{synthesis}}\right)/K_0}{K_{M1} + A}. \tag{6}$$

Make

$$A^2 - \frac{V_{\text{synthesis}} - K_1 * \text{Cdc20a} - K_0 * K_{M1}}{K_0} * A - \frac{K_{M1} * V_{\text{synthesis}}}{K_0} = 0. \tag{7}$$

Because of $(K_{M1} * V_{\text{synthesis}})/K_0 > 0$, the equation has two positive and negative solutions, which are assumed to be A_1 (<0) and A_2 (>0). Then

$$A_1 + A_2 = \frac{V_{\text{synthesis}} - K_1 * \text{Cdc20a} - K_0 * K_{M1}}{K_0}$$
$$A_1 * A_2 = \frac{K_{M1} * V_{\text{synthesis}}}{K_0}. \tag{8}$$

Therefore, formula (8) can be reduced to

$$\frac{dA}{dt} = -K_0 * \frac{(A - A_1) * (A - A_2)}{K_{M1} + A}. \tag{9}$$

Suppose $K_{M1} + A = a * (A - A_1) + b * (A - A_2)$; then $A = A_2$, $dA/dt = 0$, $A \neq A_2$; at that time,

$$\frac{dA}{dt} = -K_0 * \frac{1}{a/(A - A_2) + b/(A - A_1)}. \tag{10}$$

In the time domain solution,

$$(A - A_1)^b + |A - A_2|^{a*|A-A_2|/(A-A_1)} = R * e^{-K_0 * t} \tag{11}$$

when $t = 0$, $A = 0$, $R = (-A_1)^b + A_2^{-a}$, when $A > A_2$, $dA/dt < 0$, and when $A < A_2$, $dA/dt > 0$. When $t \to +\infty$, the steady-state value is A_2. Time constant $T = 1/(((K_1 * \text{Cdc20a})/K_{M1}) + K_0)$ mainly affected by $(K_1 * \text{Cdc20a})/K_{M1}$. If we want to change the system's response rate, we can increase the value $(K_1 * \text{Cdc20a})/K_{M1}$.

The proteins involved in all expected processes during cell division are bound to follow this synthesis modification process.

5. Simplification of Cell Division Model

Cell division was mainly manifested in the form of intracellular protein cycle changing. Because cell division is a continuous process, the content of the protein that assists in completing the cell division is fluctuating. The main factor in maintaining this ongoing process was determined by the number of telomeres in the cell. When the number of telomeres was sufficient, the cells will divide when they receive excitation information. Until the number of telomeres was not sufficient to allow the cells to split, the division stopped.

But because of the limited telomere length, the process of cell division would cease when the telomere number drops to a certain extent. So we assumed that the number of intracellular telomeres was a linear descent process, assuming that the initial telomere number was K, each time in the division, the cells automatically shed j telomeres; when the remaining amount of telomeres reaches the set threshold the cell division ends.

In the experiment, we assumed that, in the case of cells that had been activated, it was found that there was an internal cycle between cyclin A and cyclin B from a small cycle of division within the cell. We showed it in Figure 3. Assuming that the influencing factor in this cycle was sufficient, the split within a small cycle can continue. After neglecting some of the variables, we studied the periodicity of cell division, and we could show that the cell cycle was also expressed in this small cycle.

In the model without periodic protein D, the factor pRB was assumed to be zero. After the absence of the pRB factor, its effect factor E2F has been at a high level without inhibition, so we assumed that E2F had been in high expression in the cell (at a fixed concentration). So we ignored the synthetic equation of E2F and its compounds. In the whole model, there was a cell growth factor, and because the growth factor only affects the stability of the first stage of cyclin D, we neglected cyclin D, since the target of cyclin D was pRB. Because of the absence of pRB, the action product of cyclin E was absent, so the first function of cyclin E was ineffective. The function of cyclin E was phosphorylated by pRBp, which promoted cyclin A (inhibited by pRBp) expression.

When the pRB factor was 0, cyclin D was ineffective, and because pRB was ineffective, cyclin E lost the phosphorylation of pRBp, simply phosphorylation of p27, and because of its modification of p27, it had little effect on p27. So we ignore cyclin E, p27.

The postactivating factors Cdh1i and Cdh1a were used in the internal regulation of cells, and their synthesis and regulation processes were neglected under the established values.

According to Figure 3, we conclude the final simplified equation:

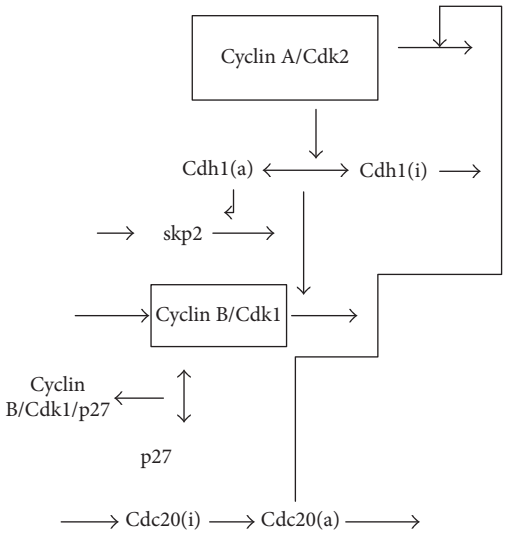


Figure 3: Simplified cell model.

$$\frac{dA}{dt} = V_{e2fa} - V_{da} * \text{Cdc20a} * \frac{A}{k_{da} + A} - k_{dda} * A$$

$$\frac{d\text{Cdh1a}}{dt} = V_{1cdh1} * \frac{1 - \text{Cdh1a}}{k_{1cdh1} + 1 - \text{Cdh1a}} - V_{2cdh1} * A * \frac{\text{Cdh1a}}{k_{2cdh1} + \text{Cdh1a}} \quad (12)$$

$$\frac{dB}{dt} = V_b - k_{db2} * \text{Cdh1a} * \frac{B}{k_{bm2} + B} - k_{db} * B$$

$$\frac{d\text{Cdc20a}}{dt} = V_{m1b} * B * \frac{1 - \text{Cdc20a}}{k_{1b} + 1 - \text{Cdc20a}} - V_{m2b} * \frac{\text{Cdc20a}}{k_{2b} + \text{Cdc20a}}.$$

(V_{e2fa} is the internal synthesis rate of protein A, V_{da} is the degradation rate, and k_{dda} is the modifier coefficient. V_{1cdh1} is the internal synthesis rate of protein Cdh1a; V_{2cdh1} is the degradation rate of protein Cdh1a. V_b is the internal synthesis rate of protein B; k_{db2} and k_{db} are the modifier coefficients. V_{m1b} is the internal synthesis rate of protein Cdc20a; V_2 is its degradation rate of protein Cdc20a; k_{1b} and k_{2b} are protein Cdc20a's modifier coefficients.)

6. Experimental System Simulation

The experimental environment for this experiment is Matlab2012a version, 64-bit Win7 system, CPUi5-4790T, core frequency 2.0 GHZ, and memory 6 G.

In the experiment of simplified model, all of our intrinsic parameters were derived from the whole model of the cell, and the parameters were adjusted to the basis of the existing experimental proof, which was beneficial to the experiment.

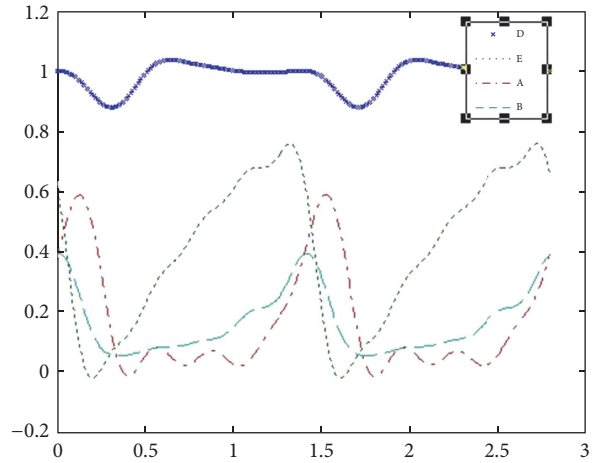


Figure 4: Complete cell division system simulation.

Table 2 shows the internal parameters of simplified equation (12) for the model.

The internal parameter was brought into the equation of the simplified model, and the function ode45s (Runge-Kutta method) was used to solve the high order differential equation of the complete model of the cell, and the simulation results were obtained.

Figure 4 shows the results of our overall model of cell division, which expressed four periodic proteins that were involved in the process of cell division. They were cyclin D, cyclin E, cyclin A, and cyclin B (the rectangle diagram in the upper right corner of Figure 4; the star line was cyclin D, short dashed line was cyclin E, the point line was cyclin A, and the long dashed line was cyclin B).

Through experimental simulations, we set the values of the variables of the various proteins in the cell and set their initial values at the same time. Our model entered the simulation with the passage of time. In the simulation results of the model, we could observe the various periodic rules of the four periodic proteins in the cells at the beginning of cell division and how their relationship changed.

The simulation results of this model completely showed the process and periodicity of the four major proteins in the cell division cycle. Cyclin D entered the G1 phase after the cells began to divide; then cyclin D began to increase gradually. After that the cells were introduced into the S phase, the expression of protein E began to rise gradually. With the increase of cyclin E, the cells began to flow through S phase and enter the G2 phase. Cycle protein D began to decrease gradually. During the S/G2 period, circulating proteins A and B began to increase. When the cells entered the S phase, protein E began to decline rapidly, because G2 to M period is very short, so the intracellular circulating protein D, protein E, and protein A rapidly decline. When the cells passed M phase and completed the division, the intracellular protein content of each cycle decreased to a minimum. After entering a number of cycles, we reach the end of the split points.

In the second step, as the overall model of the cell division system was too complex, we must simplify it in the future when applied to actual industrial equipment. According to

Table 2: Simplified parameter table.

Parameter	Simulation value
ve2fa	$0.1\,\mu M{\cdot}h^{-1}$
$v$da	$0.3\,h^{-1}$
$k$da	$0.1\,\mu M$
$k$dda	$0.1\,h^{-1}$
v1cdh1	$1\,\mu M{\cdot}h^{-1}$
v2cdh1	$4\,h^{-1}$
k1cdh1	$0.01\,\mu M$
k2cdh1	$0.01\,\mu M$
vb	$0.1\,\mu M{\cdot}h^{-1}$
$k$db2	$2\,h^{-1}$
Kbm2	$0.1\,\mu M$
$k$db	$0.5\,h^{-1}$
vm1b	$2\,h^{-1}$
vm2b	$0.7\,\mu M{\cdot}h^{-1}$
k1b	$0.1\,\mu M$
k2b	$0.1\,\mu M$

Table 2, we simulated the simplified cell model and obtained the results after ignoring some of the values that did not affect the overall rule.

In the simplified model of the model, the four main features we extracted were still inherently periodic and maintain a periodic curve similar to the global model. We could observe the cycle variation of the cyclin in the four main dominant simplified models and the change of the period after the change factor in the simulation of the simplified model. In this paper, Figure 5 shows the periodicity of the cyclin A, cyclin B, Cdc20a, and Cdh1a in the simplified model (where the blue dot was cyclin A, the red solid line was B, the cyan dotted line was Cdh1a, and the indigo dotted line was Cdc20a). Split starts from 0 seconds. When the telomere is in accordance with the rule of linear decline to a certain value, the split stopped. Starting from 0 seconds, protein A began to increase first, and cyclin A synergistic Cdh1a began to gradually increase. At about 1-second position, Cdch1a began to drop. Since cyclin A promotes cyclin B expression, cyclin A continues to rise, and protein B began expressing at the end of the first period, and the synergistic factor Cdc20a of cyclin B also began to rise. To the end of the division, they were attributed to the minimum. Because it was a simplified model, the cell division cycle was not complete but still can express the division of cells.

Figure 6 shows the simulation results of intracellular splitting after modifying the internal parameters of the model. When the internal influencing factor was modified, the expression cycle of each protein in the cell was abnormal; the expression cycle became longer but did not affect the normal division of the cell. However, when the impact factor dropped to a certain extent, it would influence the experimental results in Figure 7: the expression of the cell cycle protein was no longer cyclical (cell cycle damage: the cells are not split); we could analyze the cells effects of internal factors in cell division cycle.

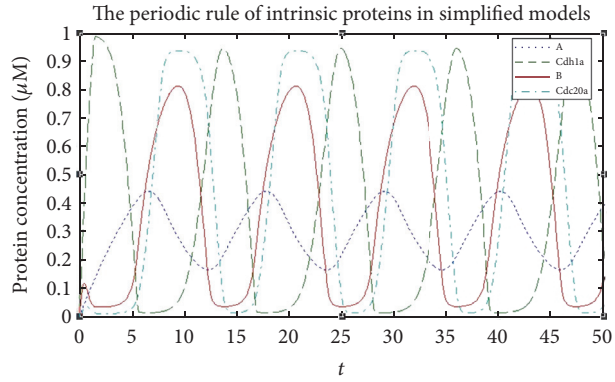


Figure 5: Periodicity of cyclin in the simplified system.

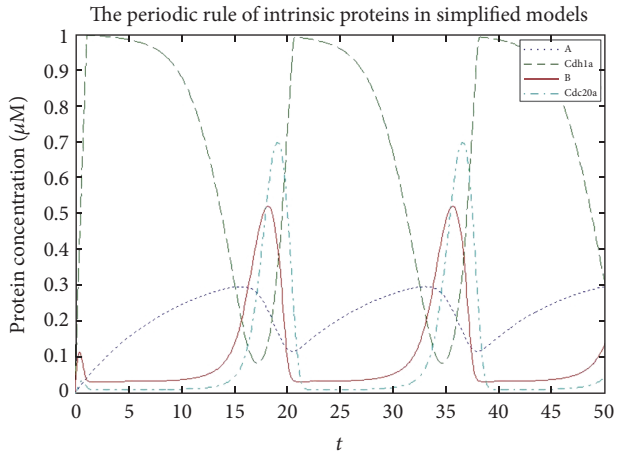


Figure 6: Periodicity of cyclin in the simplified system (changed internal parameters).

By comparing the three graphs, it was possible to draw a simple conclusion that the cell division was controlled by its internal division factor and at the expression concentration of its periodic protein. When we modify the parameters of the cell division in the model, the cell division will change (which is shown in the case where the division period is longer and the division time was different). Similarly, through real-time monitoring of internal parameters, drawing waveforms, we could clearly and intuitively observe the process of intracellular division and cell health and life conditions. From the experiment, we could have a complete understanding of the cell cycle and life cycle, so as to predict the cell life.

In modern industrial equipment, the life cycle of industrial equipment is determined by a number of similar intrinsic factors and also in the performance characteristics of some of its components. We need to understand the cell life by studying cell division and then discover the relationship between its intrinsic and performance characteristics in industrial equipment (in a particular direction in the future). It is our next task to link this conclusion to the life expectancy of large industrial equipment in our reality. Through the abstraction of cell division and lifespan prediction model, we abstract

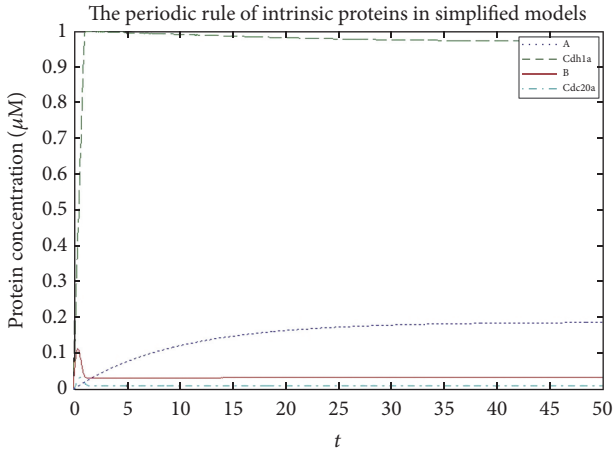


Figure 7: Termination of cell division.

a lifetime prediction model for industrial equipment and conduct bionic research on the life of industrial equipment.

7. Conclusion

Because cell division mechanism has many applications in engineering [9], the aim of this paper is to apply the model of the life cycle prediction of industrial equipment by studying the regularity of the cell life in the natural biological evolution.

Based on the experimental study and analysis, a cell division model was established to study the cell division cycle and mitotic factors. At the same time, through the observation and analysis of a variety of cell division processes, we use simplified parameters and other means to simplify the cell division model.

Finally, we used a simplified model for simulation experiments. After analyzing the results, it showed the following.

By controlling the concentration of some key factors in cell division, the effect on the normal life of a cell can be expressed intuitively. The reduction in the synthesis of cyclin A inhibits cell division and reduces the expression of cyclin B, thereby preventing cell division. But when the cyclin reaches a certain concentration the cells will continue to divide and the cell division cycle grows. Similarly, in normal industrial equipment, we can still find the parameters of the device that correspond to the factors affecting the cell division according to the model of cell division. Then, by modeling the complex system, we can predict the lifetime of the equipment.

Conflicts of Interest

The authors state that there are no conflicts of interest regarding the publication of this paper.

References

[1] C. Hu, Z. Zhou, J. Zhang, and X. Si, "A survey on life prediction of equipment," *Chinese Journal of Aeronautics*, vol. 28, no. 1, pp. 25–33, 2015.

[2] F. Ahmadzadeh and J. Lundberg, "Remaining useful life prediction of grinding mill liners using an artificial neural network," *Minerals Engineering*, vol. 53, pp. 1–8, 2013.

[3] D. McKenney and J. A. Nickel, "Mathematical model for cell division," *Mathematical and Computer Modelling*, vol. 25, no. 2, pp. 49–52, 1997.

[4] T. Hu, H. Xun, X. Zhilong et al., "Systematic modeling and analysis of cell cycle regulation mechanism of fission yeast," *Science Bulletin*, vol. 50, no. 15, pp. 1605–1612, 2005.

[5] A. d'Onofrio, "Mathematical analysis of the Tyson model of the regulation of the cell division cycle," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 62, no. 5, pp. 817–831, 2005.

[6] Y. Zhaoying and L. Jinliang, "A brief history of the development of cell cycle and its cycle control research," *Journal of Chinese Medical History*, vol. 34, no. 3, pp. 176–179, 2004.

[7] N. Wiriyasermkul, V. Boonjing, and P. Chanvarasuth, "A meiosis genetic algorithm," in *Proceedings of the 7th International Conference on Information Technology: New Generations (ITNG '10)*, pp. 285–289, Las Vegas, Nev, USA, April 2010.

[8] C. E. Dow, A. Rodger, D. I. Roper, and H. A. van den Berg, "A model of membrane contraction predicting initiation and completion of bacterial cell division," *Integrative Biology*, vol. 5, no. 5, pp. 778–795, 2013.

[9] Vaishnavikannan and S. Jhajharia, "Using cell division for text to image encryption," in *Proceedings of the 10th International Conference on Intelligent Systems and Control (ISCO '16)*, Coimbatore, India, January 2016.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 1982879, 14 pages
https://doi.org/10.1155/2017/1982879

Research Article

WOS-ELM-Based Double Redundancy Fault Diagnosis and Reconstruction for Aeroengine Sensor

Zhen Zhao,[1] Zhexu Liu,[1] Yigang Sun,[2] and Jingya Liu[3]

[1]*College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China*
[2]*College of Aerospace Engineering, Civil Aviation University of China, Tianjin 300300, China*
[3]*Tianjin Binhai International Airport, Tianjin 300300, China*

Correspondence should be addressed to Yigang Sun; caucsyg@163.com

Received 13 June 2017; Revised 8 September 2017; Accepted 13 September 2017; Published 26 November 2017

Academic Editor: Zhixing Cao

In order to diagnose sensor fault of aeroengine more quickly and accurately, a double redundancy diagnosis approach based on Weighted Online Sequential Extreme Learning Machine (WOS-ELM) is proposed in this paper. WOS-ELM, which assigns different weights to old and new data, implements weighted dealing with the input data to get more precise training models. The proposed approach contains two series of diagnosis models, that is, spatial model and time model. The application of double redundancy based on spatial and time redundancy can in real time detect the hard fault and soft fault much earlier. The trouble-free or reconstructed time redundancy model can be utilized to update the training model and make it be consistent with the practical operation mode of the aeroengine. Simulation results illustrate the effectiveness and feasibility of the proposed method.

1. Introduction

To serve as the heart of an aircraft, it is vital to ensure the security of the aeroengine [1, 2]. One way to determine whether the aeroengine is running with a fault or not is based on the data collected from the embedded sensors, which must be correct and supply the actual information. Regarding this, it is very important to employ a precise diagnosis of the aeroengine sensors' faults.

Many algorithms have been presented to diagnose the faults of the aeroengine sensors, such as Kalman filter [3], Support Vector Machine [4, 5], Neural Network [6], and wavelet [7]. Kalman filter is a widely used fault detection method. By applying Kalman filter to filter aeroengine sensors' outputs, a set of residual sequences is therefore obtained and analyzed to discover the faulty sensor. In [3], a bank of Kalman filters are used to detect and isolate sensors' faults, and another robust Kalman filter is used to detect any statistics changes happening to the sensors or the actuators. The premise of this method is to build an accurate aeroengine state variable model. However, sometimes, it is difficult or impossible to establish an accurate mathematical model. Least squares support vector machine (LS_SVM) online prediction is proposed in [4]. Through residual, which is obtained by comparing the actual outputs of the sensor and those of LS_SVM prediction model established using the real-time sensors' outputs, the sensors' faults can be detected in real-time. Wireless sensor fault diagnosis based on RBF neural network and ant colony optimization is presented in [6]. Since the connection weights, the hidden centers, and the widths have an important influence on the classification performance of the RBF neural network in the learning process, parameter selection has a close relation with the diagnosis precision. The common disadvantage of these algorithms is that they all need to set up a great deal of parameters, which increase certain difficulty for their application.

Extreme Learning Machine (ELM) [8], proposed by Huang et al., can solve this problem well. ELM only needs to set up two parameters, that is, number of hidden layer neurons and activation function [8], which may provide a more convenient way to build diagnosis model for aeroengine sensors. Furthermore, it shows faster training speed and higher model's accuracy than BP Neural Network [9, 10]. Nevertheless, ELM algorithm requires training all input data all

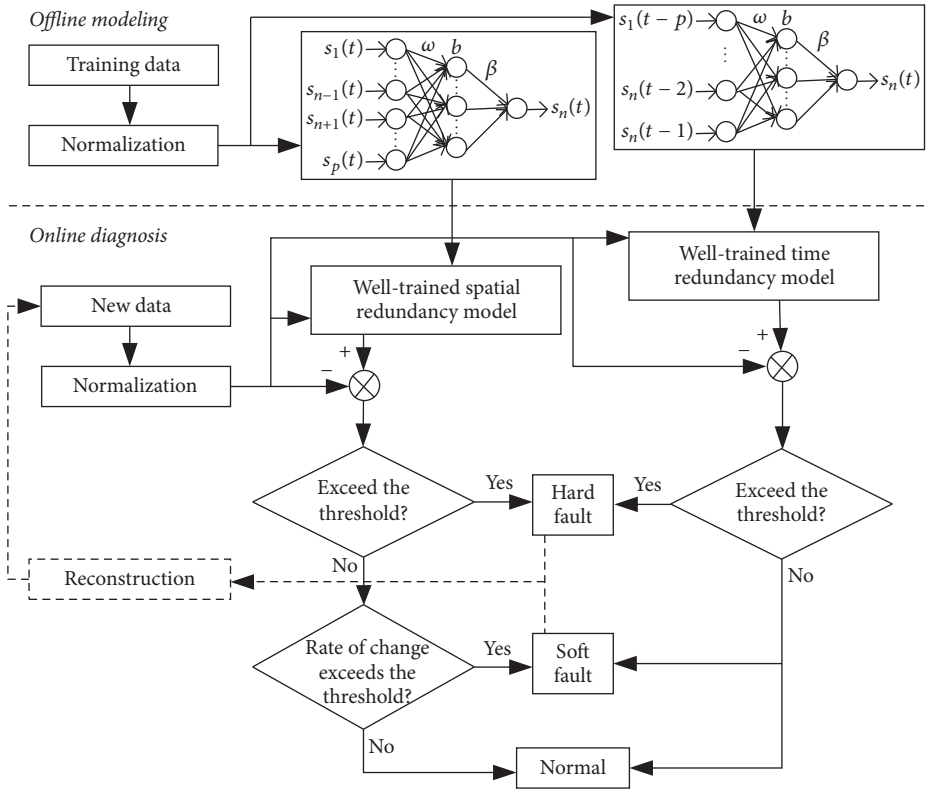


Figure 1: WOS-ELM-based double redundancy FDR method.

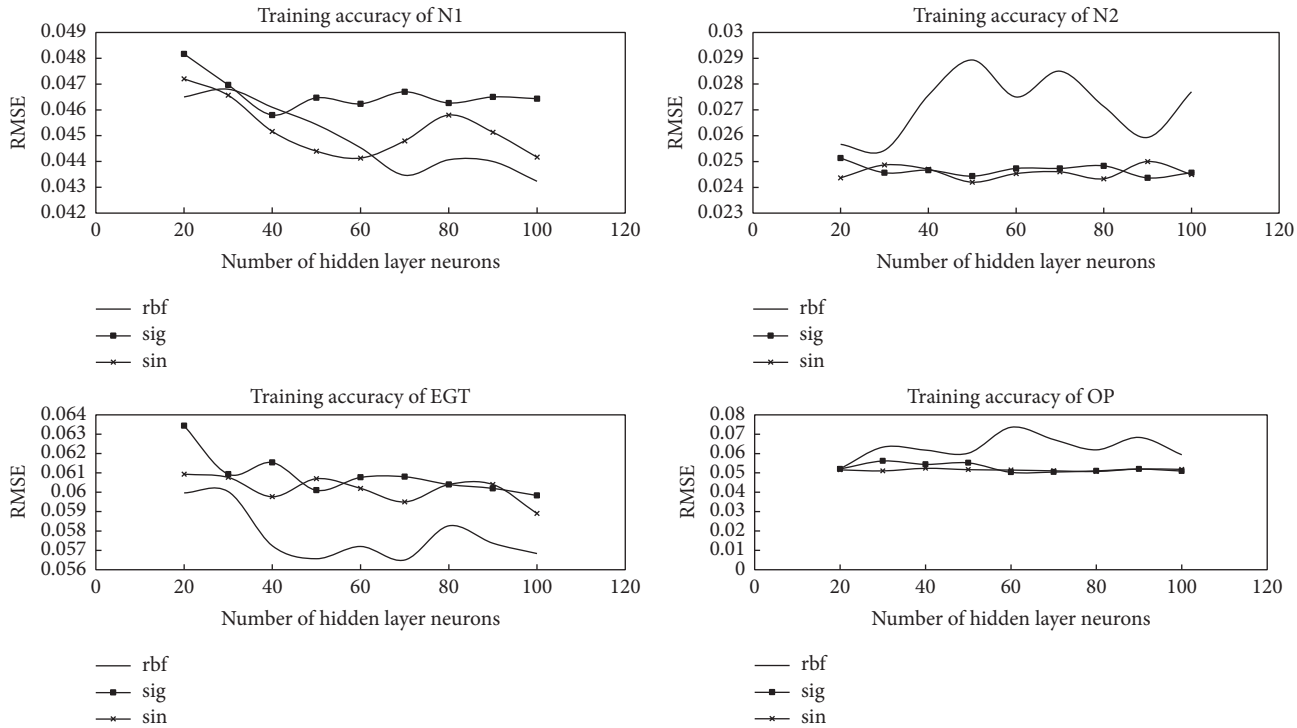


Figure 2: Training accuracy with different activation functions.

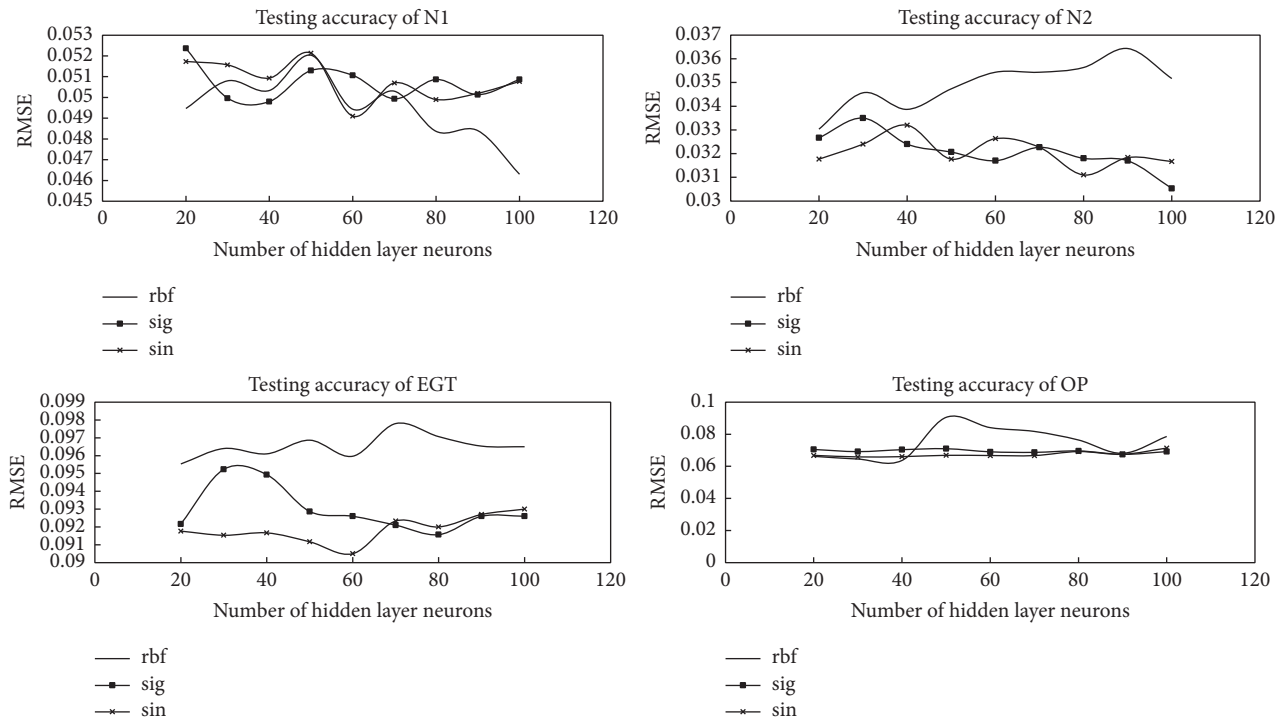


Figure 3: Testing accuracy with different activation functions.

at once. Sometimes, we cannot obtain all the sampling data at one time. When the new data comes, it has to train the model using the whole data again, which may increase its training time. Online Sequential Extreme Learning Machine (OS-ELM) [11] can enable the training data to be trained one-by-one or chunk-by-chunk, which is a better choice for real-time fault diagnosis. In recent years, OS-ELM algorithm has been used in the field of fault diagnosis widely. Yin et al. used OS-ELM algorithm to realize fault diagnosis of rotating machinery [12]. Multiple model classifiers were established, and simulation results illustrated its good ability in classification. A fault diagnosis method based on weighted dissolved extreme learning machine (WELM) is presented for transformer in [13]. WELM algorithm is proposed to deal with data imbalance in dissolved gas analysis data. As an improved algorithm, ImOs-ELM is presented in [14] to realize aeroengine sensor fault diagnosis, which has great advantages in real-time capability and accuracy. However, the aeroengine sensor model is known in advance. Considering that we only have the sensors' output data when the plane flights are under normal condition, and lack of the data when sensors' faults are happening, the achievement of aeroengine sensor fault diagnosis based on OS-ELM algorithm shows a good ability of generalization. In terms of aeroengine sensor data, sensor output data obtained in different flight phases keeps changing greatly. Therefore, weighted dealing with the old and new data can increase accuracy of diagnosis model. For this reason, this paper utilizes WOS-ELM [15, 16] to build the diagnosis model.

Generally speaking, the fault occurring in sensor can be categorized into two typical forms. One type is "hard fault," which shows obvious fault symptom and may cause substantial losses, severe safety, significant casualties, and so on. Due to its obvious fault symptom, it is relatively easy to be detected and diagnosed. The other is "soft fault," also named incipient fault, whose fault symptom is very small and the system has not been affected at the beginning, and the fault symptom gradually becomes obvious as time goes on, ultimately causing meltdowns. For "soft fault," if it could be detected and diagnosed as early as possible, subsequent safety risk, system damage, and casualties would be avoided. However, it is difficult to identify these weak fault symptoms from the noises to detect and diagnose "soft fault" at the beginning. Therefore, it is more challenging to detect and diagnose "soft fault" in practical applications.

Considering the nonlinear characteristics of aeroengine sensors' relation and the advantage of ELM, this paper introduces OS-ELM to detect and diagnose sensor faults. To ensure the modeling accuracy, WOS-ELM algorithm, assigning different weights to the old and new data, is introduced to update the training model and promotes training model closer to the real status of the aeroengine. To improve the detection and diagnosis precision and timeliness and realize multisensor composite faults, a double redundancy strategy, that is, spatial redundancy and time redundancy, is utilized. Combining ELM and double redundancy strategy, a WOS-ELM-based double redundancy fault diagnosis and reconstruction approach is presented to detect and diagnose

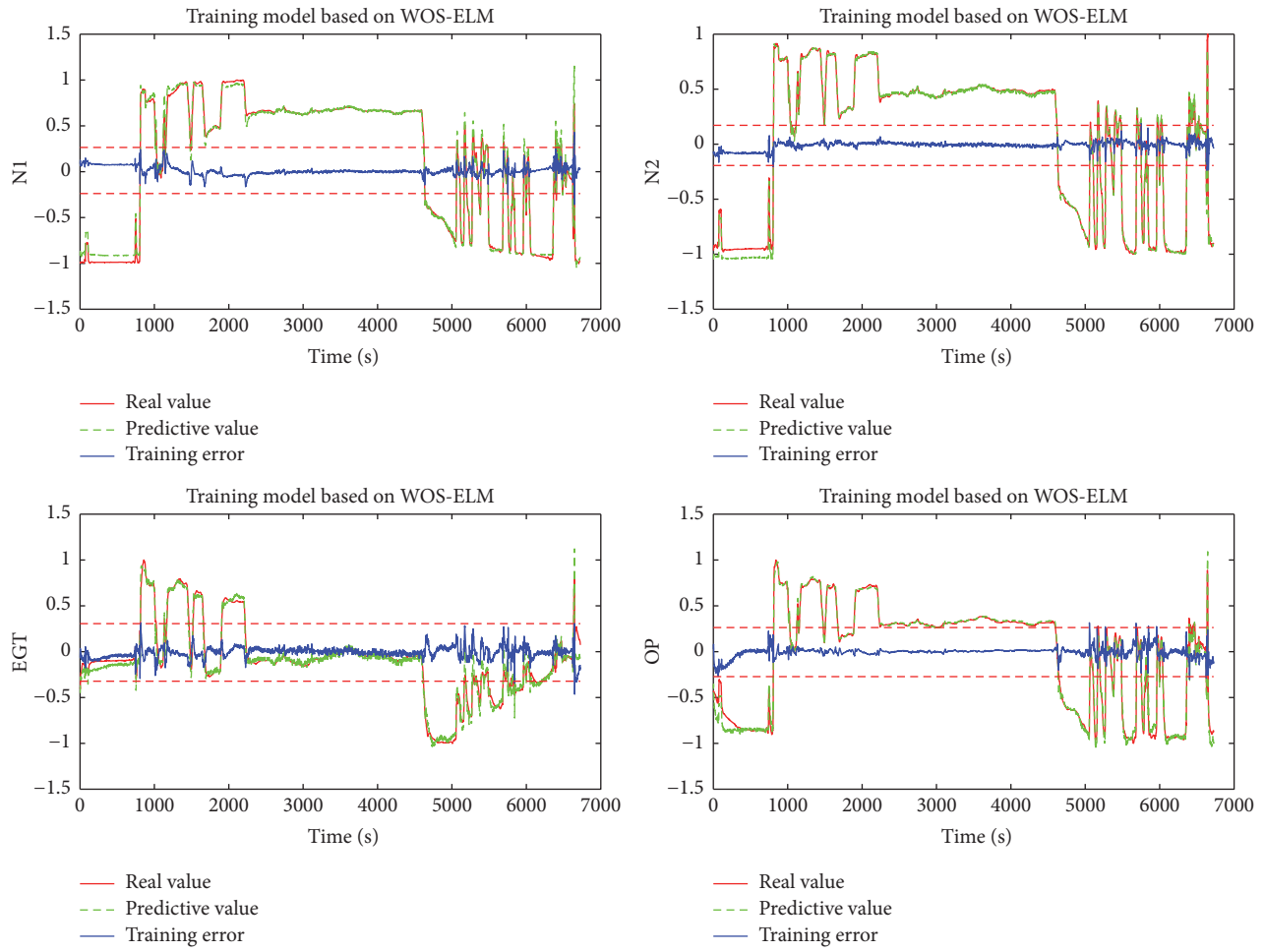


Figure 4: Training results for sensors under normal state.

both "hard fault" and "soft fault" for the aeroengine sensor. The main contributions of this paper include the following:

(1) A WOS-ELM-based double redundancy fault diagnosis and reconstruction approach is proposed for aeroengine sensor.

(2) Intermediate parameters are deduced for the WOS-ELM algorithm.

(3) Through the proposed approach, not only single sensor fault, but also multiple sensor faults can be detected and diagnosed.

The rest of this paper is organized as follows. Section 2 describes Weighted Online Sequential Extreme Learning Machine (WOS-ELM). Section 3 depicts the WOS-ELM-based double redundancy fault diagnosis and reconstruction (FDR) method for aeroengine sensor. Section 4 discusses the simulation result of aeroengine sensor fault diagnosis and reconstruction. Section 5 draws conclusions for the whole paper.

2. WOS-ELM

Considering that the training data may arrive chunk-by-chunk or one-by-one (a special case of chunk), Online Sequential Extreme Learning Machine (OS-ELM) was proposed by Liang et al. [11] for single-hidden layer feedforward neural network (SLFN) with additives and RBF hidden nodes. Given a series of arbitrary independent samples $(x_i, y_i) \in R^n \times R^m$, $i = 1, 2, \ldots, N$, the actual outputs of the SLFN with L hidden nodes for these N samples are given as follows [11].

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i G(\omega_i, b_i, x_j) = y_j, \quad j = 1, 2, \ldots, N, \tag{1}$$

where ω_i is the input weight vector connecting the input layer neuron to the ith hidden neuron, b_i is the bias of the ith hidden node, β_i is the output weight vector connecting the ith hidden neuron to the output layer neuron, $G(\omega_i, b_i, x_j)$ is the output of the ith hidden neuron with respect to the input x_j, and L is the number of hidden nodes.

OS-ELM neglects the time factor when dealing with the training data. It means that old and new data are thought to be the same in reflecting the operating statuses of some equipment (i.e., aeroengine sensors in this paper). WOS-ELM, considering the time factor and endowing different weights to data sampled at different time period, is an improvement of OS-ELM. The same as OS-ELM algorithm,

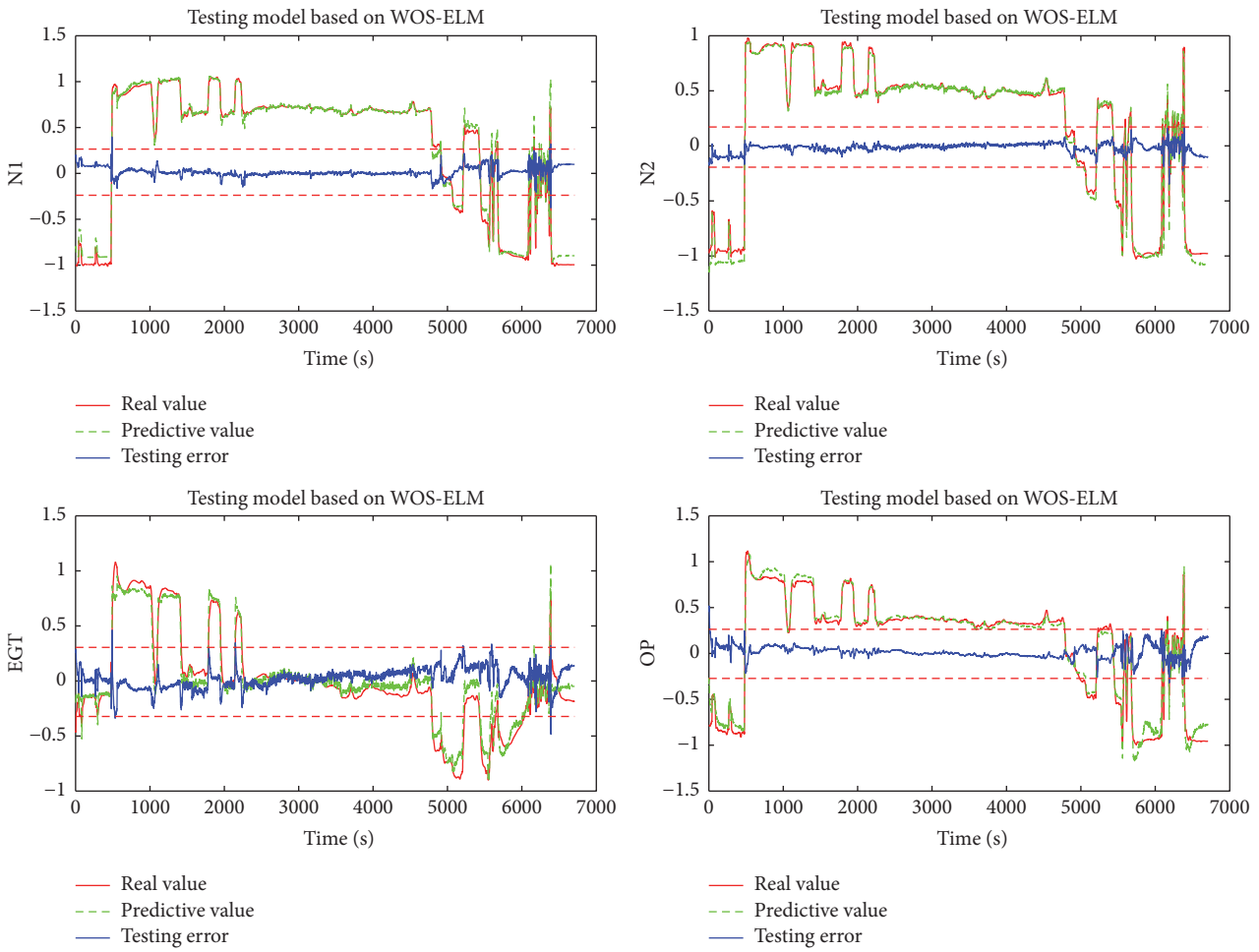


Figure 5: Testing results for sensors under normal state.

WOS-ELM can also be divided into two phases, that is, initialization phase and sequential learning phase.

2.1. Initialization Phase. Using a small chunk of training data $\aleph_0 = \{(x_i, y_i)\}_{i=1}^{N_0}$ from the given training set $\aleph = \{(x_i, y_i) \mid x_i \in R^n, y_i \in R^m, i = 1, \ldots, N\}$ for the initial learning, N_0 respects the number of initial data, and $N_0 \geq L$. The initialization phase of WOS-ELM includes four steps as follows.

Step 1. Randomly assign input weight ω_i and bias of hidden neuron b_i $(i = 1, \ldots, L)$ and give activation function and number of hidden nodes L. Normally, different number of hidden nodes and activation function show different precision, which will be proved in the following simulation.

Step 2. Calculate the initial hidden layer output matrix H_0.

$$H_0 = \begin{bmatrix} G(\omega_1, b_1, x_1) & \cdots & G(\omega_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(\omega_1, b_1, x_{N_0}) & \cdots & G(\omega_L, b_L, x_{N_0}) \end{bmatrix}_{N_0 \times L}. \quad (2)$$

Step 3. Deduce the initial output weight $\beta^{(0)}$.

$$\beta^{(0)} = P_0 H_0^T Y_0, \quad (3)$$

where $P_0 = (H_0^T H_0)^{-1}$, and $Y_0 = [y_1 \cdots y_{N_0}]^T_{N_0 \times m}$.

Step 4. Set $K_k = H_k^T H_k = P_k^{-1}$ $(k = 0)$.

2.2. Sequential Learning Phase. For the $(k+1)$th chunk of new data, $\aleph_{k+1} = \{(x_i, y_i)\}_{i=(\sum_{j=0}^{k} N_j)+1}^{\sum_{j=0}^{k+1} N_j}$ (N_{k+1} stands for the number of data in the $(k+1)$th chunk); the sequential learning phase is as in the following steps.

Step 1. Calculate the $(k+1)$th partial hidden layer output matrix H_{k+1}.

$$H_{k+1} = \begin{bmatrix} G\left(\omega_1, b_1, x_{(\sum_{j=0}^{k+1} N_j)+1}\right) & \cdots & G\left(\omega_L, b_L, x_{(\sum_{j=0}^{k+1} N_j)+1}\right) \\ \vdots & \ddots & \vdots \\ G\left(\omega_1, b_1, x_{\sum_{j=0}^{k+1} N_j}\right) & \cdots & G\left(\omega_L, b_L, x_{\sum_{j=0}^{k+1} N_j}\right) \end{bmatrix}_{N_{k+1} \times L}. \quad (4)$$

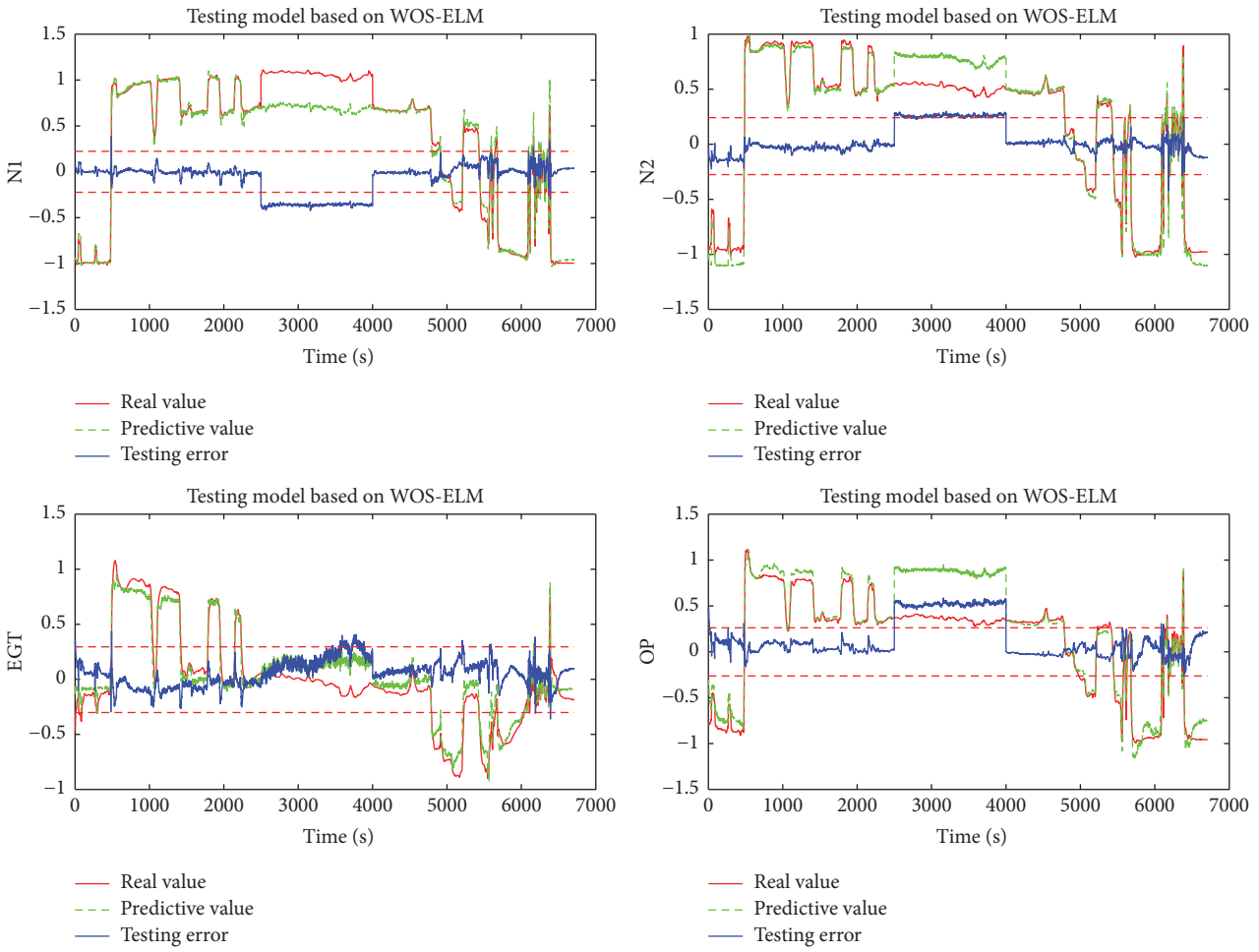


Figure 6: Results of spatial redundancy submodels for step fault of N1.

Step 2. Calculate the $(k+1)$th output weight $\beta^{(k+1)}$.

The target output can be expressed as

$$Y_{k+1} = \left[y_{(\sum_{j=0}^{k} N_j)+1} \cdots y_{\sum_{j=0}^{k+1} N_j} \right]^T_{N_{k+1} \times m}. \tag{5}$$

In terms of OS-ELM algorithm, K_{k+1} can be expressed as follows.

$$K_{k+1} = \begin{bmatrix} H_k \\ H_{k+1} \end{bmatrix}^T \begin{bmatrix} H_k \\ H_{k+1} \end{bmatrix} = K_k + H_{k+1}^T H_{k+1}. \tag{6}$$

From (6), it can be seen that OS-ELM algorithm handles new data and old data with equal weight, which ignores the influence of time factor. Considering the fact that sensors' output data often changes continuously and greatly when the plane is in the different flight phases, this paper utilizes weighted method to deal with the input data to improve the accuracy of the diagnosis model. Thus, K_{k+1} in the WOS-ELM can be described as follows.

$$K_{k+1} = \begin{bmatrix} \lambda H_k \\ H_{k+1} \end{bmatrix}^T \begin{bmatrix} \lambda H_k \\ H_{k+1} \end{bmatrix} = \lambda^2 K_k + H_{k+1}^T H_{k+1}, \tag{7}$$

where λ stands for weight of the old data, whose value can be determined according to the root mean square error (RMSE) at previous step. RMSE can be written as

$$\text{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} d_i^2}, \tag{8}$$

where d_i is the deviation between the predicted values of the model and the real values. Define RMSE $\in (a, b)$ (a, b can be set according to practical application). When RMSE $< a$, $\lambda = 1.005$; when RMSE $> b$, $\lambda = 0.995$; else $\lambda = 1$. Then, K_{k+1}^{-1} can be deduced as

$$\begin{aligned} K_{k+1}^{-1} &= \left(\lambda^2 K_k + H_{k+1}^T H_{k+1}\right)^{-1} = \frac{1}{\lambda^2}\left(K_k^{-1}\right. \\ &\left. - K_k^{-1} H_{k+1}^T \left(\lambda^2 I + H_{k+1} K_k^{-1} H_{k+1}^T\right)^{-1} H_{k+1} K_k^{-1}\right). \end{aligned} \tag{9}$$

Considering that $K_{k+1}^{-1} = P_{k+1}$ and $K_k^{-1} = P_k$, we can derive

$$P_{k+1} = \frac{1}{\lambda^2}\left(P_k - \frac{Q_k Q_k^T}{\lambda^2 I + H_{k+1} Q_k}\right), \tag{10}$$

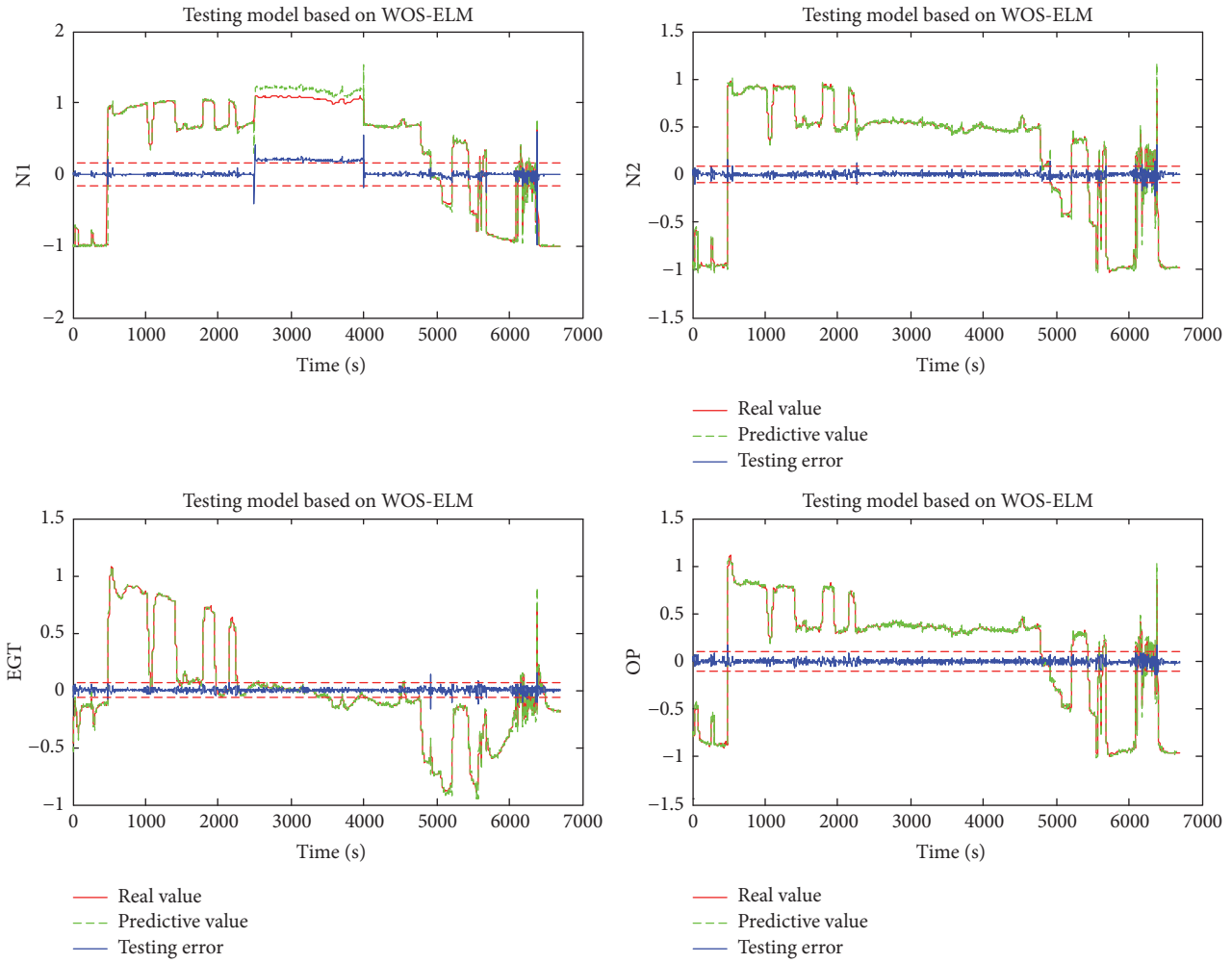


Figure 7: Results of time redundancy submodels for step fault of N1.

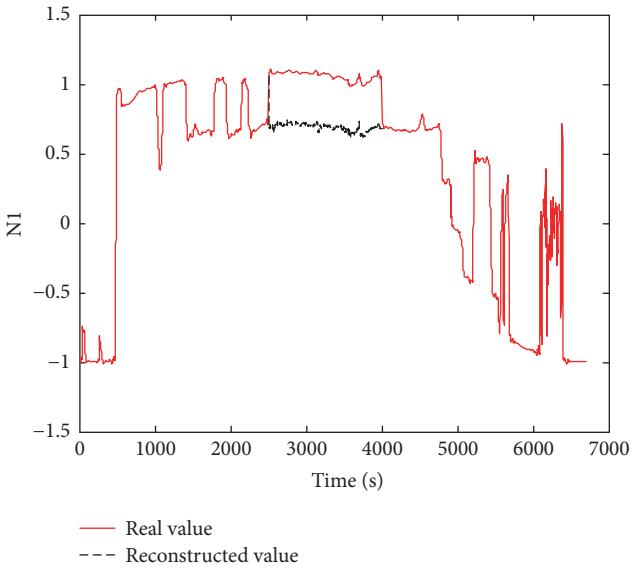


Figure 8: Reconstructed value for N1.

where $Q_k = P_k H_{k+1}^T$. Finally, $\beta^{(k+1)}$ can be written as

$$\beta^{(k+1)} = \beta^{(k)} + P_{k+1} H_{k+1}^T \left(Y_{k+1} - H_{k+1}\beta^{(k)}\right). \quad (11)$$

Step 3. Set $k = k + 1$, go to Step 1 in the sequential learning phase, and continue sequential learning until finishing data learning.

3. WOS-ELM-Based Double Redundancy FDR Method

In this section, the proposed WOS-ELM is combined with double redundancy strategy to realize multiple sensors' FDR. Double redundancy refers to spatial redundancy and time redundancy. The former utilizes the redundancy information among all the sensors, while the latter uses the redundancy information of each sensor over consecutive time. In this paper, we set up two submodels for each sensor and use them to realize sensor fault diagnosis. The whole fault diagnosis process of the nth sensor is shown in Figure 1. From Figure 1,

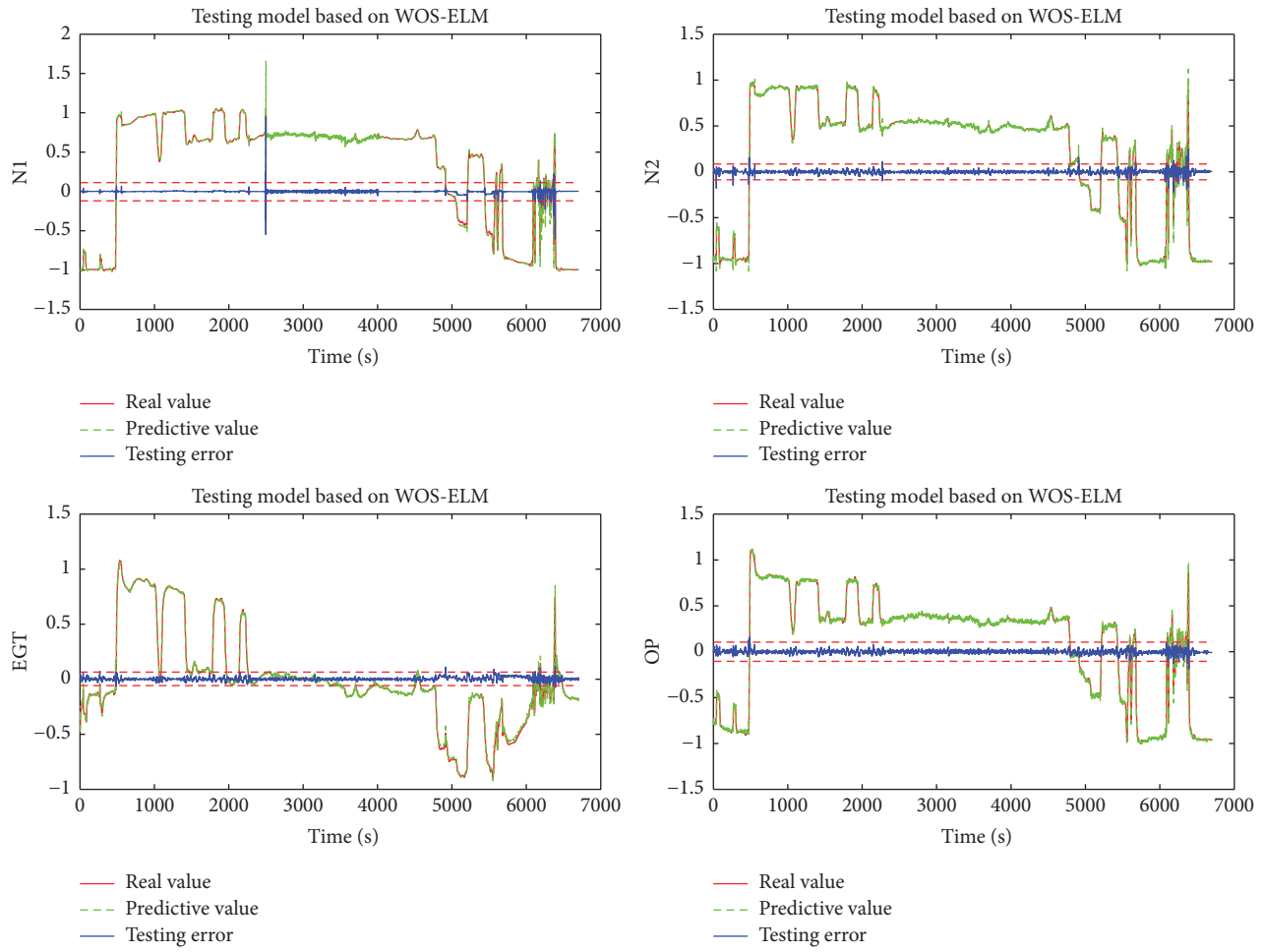


Figure 9: Time redundancy results after reconstruction.

it can be seen that the whole diagnosis process contains two stages, namely, offline modeling stage and online diagnosis stage.

3.1. Offline Modeling. The offline modeling stage of the proposed WOS-ELM-based double redundancy FDR is illustrated above the dashed line in Figure 1. The detailed description of the offline modeling stage is as follows.

(1) Before training the WOS-ELM model, wavelet denoising method is utilized to deal with the training data sampled from sensors. In this way, the training data is obtained.

(2) Maximum and minimum normalization method is used to normalize the denoised training data, and the data is standardized to be in $[-1, 1]$.

(3) Input the normalized training data into spatial redundancy submodel and time redundancy submodel. For each spatial redundancy submodel, the input is the output of all sensors except the nth sensor at time t, and the output is the nth sensor's output at time t. For each time redundancy submodel, the input is the output of the nth sensor from time $t - p$ to $t - 1$, and the output is the nth sensor's output at time t. Then, we can get n well-trained spatial redundancy submodels and n time redundancy submodels.

3.2. Online Diagnosis. When implementing the online monitoring, follow the following steps to realize the online detection of sensors' fault after receiving the present sensor data $x_{\text{new}}(1 \times n)$ in real time.

(1) The data sampled from the aeroengine sensors is firstly preprocessed as that at the offline modeling stage.

(2) Then, the preprocessed data is input into the well-trained spatial redundancy submodels and time redundancy submodels. These submodels' outputs are the evaluated value of each sensor at current time, from which we can calculate the residuals by comparing with its real value.

(3) Thresholds are used to determine whether there are sensors' faults. Here, thresholds, which have great influence on the accuracy of fault detection and diagnosis, should be set according to practical applications. In this paper, the residual sequence's standard variance while the system is under the normal situation is used to determine the thresholds. Generally, choose m times of the standard variance. Assuming that the meaning of the residuals sequence is μ, the standard variance is σ; then the range of the threshold can be set as $[\mu - m\sigma, \mu + m\sigma]$ [14].

(4) As shown in Figure 1, when the spatial redundancy residuals and the time redundancy residuals both exceed their

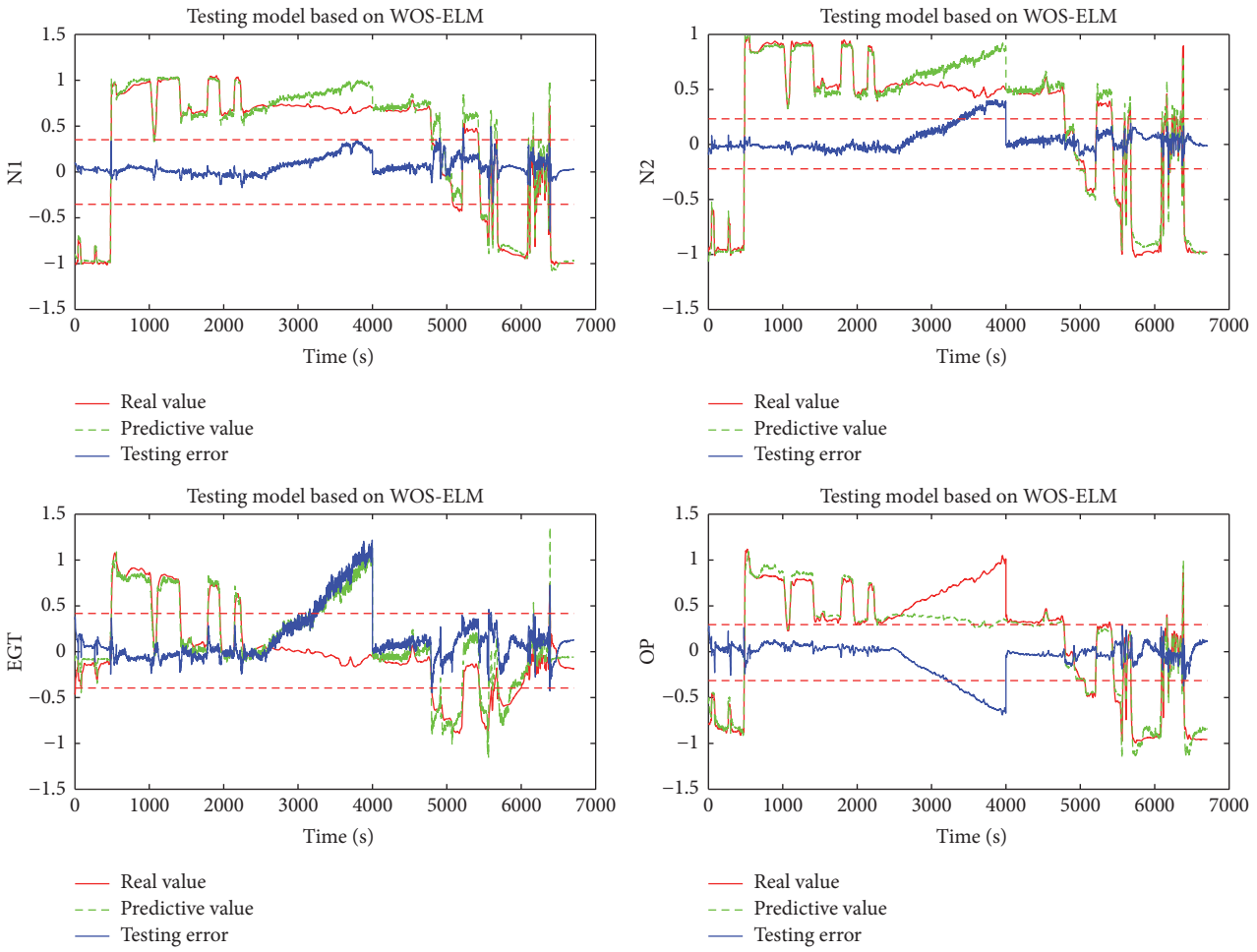


Figure 10: Spatial redundancy results for soft fault.

thresholds, a hard fault of the nth sensor occurs. Suppose the nth sensor encounters a hard fault. On the one hand, as the input of each spatial redundancy submodel is the other sensors' signals, which are all normal, the nth sensor's evaluated value is correct, while the other spatial redundancy submodels' outputs are not. Since the nth sensor's real value shows fault signals, and the other sensors do not, all the residuals exceed the thresholds. On the other hand, the time redundancy submodels could give us a definite answer of which sensor is running with fault. The sensor, whose time redundancy residual exceeds the threshold, is the faulty one. Furthermore, if there are more than one sensor falling in hard faults, all the spatial redundancy submodels' outputs exceed the corresponding thresholds. For time redundancy submodels, only the faulty sensor-related time redundancy residuals exceed the corresponding thresholds.

(5) If neither the spatial redundancy residuals nor the time redundancy residuals exceed the thresholds, but the change rate of the spatial redundancy residual exceeds the threshold for a long period of time, a soft fault may occur to the related sensor. Otherwise, it may be treated as a false alarm due to the change of flight phases. Considering that the change rate of residual is small for short period, the time interval is set to be 200 s, and the expression is shown as

$$c(i) = \varepsilon(i + 200) - \varepsilon(i), \tag{12}$$

where $\varepsilon(i)$ is the spatial redundancy residuals at time i.

If there are more than one sensor falling in soft fault, then the faulty sensor-related change rate of residual would exceed its corresponding thresholds.

(6) When the spatial redundancy residuals, the time redundancy residuals, and the change rate of spatial redundancy residual do not exceed the thresholds, it seems to be operated under normal status and there is no sensor fault.

During the FDR process, if the sensors' data are normal, the right data can be used as training data to make the spatial redundancy submodels and the time redundancy submodels more accurate and more suitable in the current flight phase. The trained process is depicted in Section 2.2.

After the faulty sensors have been detected and diagnosed, the faulty sensor-related time redundancy submodels' outputs can be used to reconstruct the fault signal. That is to say, we used trouble-free or reconstructed sensors' signals to update the training models, which make the diagnosis models

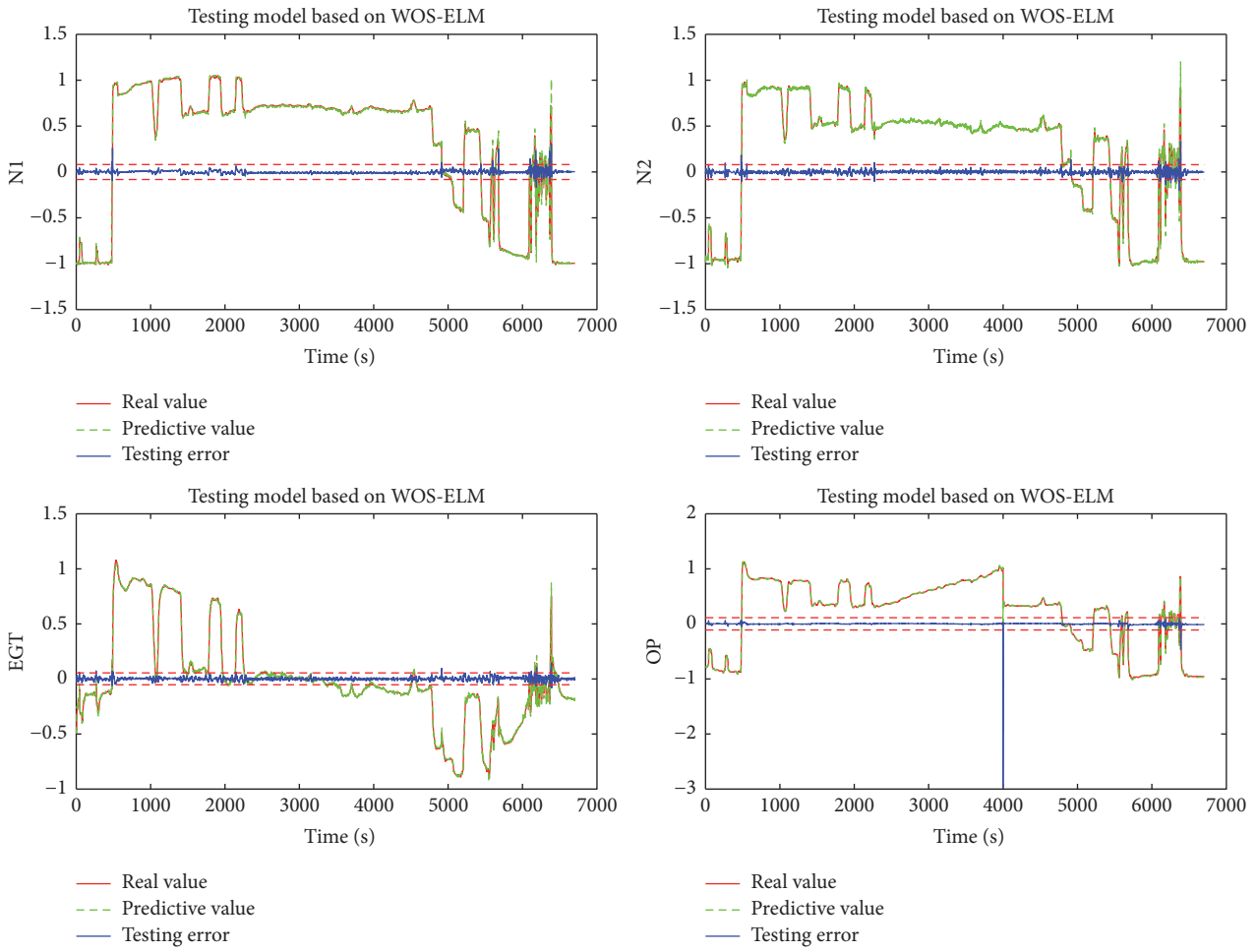


Figure 11: Time redundancy results for soft fault of OP.

much closer to the actual sensor operating environment. In order to ensure the diagnosis preciseness and system safety, the weight for reconstructed sensors' data should be appropriately reduced.

4. Illustration and Discussion

In this section, the data is acquired in real time from the sensors when the plane is in flight status. The simulation tool is MATLAB (R2010a). Four sensors' outputs are chosen in this paper, including low pressure rotor speed (N1), high pressure rotor speed (N2), engine exhaust temperature (EGT), and oil press (OP). The sampling time interval is 1 second. The number of the training data in the offline modeling is 6700, and the number of the testing data in the online modeling is about 6700.

4.1. Parameter Selection. In order to achieve better training effect, the activation function and the number of hidden layer neurons should be selected properly. To build the spatial redundancy submodels, this paper chooses RBF, sin, and sig as activation function in WOS-ELM algorithm, and set N_0 = 800 and N_{k+1} = 6 based on lots of test. The training accuracy and testing accuracy with different activation functions can be seen in Figures 2 and 3.

Table 1: Activation function and the number of hidden nodes for each sensor.

Sensor	Type of activation function	Number of hidden layer neurons
N1	RBF	70
N2	sin	50
EGT	sin	60
OP	sin	30

In Figures 2 and 3, the x-axis is the number of hidden nodes and the y-axis is the RMSE. From Figures 2 and 3, we can see that different activation functions have different influences on the accuracy of training process and testing process. In order to achieve higher modeling accuracy, the activation function and the number of hidden nodes for each sensor are selected according to a large number of trials, whose final results are shown in Table 1.

The training and testing results of these four sensors under normal state are shown in Figures 4 and 5.

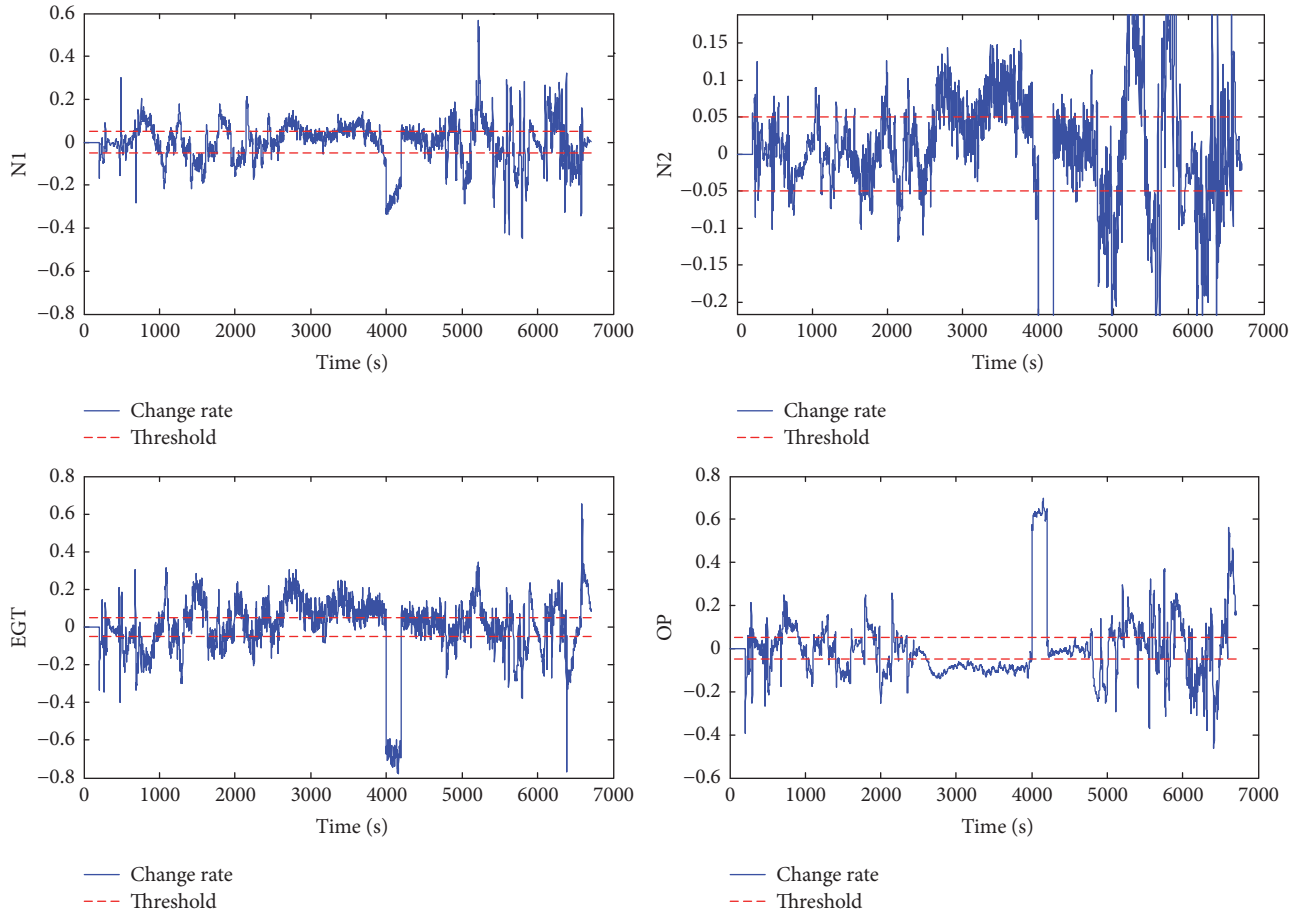


Figure 12: Change rate of spatial redundancy residuals.

Figure 4 shows that WOS-ELM algorithm has a good ability of linear fitting, and it can track sensors' output signal perfectly. That is to say, using this algorithm can get well-trained submodels.

Figure 5 shows that good effect can be received if these well-trained models are applied to evaluate the testing sensors' outputs, which is well used in the sensors' fault diagnosis process.

4.2. Hard Fault Diagnosis. Hard fault refers to instantaneous bigger step or mutations of the normal sensors' outputs, including step fault or pulse fault of sensors. Due to the fact that the real aeroengine sensors' faulty data is difficult to obtain for many reasons, for example, protection measures would be implemented when fault occurs, this paper adds a constant gain α to the original normal N1 signal to simulate step fault. In this paper, α is set to be 1.15 from 2500 s to 4000 s. The results of the spatial redundancy submodels are shown in Figure 6.

Figure 6 shows that when the step fault occurs to N1, residuals of the four sensors all exceed their corresponding thresholds from 2500 s to time 4000 s. Therefore, it cannot distinguish faulty sensor from the spatial redundancy results. As described in Section 3.2, the time redundancy models are introduced to determine which sensor is faulty. The results of the time redundancy submodels are shown in Figure 7.

From Figure 7, it is easy to see that the residual sequence of N1 sensor exceeds its threshold at 2500 s, while the residuals of other sensors are within their corresponding thresholds. Combining spatial redundancy model and time redundancy model, we can conclude that N1 sensor encounters a hard fault at the 2500 s. When sensors' faults are diagnosed, reconstruction should be done. The reconstruction of N1 can be seen in Figure 8.

Figures 8 and 9 show that WOS-ELM algorithm can be used to reconstruct fault signal, and the reconstructed signal can also be input into the well-trained models for further training.

4.3. Soft Fault Diagnosis. Soft fault refers to slow changing of the sensors' outputs, such as drift fault. In general, soft fault is mainly caused by aging and zero drift of the aeroengine sensor. In this paper, we can add a constant gain α and a constant deviation γ for OP from 2500 s to 4000 s. The results of spatial redundancy submodels are shown in Figure 10 (α = 0.03, γ = 0.05).

Compared to hard fault, it is not easy to diagnose soft fault due to tiny failure symptom at the early stage. In Figure 10,

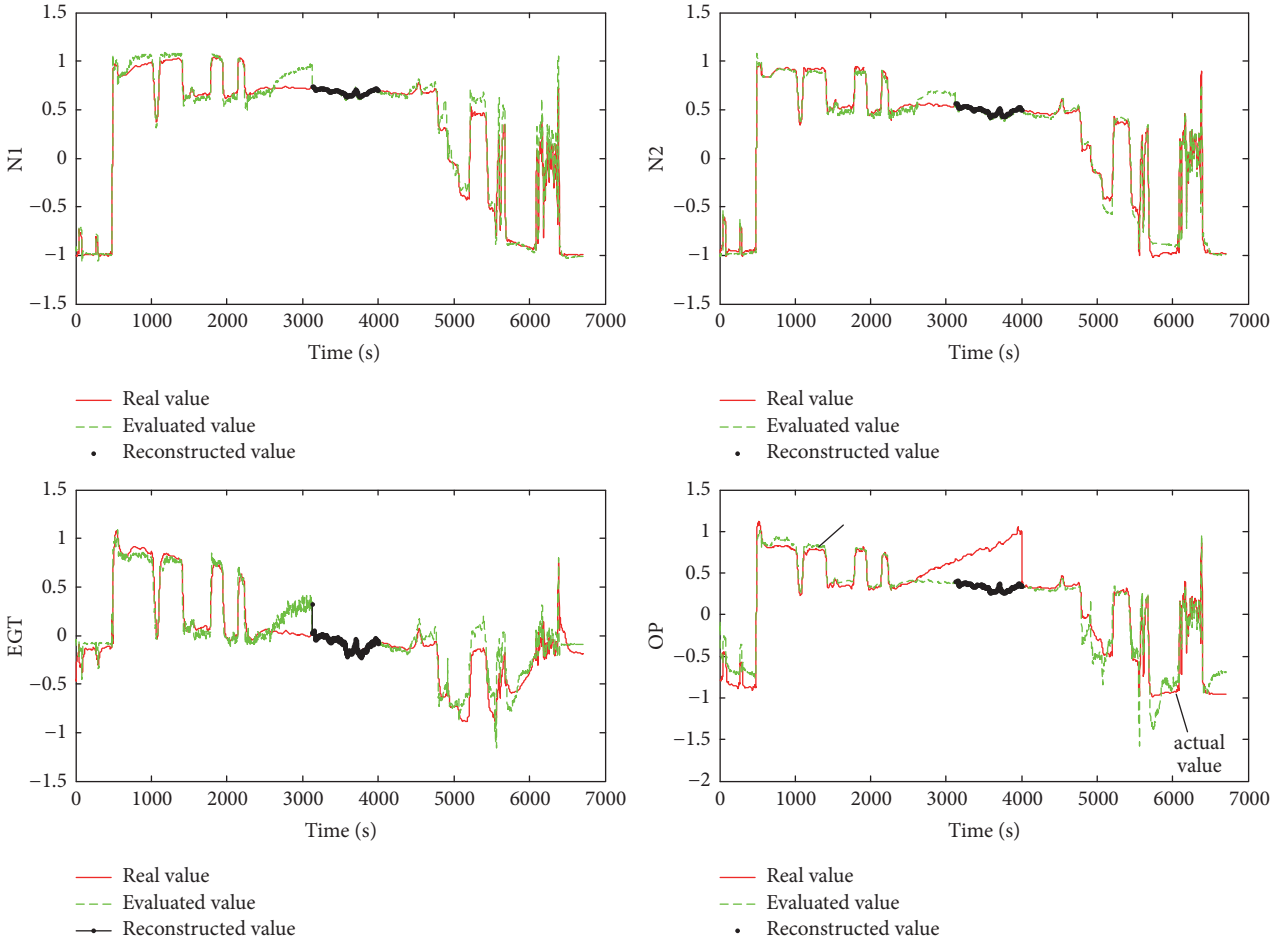


Figure 13: Reconstruction results.

it can be seen that all the spatial residuals are within the thresholds ranges until around 3300 s. After around 3300 s, all residuals exceed their corresponding thresholds except N1 (whose residual keeps growth and is near the threshold), which implies that some faults may occur. However, it is difficult to make decision of which sensor is faulty.

Turn to look at the time redundancy submodels' results, which are shown in Figure 11. From Figure 11, it is also difficult to determine the faulty sensor. As described in Section 3.2, we observe the change rate of the spatial redundancy residual, which is shown in Figure 12. In Figure 12, the thresholds are set to be −0.05 and 0.05. Considering that the development of soft fault would go through a long period, there a soft fault occurs only when the duration of change rate exceeds the upper threshold or the lower threshold lasts for a long period. In this paper, the duration exceeding the threshold is set as 300. Then, we can get the period of every sensor exceeding the thresholds for Figure 12, shown in Table 2.

In Table 2, "—" stands for never exceeding the threshold. According to Figure 12 and Table 2, we can confirm that the OP sensor has a soft fault at time 3125 s (2625 + 200 + 300). Here, 200 stands for the time lag for detecting the residual errors, and 300 stands for the duration time for residual change amplitude exceeding the thresholds. When detecting the soft fault, it may have continued for a period. However, considering that it had little effect on the flight of aircraft at the beginning of the soft fault, time delay for detecting soft fault is accepted.

Table 2: The period of every sensor exceeding the thresholds.

Sensor	Exceed the upper threshold period	Exceed the lower threshold period
N1	—	—
N2	—	—
EGT	—	—
OP	—	2625 s~3923 s

Figure 13 illustrates the reconstructed results of the four sensors when detecting the soft fault of OP. After reconstruction, the results of its time redundancy submodel can be shown as in Figure 14. The change of the spatial redundancy residual is more efficiently applied to sensors' soft fault detection.

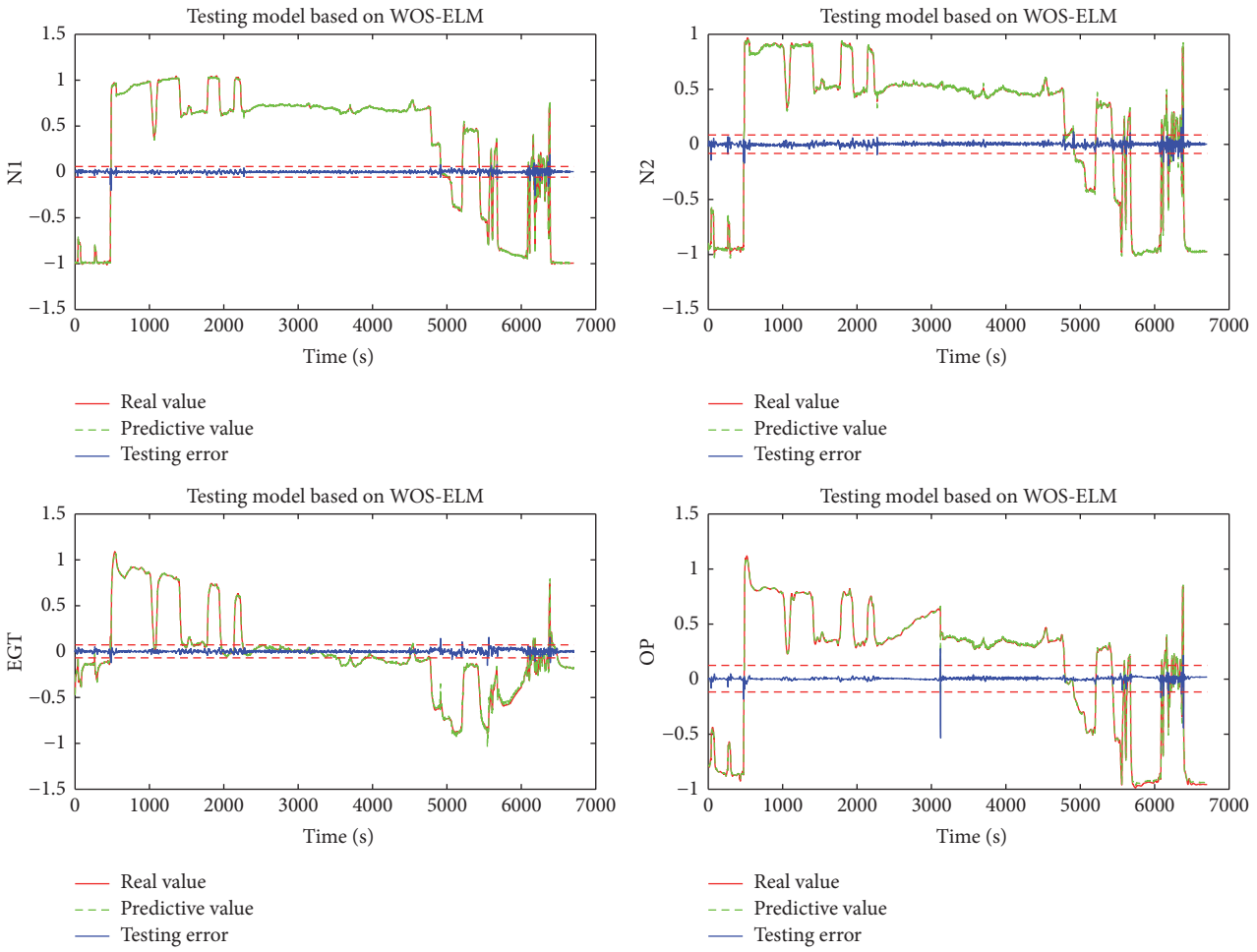


Figure 14: Time redundancy results after reconstruction.

5. Conclusion

In this paper, a double redundancy diagnosis approach based on WOS-ELM is proposed to diagnose sensor fault for the aeroengine. The proposed method contains two stages, that is, offline modeling stage and online detection stage. At the offline modeling stage, two series of models, namely, spatial redundancy submodels and time redundancy submodels, are trained to evaluate the sensors' value for further online detection, diagnosis, and reconstruction. At the online diagnosis stage, new sampled data is preprocessed to be input into these well-trained submodels. According to the outputs of these submodels and the practical sensors' value, several residuals are generated to diagnose sensors' faults. After that, WOS-ELM, which assigns different weights to the old and new data, is proposed to adjust parameters of ELM to improve the modeling accuracy. Simulation results show that the proposed method can be well used in sensors fault diagnosis.

The presented WOS-ELM-based double redundancy fault diagnosis and reconstruction method can also be extended using practical sampled sensors' outputs. Furthermore, the proposed method can be extended to other similar application fields.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Science Foundation for Young Scientists of China (Grant no. 61703406), the project 973 (Grant no. 2014CB744904), civil aviation science and technology major projects (Project no. MHRD20130112), Joint Funds of the National Natural Science Foundation of China and Civil Aviation Administration of China Key Project (Grant no. U1533201), and the Fundamental Research Funds for the Central Universities (Project no. 3122014C010, 3122015C011).

References

[1] K. Cai L, Y. Sun F, and W. Yao W, "Fault diagnosis and adaptive reconfiguration control for sensors in aeroengine," *DIanguang YU Kongzhi*, vol. 16, no. 6, pp. 57–61, 2009.

[2] B. Zhao W, "Research on aircraft engine sensor fault diagnosis and signal reconstruction," *Nanjing University of Aeronautics and Astronautics*, pp. 34–42, 2011.

[3] K. Takahisa and S. Donald L, "Application of a bank of Kalman filters for aircraft engine fault diagnosis," *Atlanta: ASME Turbo Expo Power for Land, Sea, and Air*, pp. 1–10, 2003.

[4] X. Lishuang, C. Tao, and D. Fang, "Sensor fault diagnosis based on least squares support vector machine online prediction," in *Proceedings of the 2011 IEEE 5th International Conference on Robotics, Automation and Mechatronics, (RAM '11)*, pp. 275–279, Qingdao, China, September 2011.

[5] S. Duan, Q. Li, and Y. Zhao, "Fault diagnosis for sensors of aero-engine based on improved least squares support vector regression," in *Proceedings of the 2011 8th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2011, Jointly with the 2011 7th International Conference on Natural Computation, (ICNC '11)*, pp. 1962–1966, Shanghai, China, July 2011.

[6] H. Zhang, T. Chen, and W. Li, "Abrupt sensor fault diagnosis based on wavelet network," in *Proceedings of the 2006 IEEE International Conference on Information Acquisition, (ICIA '06)*, pp. 111–116, Weihai, China, August 2006.

[7] X. Wang M and Z. Ren, "a sensor fault diagnosis method research based on wavelet transform and hilbert-huang transform," *Hong Kong: Conference on Measuring Technology and Mechatronics Automation*, pp. 81–84, 2013.

[8] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 985–990, Budapest, Hungary, July 2004.

[9] Y. Sun G, J. Liu Y, and Z. Zhan, "Extreme learning machine-based aircraft engine sensor fault diagnosis," *Sensor and Micro System*, vol. 33, pp. 23–26, 2014.

[10] X. Zhang and L. Wang H, "Have a choice and forgetting mechanism of extreme learning machine in the application of time series prediction," *Acta Physica Sinica*, vol. 60, no. 8, Article ID 080504, 2011.

[11] N. Liang, G. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 17, no. 6, pp. 1411–1423, 2006.

[12] G. Yin, Y. Zhang T, and N. Li Z, "Fault diagnosis method based on online sequential extreme learning machine," *Vibration, Measurement and Diagnosis*, vol. 33, no. 2, pp. 325–329, 2013.

[13] B. Yu J and L. Zhu Y, "Transformer fault diagnosis using weighted extreme learning machine," *Computer Engineering and Design*, vol. 34, pp. 4340–4344, 2013.

[14] B. Y. Li, Q. H. Li, J. K. Wang, and X. H. Huang, "Sensor fault adaptive diagnosis of aero-engines based on ImOS-ELM," in *Hangkong Xuebao*, vol. 34, pp. 2316–2324, 10 edition, 2013.

[15] B. Mirza, Z. Lin, and K.-A. Toh, "Weighted online sequential extreme learning machine for class imbalance learning," *Neural Processing Letters*, vol. 38, no. 3, pp. 465–486, 2013.

[16] B. Mirza, Z. Lin, J. Cao, and X. Lai, "Voting based weighted online sequential extreme learning machine for imbalance multi-class classification," in *Proceedings of the IEEE International Symposium on Circuits and Systems, (ISCAS '15)*, pp. 565–568, Lisbon, Portugal, May 2015.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 3614790, 12 pages
https://doi.org/10.1155/2017/3614790

Research Article

Parameter Selection Method for Support Vector Regression Based on Adaptive Fusion of the Mixed Kernel Function

Hailun Wang[1,2] and Daxing Xu[1,3]

[1]*College of Electrical and Information Engineering, Quzhou University, Quzhou 324000, China*
[2]*Logistics Engineering College, Shanghai Maritime University, Shanghai 200000, China*
[3]*Department of Automation, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Daxing Xu; daxingxu@163.com

Received 30 June 2017; Revised 14 September 2017; Accepted 8 October 2017; Published 2 November 2017

Academic Editor: Yuan Yao

Support vector regression algorithm is widely used in fault diagnosis of rolling bearing. A new model parameter selection method for support vector regression based on adaptive fusion of the mixed kernel function is proposed in this paper. We choose the mixed kernel function as the kernel function of support vector regression. The mixed kernel function of the fusion coefficients, kernel function parameters, and regression parameters are combined together as the parameters of the state vector. Thus, the model selection problem is transformed into a nonlinear system state estimation problem. We use a 5th-degree cubature Kalman filter to estimate the parameters. In this way, we realize the adaptive selection of mixed kernel function weighted coefficients and the kernel parameters, the regression parameters. Compared with a single kernel function, unscented Kalman filter (UKF) support vector regression algorithms, and genetic algorithms, the decision regression function obtained by the proposed method has better generalization ability and higher prediction accuracy.

1. Introduction

The core components and important mechanical structures of mechanical equipment will inevitably be subject to varying degrees of failure with the complex operating conditions and bad working environment. It may cause huge economic losses and casualties when equipment fails. Rolling bearing is widely used in rotating machinery, and its running state directly affects the accuracy, reliability, and life of the machine. Timely and accurate diagnoses for the fault of rolling bearing are helpful to improve the reliability of equipment and reduce the probability of accidents. Due to the highly nonlinear nature between the fault and the characteristic, fault diagnosis methods based on machine learning have been applied more and more widely in the field of automation in recent years [1–3]. Vapnik invented the superior performance of support vector regression method (support vector regression, SVR) [4]. The improved SVR algorithm will greatly improve the accuracy of fault diagnosis.

The key to SVR is the kernel function and parameter selection. There are several methods that have been used to optimize the kernel and select the regression parameters, such as cross validation learning [5, 6], gradient descent learning [7, 8], evolutionary learning [9, 10], and positive semidefinite programming learning [11, 12]. The model of support vector regression and the selection of kernel parameters are relatively few, and it primarily uses the grid search cross validation method and evolutionary method. But the efficiency of these methods is very low because of the exhaustive searching for optimal parameters [13]. When the number of parameters is more than two, it becomes almost impossible to operate, such as the genetic algorithm [14] and particle swarm optimization algorithm [15]. The more serious case is that the evolutionary algorithm may easily fall

into local optimization; that is, it only obtains a suboptimal solution, rather than the optimal solution. Literature [16] provides another way to estimate kernel parameters by using the adjustment of multiparameters of LS-SVM as the parameter identification problem of a nonlinear dynamic system. By using the smoothness of the system model, the kernel parameters and the regression parameters are automatically adjusted by the extended Kalman filter (EKF). Reference [17] puts forward a new method based on the unscented Kalman filter support vector regression model selection method (UKF-SVR) to solve the loss function on the hyperparameters of a nondifferentiable problem. However, considering the accuracy of UKF, it is difficult to meet the needs of practical applications, and when the system dimensions are high, the performance of UKF is significantly degraded, leading to a curse of dimensionality. In 2009, Arasaratnam and Haykin proposed a cubature Kalman filter (CKF) that uses radial quadrature rules for optimization sigma points and weights, enhancing the ability to handle high dimensional nonlinear state estimated accuracy and improving stability. CKF is based on the 3rd degree of radial integration rule, and filtering accuracy is still limited [18]. Recently, [19] proposed a class of high-order spherical and radial integral cubature Kalman filters and proved that the five-order cubature Kalman filter (5th-degree CKF) can obtain high accuracy and high stability with low computational cost.

Although the above actions to resolve specific problems have improved the algorithm, they are single kernel function based on support vector regression. Kernel functions can be divided into local kernel functions and global kernel functions. The learning ability of local kernel functions is stronger, and the extrapolation ability of global kernel functions is stronger. The selection of kernel functions significantly affects the generalization ability of support vector regression. Using only one kernel function will often have limitations. Based on the characteristics of the original kernel function, linear fusion of a local kernel function and a global kernel function constitutes a new kernel function, the mixed kernel function, and the kernel function learned from the advantages of local and global kernel functions that can accurately reflect the actual characteristics of a sample. Hybrid kernel function is introduced to determine the local kernel and the global kernel function fusion coefficient. An appropriate fusion coefficient can better exert the advantages of hybrid kernel function. At present, combined weight values are often determined by experience [20, 21]. Reference [20] refers to the removing of pulmonary nodules which indicates that the initial selection of mixed kernel function coefficients can be based on Gray features, morphological features, and texture features. And the support vector regression based on the constructed hybrid kernel function cannot ensure the best performance.

Because of this, this paper uses the mixed kernel function as the kernel function of the SVM and the 5th-degree CKF as the basic framework, adaptively adjusting the fusion coefficient, kernel parameters, and regression parameters of the mixed kernel function. The remainder of this article is as follows: the first part is the description of the problem; the second part is a review of the typical kernel function; the third part is the parameter selection method of support vector regression based on the adaptive fusion of the mixed kernel function; the fourth part is the analysis of the algorithm; the fifth part is a simulation example; and the sixth part is the summary.

2. Problem Description

2.1. Support Vector Regression. The ultimate goal of the support vector regression is to find a regression function $f: R^D \to R$:

$$y = f(x) = w^T \varphi(x) + b, \tag{1}$$

where $\varphi(x)$ is a function that can map data x from low dimension to high dimensional feature space, w is a weight vector, and b is a numeric value that can be up or down. Standard support vector regression adopts ε−insensitive function. It is assumed that all the training data are fitted with a linear function in the accuracy of ε. The problem is translated into an objective function to optimize the objective function minimization problem as follows [22]:

$$\begin{aligned} \min_{w \in R^n, \xi, \xi^*, b \in R} \quad & \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{l}\left(\xi_i^2 + \xi_i^{*2}\right) \\ \text{s.t.} \quad & w^T\varphi(x_i) + b - y_i \le \varepsilon + \xi_i, \\ & \qquad\qquad i = 1, 2, \ldots, l \\ & y_i - w^T\varphi(x_i) - b \le \varepsilon + \xi_i^*, \\ & \qquad\qquad i = 1, 2, \ldots, l \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1, 2, \ldots, l, \end{aligned} \tag{2}$$

where ξ_i, ξ_i^* is the relaxation factor. When there is an error in fitting, ξ_i, ξ_i^* are greater than 0. If not, ξ_i, ξ_i^* are all equal to 0. The first term of the optimization function further smooths the fitting function to improve generalization. The second item is to reduce the error; when constant $C > 0$, it indicates the extent of the penalty for a sample out of error ε.

The performance of support vector regression is affected by the error penalty parameter C, which is the degree of punishment that is used to process the mistakenly divided sample. C is a tradeoff between the algorithm complexity and degree of mistakenly classified samples. When the value of C is small, it means that the punishment for the empirical error of the original data is small. Machine learning complexity is small, but the experience risk is high. When the value of C is larger, the empirical error penalty is larger, and the experience risk is small. However, this can lead to high computational complexity and poor generalization ability. Therefore, it is very important to choose the appropriate punishment coefficient C for practical problems.

The structure of the support vector regression is shown in Figure 1. Another key factor that affects the performance of the support vector regression is the kernel function and its parameters. The core of the support vector regression algorithm is the introduction of kernel function. The kernel

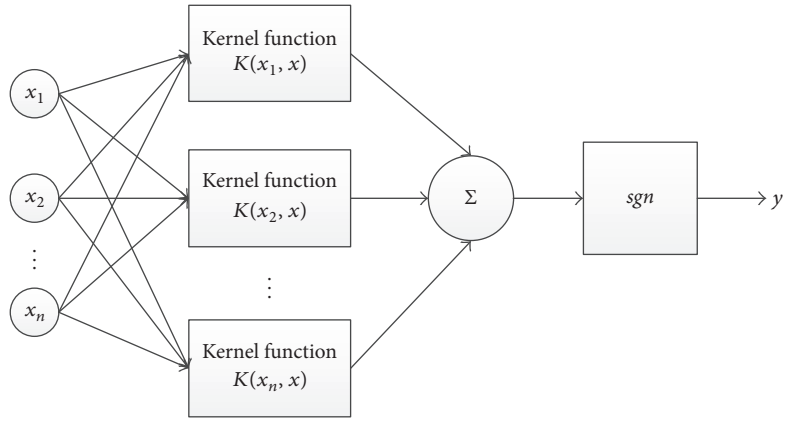


Figure 1: Schematic diagram of support vector regression.

function has two aspects: the construction of the kernel function and the selection of the kernel function model. In fact, the appropriate choice of the model is the key to improve the performance of support vector regression. Model selection determines the kernel function that is more suitable for the data characteristics of the original sample data before training. The kernel function involves two steps: first, determine the type of kernel function, and then select the relevant parameters of the kernel function. Current research is focusing on the choice of the kernel function model. Because different samples may have different characteristics, the construction of the kernel function is more important than the choice of kernel function. Construction of a good kernel function is still a challenging problem.

High performance of support vector regression is difficult to obtain with a single kernel function. The characteristics of the actual sample are complicated and changeable and cannot be completely characterized by the local kernel function or the global kernel function. The mixed kernel function combines the global kernel function and the local kernel function according to a ratio that can accurately reflect the characteristics of the actual sample based on the local and global kernel functions. Therefore, the mixed kernel function has good learning ability and good generalization ability.

2.2. 5th-Degree CKF Principle. Different nonlinear filters have different performance characteristics. The estimation performance of a nonlinear filter is dependent on the specific nonlinear filter type. 5th-degree cubature Kalman filtering algorithm can obtain high accuracy and high stability with low computational cost; thus, it is selected to adaptively adjust the fusion coefficient, kernel parameters, and regression parameters. The general nonlinear system is given as follows:

$$\begin{aligned} x_k &= \mathbf{f}\left(x_{k-1}\right) + w_k \\ z_k &= \mathbf{h}\left(x_k\right) + v_k, \end{aligned} \tag{3}$$

where, x_k is n-dimensional state vector; z_k is m-dimensional observation vector. $\mathbf{f}$ and $\mathbf{h}$ are known nonlinear functions. Both $\{w_k\}$ and $\{v_k\}$ are independent zero mean Gaussian white noise.

Similar to the CKF, the structure of 5th-degree CKF is also divided into two steps: state prediction (time updating) and measurement update. The core of the difference is that the high-order volume Kalman filter uses the phase cubature rule and weight coefficient of high dimension to solve the problem above in introduction. High-order cubature rule satisfies

$$\begin{aligned} I_{U_n} = \overline{w}_{s1} \sum_{j=1}^{n(n-1)} &\left(g_s\left(s_j^+\right) + g_s\left(-s_j^+\right) + g_s\left(s_j^-\right)\right. \\ &\left. + g_s\left(-s_j^-\right)\right) + \overline{w}_{s2} \sum_{j=1}^{n} \left(g_s\left(e_j\right) + g_s\left(-e_j\right)\right), \end{aligned} \tag{4}$$

where e_j is the jth column of the unit vector matrix of n-dimensional space. s_j^+ and s_j^- are the set of points as follows:

$$\begin{aligned} \left\{s_j^+\right\} &= \left\{\sqrt{\frac{1}{2}}\left(\mathbf{e}_k + \mathbf{e}_l\right) : k < l, \quad k < l = 1, 2, \ldots, n\right\} \\ \left\{s_j^-\right\} &= \left\{\sqrt{\frac{1}{2}}\left(\mathbf{e}_k - \mathbf{e}_l\right) : k < l, \quad k < l = 1, 2, \ldots, n\right\}. \end{aligned} \tag{5}$$

The weights $\overline{w}_{s1}$ and $\overline{w}_{s2}$ are

$$\begin{aligned} \overline{w}_{s1} &= \frac{A_n}{n\left(n+2\right)} \\ \overline{w}_{s2} &= \frac{\left(4-n\right)A_n}{\left(2n\left(n+2\right)\right)}, \end{aligned} \tag{6}$$

where $\Gamma(z) = \int_0^\infty \exp(-\lambda)\lambda^{z-1}d\lambda$,

$A_n = 2\sqrt{\pi^n}/\Gamma(n/2)$ is the surface area of the unit sphere. According to the moment matching method, when $n = 2$, the weight is

$$\begin{aligned} w_1 &= \frac{\Gamma\left(n/2\right)}{\left(n+2\right)} \\ w_2 &= \frac{n\Gamma\left(n/2\right)}{\left(2\left(n+2\right)\right)}. \end{aligned} \tag{7}$$

Table 1: Four types of kernel functions and their characteristics.

Kernel function	Characteristics
Linear kernel function: $K(x_i, x_j) = x_i \cdot x_j$.	It is a special case of the kernel function. The parameters are few and the speed is fast [28].
Polynomial kernel function: $K(x_i, x_j) = ((x_i \cdot x_j) + c)^q$, where c and q are the kernel parameters and satisfy the condition $c \geq 0$, $q \in N$.	It is a global kernel function. And it becomes a linear kernel function when $q = 1$. The greater the value of q, the higher the dimension of mapping, and the greater the amount of computation. When q is too large, the complexity of the learning machine is also increased. The promotion ability of the support vector regression is reduced, and it is easy to introduce the phenomenon of overfitting [29].
Gauss kernel function (RBF kernel): $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/\sigma^2)$, where $\sigma > 0$.	RBF kernel function is a strong local kernel function, and the external pushing ability decreases with the increase of parameters. Compared with the general kernel functions, Gauss kernel function only needs to determine a parameter, and constructing the kernel function model is relatively simple. Therefore, RBF kernel function is currently the most widely used one [30].
Sigmoid kernel function: $K(x_i, x_j) = \tanh(\lambda(x_i \cdot x_j') + \varphi)$, where λ, φ are the kernel parameters and satisfy the condition $\lambda > 0$, $\varphi < 0$.	The theoretical basis of support vector regression determines the global optimal value of the support vector regression rather than the local minimum value. It also ensures that it will not cause an overlearning phenomenon because of good generalization ability of unknown samples [31].

This paper first chooses the mixed kernel function as the kernel function of the support vector regression. The problem with constructing the kernel function is to select the fusion coefficients of the local kernel function and the global kernel function, the kernel parameters, and the penalty parameter C of the mixed kernel function. Then, the fusion coefficients are imbedded into the super kernel parameters as the state vector so that the construction of the kernel function and the selection of the parameters of the kernel function can be transformed into a nonlinear filtering problem that can be solved by the 5th-degree CKF. Finally, the adaptive adjustment of the fusion coefficient as well as the estimation of the kernel parameters and the penalty parameters is determined.

3. Review of Classic Kernel Functions

The key to support vector regression is the introduction of kernel function. When the data set is in a low dimensional space, it is usually difficult to separate; when the data set is mapped to a high dimensional space, the formation of new data sets is more easily separated, but the computational effort for this method is huge. The introduction of kernel function reduces computation in the high dimensional feature space directly after the transformation that avoids the "curse of dimensionality" problem. The kernel function is denoted as $K(x_i, x_j)$, where x_i, x_j are the sample data. Four types of kernel functions are widely used in the research and application of support vector regression [21], as shown in Table 1.

Different kernel functions are selected to form different support vector regressions. The linear kernel function, the polynomial kernel function, and the Gauss kernel function have been widely used. The most widely used one is the RBF kernel function with good learning ability. No matter what the conditions are, low dimensional, high dimensional, small samples, and large samples, RBF kernel functions are applicable. RBF kernel function has a wide convergent region, and it is an ideal classification basis function. Sigmoid kernel function from neural networks in practical application is limited. Only under specific conditions (parameters V and C satisfy certain conditions) can the sigmoid kernel function meet the conditions of symmetric and positive definite kernel function. Sigmoid kernel function is proven to have good global classification performance in the application of neural networks, but the classification performance of the application in SVM needs further research [23].

Kernel function skillfully solves the low dimensional vectors that are mapped into a high dimensional curse of dimensionality problem and improves machine learning nonlinear processing ability. However, each kind of kernel function has its own characteristics based on different kernel functions of support vector regression with different generalization abilities. At present, the kernel function is divided into two categories: global kernel function and local kernel function. The local kernel function is effective in extracting the local character of the sample. The value of kernel function is affected by the data points at a very close distance, and the interpolation ability is stronger. Therefore, learning ability is strong. The Gauss kernel function is a local kernel function. Global kernel function is effective in extracting the global characteristics of the samples. Kernel function is only affected by the distance of data points of the value, so generalization ability is strong [24]. Compared with the local kernel function, the global kernel function is weak. The linear kernel function, the polynomial kernel function, and the sigmoid kernel function all are global kernel functions. In short, the learning ability of a local kernel function is strong, and its generalization ability is weak. The generalization ability of a global kernel function is strong, but its learning ability is weak.

4. Parameter Selection Method for Support Vector Regression Based on Adaption Fusion of Mixed Kernel Function

4.1. Mixed Kernel Function. Based on the above analysis, we fuse the local and global kernel functions so that the

mixed kernel function is of both strong learning ability and generalization ability. In this section, we propose a mixed kernel function based on adaptive fusion. The adaption means that the weight of every kernel function in the mixed kernel function is estimated by a filter rather than determined according to past experiences.

Theorem 1. *Denote the local and global kernel functions by $K_{local}(x_i, x_j)$ and $K_{global}(x_i, x_j)$, respectively. Then, the mixed kernel function can be expressed by*

$$K_{mix}\left(x_i, x_j\right) = p1 \cdot K_{local}\left(x_i, x_j\right) + p2 \cdot K_{global}\left(x_i, x_j\right), \tag{8}$$

where $p1$ and $p2$ are the weights of the two kernel functions in the mixed kernel function and $0 \le p1, p2 \le 1$, $p1 + p2 = 1$. The mixed kernel function is still a Mercer kernel.

Proof. Since $K_{\text{local}}(x_i, x_j)$ and $K_{\text{global}}(x_i, x_j)$ are local and global kernel functions, they both satisfy the Mercer condition [23]; that is, for any $\varphi(x) \neq 0$ and $\int \varphi^2(x)dx < 0$, (9) is satisfied.

$$\int K_{\text{local}}\left(x, x'\right)\varphi(x)\varphi\left(x'\right)dxdx' > 0$$
$$\int K_{\text{global}}\left(x, x'\right)\varphi(x)\varphi\left(x'\right)dxdx' > 0. \tag{9}$$

Since $0 \le p1, p2 \le 1$, it can be derived that

$$p1 \cdot \int K_{\text{local}}\left(x, x'\right)\varphi(x)\varphi\left(x'\right)dxdx' + p2 \cdot \int K_{\text{global}}\left(x, x'\right)\cdot\varphi(x)\varphi\left(x'\right)dxdx' > 0; \tag{10}$$

that is, $\int K_{\text{mix}}(x, x')\varphi(x)\varphi(x')dxdx' > 0$. □

It has already been proven that any function can be chosen as a kernel function as long as it satisfies the Mercer condition. Therefore, mixed kernel function (8) can be chosen as a kernel function since it satisfies the Mercer condition. The mixed kernel function is the convex combination of the local and global kernel functions. The introduction of the mixed kernel function eliminates the deficiencies in using a single global or local kernel function.

When $p1 = 0$ or $p2 = 0$, the mixed kernel function becomes a single kernel function. The model selection of single kernel function based support vector regression only concerns the selection of the internal parameters of the kernel function. However, the model selection of mixed kernel function based support vector regression concerns not only the selection of the internal parameters of both the local and global kernel functions but also the fusion coefficients of the two kernel functions to make sure that the performance of the mixed kernel function based support vector regression is best. Before we train the support vector regression, we need to determine the weighed fusion coefficients. The coefficients $p1$ and $p2$ in (6) are usually determined by past experiences. Since the mixed kernel function does not describe the properties of the training samples very well, regression forecast performance will be degraded. Currently, there is no analytical method for the selection of fusion coefficients. The fusion coefficients are usually selected according to experience and it is difficult to estimate the coefficients on line.

4.2. The Establishment of Parameter Filter Model

4.2.1. Mixed Kernel Function Based Support Vector Regression. Differing from the support vector regression based on a single kernel function, the support vector regression based on a fused kernel function utilizes a fused kernel function containing both local and global kernel functions; that is, $\phi(x)$ in (2) is a high dimensional feature space function composed of the mixed kernel function. To solve the convex quadratic optimization problem defined by (2), we introduce the Lagrange multipliers α_i, α_i^*. Then, the optimization problem can be transformed into a dual problem as follows [25]:

$$\begin{aligned}\min_{w\in R^n,\xi,\xi^*,b\in R}\quad & \frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right)\left(x_i \cdot x_j\right) \\ & + \varepsilon\sum_{i=1}^{l}\left(\alpha_i^* + \alpha_i\right) - \sum_{i=1}^{l} y_i\left(\alpha_i^* - \alpha_i\right) \\ \text{s.t.}\quad & \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0 \\ & 0 \le \alpha_i, \alpha_i^* \le C,\ i = 1, 2, \ldots, l.\end{aligned} \tag{11}$$

By solving the dual problem above, we can derive the solution $\overline{\alpha} = (\overline{\alpha}_1, \overline{\alpha}_1^*, \overline{\alpha}_2, \overline{\alpha}_2^*, \ldots, \overline{\alpha}_l, \overline{\alpha}_l^*)^T$ of the original optimization problem defined by (2). Replacing the inner product $(x_i \cdot x_j)$ in objective function (11) by the mixed kernel function $K_{\text{mix}}(x_i, x_j)$, we can construct a decision function as follows:

$$f(x) = \sum_{i=1}^{l}\left(\overline{\alpha}_i^* - \overline{\alpha}_i\right)K_{\text{mix}}\left(x_i \cdot x\right) + \overline{b}. \tag{12}$$

where $\overline{b}$ is calculated in the following way. Select $\overline{\alpha}_j$ or $\overline{\alpha}_k^*$ in an open interval. If $\overline{\alpha}_j$ is selected, then

$$\overline{b} = y_j - \sum_{i=1}^{l}\left(\overline{\alpha}_i^* - \overline{\alpha}_i\right)K_{\text{mix}}\left(x_i, x_j\right) + \varepsilon. \tag{13}$$

If $\overline{\alpha}_k^*$ is selected, then

$$\overline{b} = y_k - \sum_{i=1}^{l}\left(\overline{\alpha}_i^* - \overline{\alpha}_i\right)K_{\text{mix}}\left(x_i, x_j\right) - \varepsilon. \tag{14}$$

4.2.2. Predictive Output Function. Suppose the sample data set of support vector regression is $D = \{(x_i, y_i) \mid i \in I\}$, where $I = \{1, 2, \ldots, N\}$ is the index set and y_i is the objective vector of the data. Divide the sample data into k groups by the k-fold cross validation method; that is,

$$D_j = \left\{\left(x_i, y_i\right) \mid i \in I_j\right\}, \tag{15}$$

where $j \in \{1, 2, \ldots, k\}$, $I_1 \cup I_j \cup \cdots \cup I_k = I$, $D_1 \cup D_2 \cup \cdots \cup D_k = D$. In each iteration randomly choose a group of data D_p as the prediction and the remaining $k-1$ groups as the training database. Given the initial parameter γ_0, we use LIBSVM [26] to train the support vector regression. Suppose the training result is $\widetilde{\alpha}$ and $\widetilde{b}$. Then, the decision function is

$$f(x) = \sum_{i=1}^{l} \left(\widetilde{\alpha}_i^* - \widetilde{\alpha}_i\right) K\left(x_i \cdot x\right) + \widetilde{b}, \tag{16}$$

where $\widetilde{\alpha} = (\widetilde{\alpha}_1, \widetilde{\alpha}_1^*, \widetilde{\alpha}_2, \widetilde{\alpha}_2^*, \ldots, \widetilde{\alpha}_l, \widetilde{\alpha}_l^*)^T$.

Substitute D_p into (16) and we can derive the prediction output of D_p as follows:

$$\widetilde{y}_p(x) = \sum_{t \in I_p} \left(\widetilde{\alpha}_p^* - \widetilde{\alpha}_p\right) K\left(x_p \cdot x\right) + \widetilde{b}. \tag{17}$$

Choose D_i, $i \in \{1, 2, \ldots, k\}$ as the prediction group and the remaining groups $D_1, \ldots, D_{i-1}, D_{i+1}, \ldots, D_k$ as the training groups. After k-fold cross validation regression prediction, all data in sample data set D has one and only one prediction output. Therefore, for parameter vector γ, we can define a prediction output function as follows:

$$\widetilde{y} = h(\gamma). \tag{18}$$

4.2.3. The Establishment of Parameter Filter Model. In this paper, the kernel function weighted fusion coefficients $p1$, $p2$, the local kernel function parameters, the parameters of the global kernel function, and the penalty parameter C are combined together as the support vector regression parameters denoted as γ. For convenience, let k_1, k_2 be the kernel parameters of local kernel function and the kernel parameters of the global kernel function, respectively. The selection of the whole parameter can be used as a filter estimation problem for a nonlinear dynamic system. The establishment of parameters of nonlinear system is as follows:

$$\gamma(k) = \gamma(k-1) + w(k) \tag{19}$$

$$y(k) = h(\gamma(k)) + v(k), \tag{20}$$

where $\gamma(k)$ is an n-dimensional state vector parameter; $y(k)$ is the output observation. Process noise $w(k)$ and observation noise $v(k)$ are both Gaussian white noise sequences with zero mean and known variances Q and R.

Because solving the optimal kernel parameters can be considered a fixed constant for a specific practical object, we can establish the linear state equation with respect to the parameters described in formula (19). For any state vector, all primitive data has a predictive output after being trained and predicted by LIBSVM, so a nonlinear observation equation can be established as formula (20). For operation of the 5th-degree cubature Kalman filtering algorithm, artificial process noise and observation noise need to be added to the system model.

4.3. Parameter Selection Method for Support Vector Regression Based on Adaptive Fusion of Mixed Kernel Function. In this section, the method for selecting model parameters of support vector regression and the specific steps of the proposed algorithm are described. The design of the parameter adjustment system is shown in Figure 2.

The parameter selection method for support vector regression based on adaptive fusion of mixed kernel function
Initialization:
(1) For original data set D, select the mixed kernel function, and set the initial parameter state value γ_0.
(2) Divide D into k groups by using k fold cross validation method denoted by $D_1, D_2, \ldots, D_k$.
While (Parameter state value does not meet the set conditions) do
Time update process:
(3) Calculate weights $\overline{w}_{s1}$, $\overline{w}_{s2}$, w_1, w_2 using formulas (6)-(7).
Measurement update process:
(4) Decompose one step prediction error covariance matrix $P_{k|k-1}$ and evaluate the cubature point ξ_i according to formula (14) in reference [19].
(5) Train the data set based on the LIBSVM algorithm to obtain the final prediction output.
(6) Combining predict $\bar{y}$, compute one step prediction by using formula (12).
(7) Use formula (8)–(14) of reference [19] to implement the subsequent measurement update.
End while
End

Algorithm 1: The detailed algorithm steps of the parameter selection method for support vector regression based on adaptive fusion of mixed kernel function.

First, the k-fold cross validation method is used to divide the original data set into k groups. Select the local kernel function and the global kernel function to determine the mixed kernel function. Train this data set with k sub-LIBSVM based on the mixed kernel function. Then, the predictive output is input to the 5th-degree Kalman filter. All parameters of the model are used as the state vector of the system; thus the selection of model parameters is a nonlinear state estimation problem.

In parameter system model (19) and (20), the real value of observation vector $y(k)$ in each iteration is unchanged. The observation vector is the original sample data of the target value vector $y(k) = (y_1, y_2, \ldots, y_N)^T$. We can make the optimal state estimation for the parameter state vector γ according to the observation vectors of real values and predicted output values to obtain a minimum variance between the real values $y(k)$ and predicted output values $\bar{y}$. 5th-degree cubature Kalman filter algorithm is used to estimate γ. The algorithm of the parameter selection method for support vector regression based on adaptive fusion of mixed kernel function includes two processes: the time update process and the measurement update process as shown in Algorithm 1.

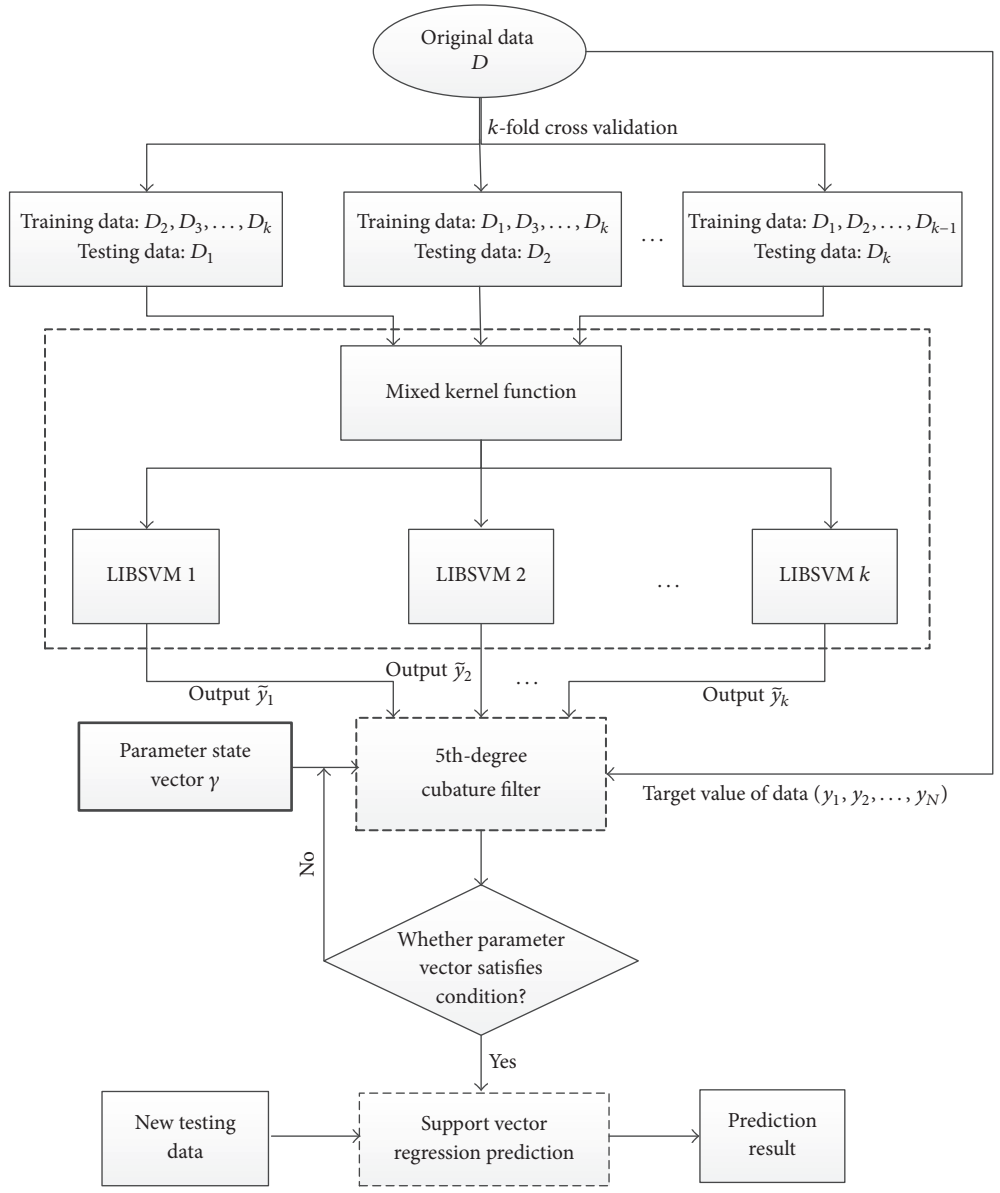


Figure 2: Parameter adjustment structure of support vector regression based on adaptive fusion of the mixed kernel function.

The proposed algorithm combines the mixed kernel function fusion coefficient, the kernel parameter, and penalty parameter C as the state parameter vector then obtains the predicted output of the data set using the k-fold cross validation method based on LIBSVM. Finally, calculate the optimal parameter state vector iteratively by the 5th-degree cubature filtering algorithm. The goal of the proposed algorithm in this paper is to search the optimal state vector γ recursively and obtain the minimum error variance between the real target value of the sample $y(k)$ and the predictive output of the support vector regression $\tilde{y}$.

5. Algorithm Analysis

The value of the kernel function is influenced by the data points that are close to each other in the algorithm of support vector regression based on local kernel function, while in the support vector regression based on local kernel function algorithm the data points that are far from each other have an effect on the value of the kernel function. Using only the global kernel function or local kernel function has some limitations in solving practical problems. It often cannot accurately describe the characteristics of the actual sample. Thus, it leads to poor performance of the support vector regression. The mixed kernel function contains two different types of kernel function, the local kernel function and global kernel function. This new kernel function can greatly improve the actual sample properties. The support vector regression based on the mixed kernel function has good learning ability and good generalization ability. However, choosing a mixed kernel function coefficient fusion remains a difficult problem. Prior knowledge of experts and simple cross validation operation

are commonly used, but these methods cannot achieve high performance support vector regression.

In this paper, we use a combination of parameters to construct the kernel function. The parameters of the local kernel function, the global kernel function, and the penalty parameter C together form the parameters of the support vector regression. Thus, the mixed kernel function can accurately describe the actual samples according to the specific characteristics of the sample, by adjusting the weighted fusion coefficient of the local kernel function and the global kernel function. The support vector regression based on the mixed kernel function has stronger extrapolation ability due to the highly integrated radial and spherical integral method applied to 5th-degree cubature Kalman filter algorithm. This algorithm has higher parameter estimation accuracy, compared with the UKF algorithm and 3rd-degree cubature Kalman filter algorithm. Therefore, the estimated result of the parameter state vector is more accurate, and the parameters brought into the support vector regression are better, and the predictability of support vector regression is better. We can understand the proposed support vector regression parameter adjustment algorithm in another way. All the state parameter vectors can be regarded as the kernel parameters of the mixed kernel function, including the combination parameter, the parameters of the local kernel function, the parameters of the global kernel function, and the penalty parameter C. The state estimation of the parameter vector is performed based on the 5th-degree cubature Kalman filter with high precision, and the optimal kernel parameter values of the support vector regression are obtained.

6. Simulation Example

6.1. Subjects. We selected the experimental data of rolling bearings for Electrical Engineering Laboratory of Case Western Reserve University for analysis and verification [27]. The measured rolling bearing type is SKF6205. Single point faults are, respectively, arranged on the driving end, the bearing outer ring, the inner ring, and the rolling body with spark erosion technique. The fault depth is 0.2794 mm and diameter 0.1778 mm. The number of balls is 9. The rolling bearing works under four states, including normal, inner ring fault, outer ring fault, and rolling body fault. Acceleration sensors are used to measure vibration signals with the traditional sampling method of signal acquisition and sampling frequency 12 kHz. The data obtained is shown in Figure 3.

6.2. Feature Extraction. The study on extraction is used to represent the fault state characteristic parameters of fault size with 12, as shown in Table 2. We take 50 groups of data from drive end of the vibration data directly. And each set of data has intercepted 4096 sample points. Then, calculate the characteristic parameters of each data.

6.3. Algorithms Comparison. In this simulation, the local kernel function of the mixed kernel function is the RBF kernel function, and the global kernel function is the sigmoid kernel function. The parameter vector of the mixed kernel function

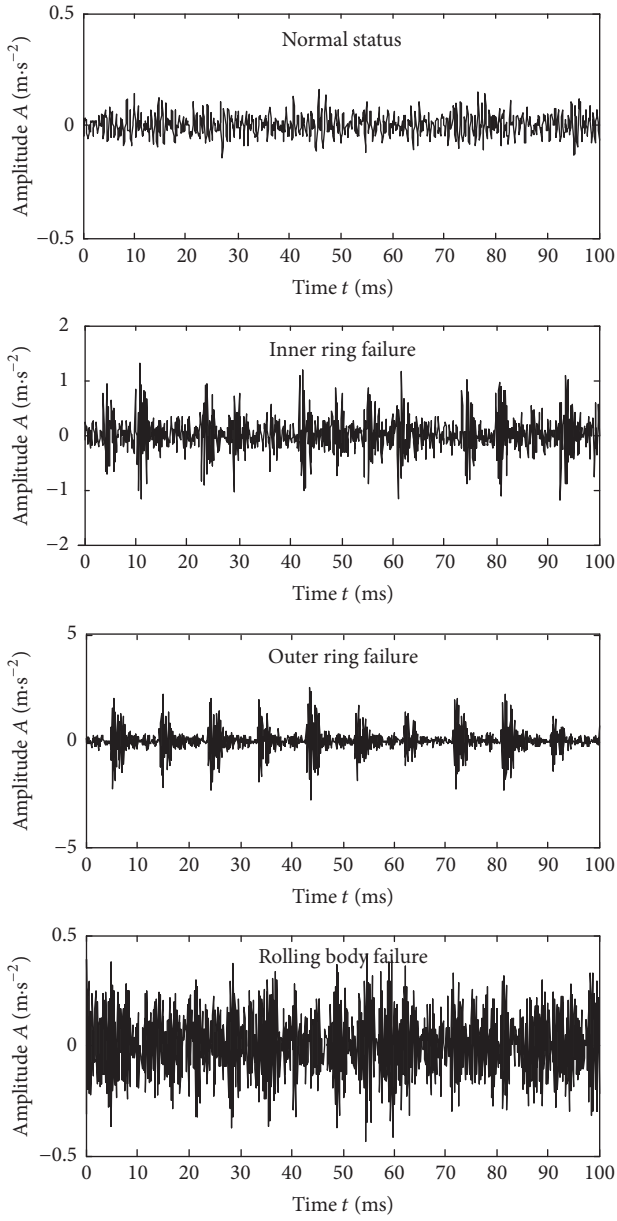


Figure 3: Vibration signals for four different types of faults.

is $\gamma = [p1, p2, \sigma, \lambda, \varphi, C]^T$, while the parameter vector based on the kernel function of a single RBF is $\gamma = [\sigma, C]^T$. In order to illustrate the effectiveness of the proposed algorithm, support vector regression algorithm for single RBF kernel function based on genetic algorithm (RBF-GA-SVR), support vector regression algorithm for single RBF kernel function based on UKF algorithm (RBF-UKF-SVR), support vector regression algorithm for single RBF kernel function based on CKF algorithm (RBF-CKF-SVR), mixed kernel function based on UKF algorithm (MKF-UKF-SVR), and support vector regression algorithm for mixed kernel function based on CKF algorithm (MKF-CKF-SVR) are compared. The prediction results for four states are shown as Figures 4–7.

For the convenience of description, we have the following simplified definitions about the filtering algorithms:

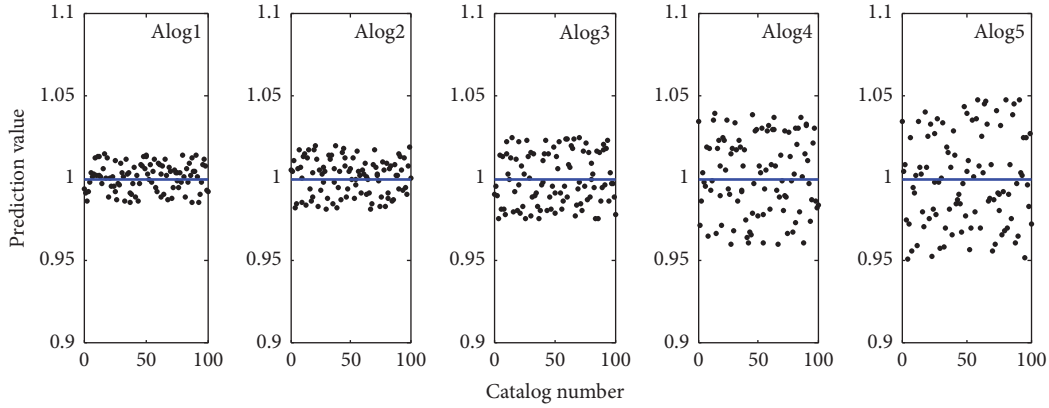


Figure 4: Prediction result for state 1.

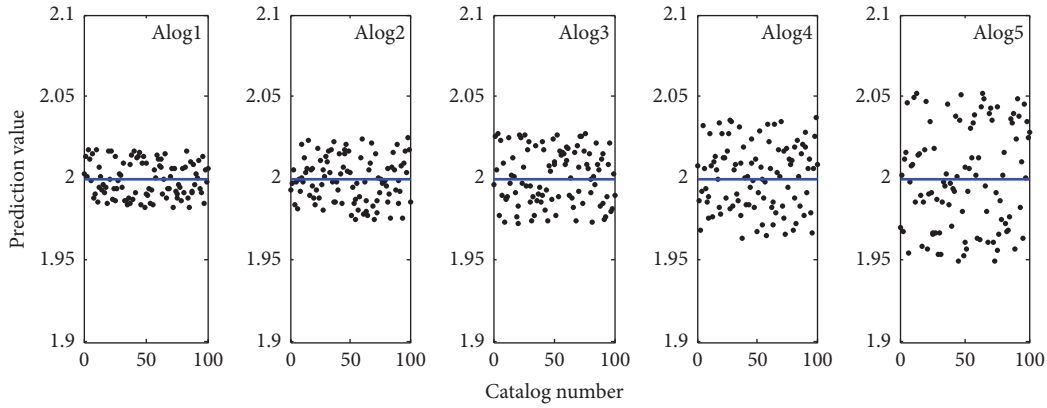


Figure 5: Prediction result for state 2.

Table 2: Sensitive features parameters.

Quantity symbol	Characteristic index
T1	Mean
T2	Absolute average
T3	Peak
T4	Square root amplitude
T5	Root mean square value
T6	Variance
T7	Crooked
T8	Kurtosis factor
T9	Waveform factor
T10	Margin factor
T11	Peak factor
T12	Pulse factor

Algo1: MKF-5th-degree-CKF-SVR algorithm

Algo2: MKF-UKF-SVR algorithm

Algo3: RBF-5th-CKF-SVR algorithm

Algo4: RBF-UKF-SVR algorithm

Algo5: RBF-GA-SVR algorithm.

Table 3: Results of parameter estimation.

Algorithms	Parameter vector γ
Algo5	[99.992, 99.999]
Algo4	[1.683, 27.723]
Algo3	[0.681, 18.230]
Algo2	[0.912, 0.088, 1.501, 24.342]
Algo1	[0.902, 0.098, 0.708, 1.034, 2.127, 73.364]

In order to show the robustness of the proposed method in front of typical noisy levels, we chose three different levels of noise to experiment with $R1 = 0.1$, $R2 = 0.3$, and $R3 = 0.5$, respectively. The prediction results of these algorithms are shown in Figures 8–10. Kernel parameter estimation results and the prediction error results are shown in Tables 2 and 3, respectively.

From the simulation results in Figures 4–7 and Tables 3 and 4, it can be seen that the kernel parameter of Algo5 is larger, so it has poor generalization ability and larger prediction error. Compared with the Algo5, the Algo4 algorithm has higher prediction accuracy. This is mainly due to the use of the filtering framework to estimate the kernel parameters. Because of the high accuracy of 5th cubature Kalman filter algorithm to estimate the parameters, kernel parameter

Table 4: Results of sample prediction error.

Data	Statistical indicators	Algorithms				
		Algo1	Algo2	Algo3	Algo4	Algo5
Train data	MAE	0.0053	0.0073	0.0094	0.0156	0.0200
	RMSE	0.0098	0.0146	0.0198	0.0298	0.0399
	SD	0.0060	0.0090	0.0110	0.0179	0.0228
Test data	MAE	0.0095	0.0128	0.0171	0.0208	0.0274
	RMSE	0.0111	0.0152	0.0195	0.0241	0.0319
	SD	0.0110	0.0150	0.0203	0.0252	0.0331

Here, MAE represents mean absolute error, RMSE represents mean square error, and SD represents standard deviation.

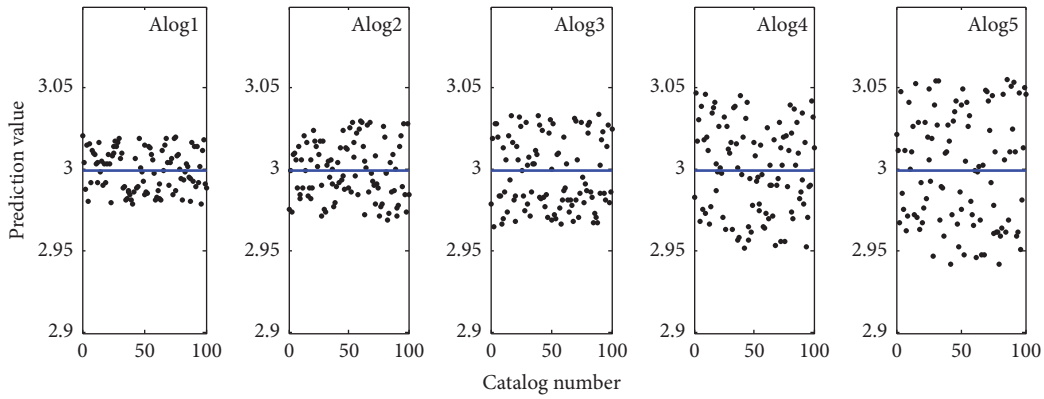


Figure 6: Prediction result for state 3.

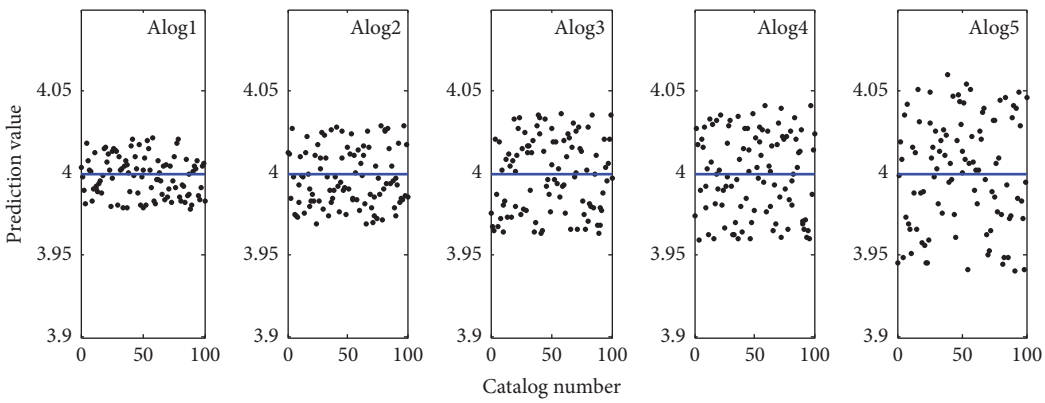


Figure 7: Prediction result for state 4.

values of Algo3 are smaller, and the predictive ability is better than the Algo5 algorithm and Algo4 algorithm. But its accuracy is lower than Algo2 due to the effect of the mixed kernel. The proposed Algo1 algorithm characterizes the sample information by using local and global kernel function information. The Algo1 algorithm has stronger generalization ability, and the prediction error is the least. From the perspective of the fusion coefficient, coefficient of the local kernel function RBF is 0.902. This is mainly because the actual sample prefers to use the local kernel function, but compared with the Algo3, the predicted accuracy of Algo1 algorithm is greatly improved. This is the key role played by the global kernel function in the mixed kernel function, which makes it more complete and accurate to describe actual sample information.

From Figures 8, 9, and 10, the kernel parameter estimation error becomes larger with the increasing noise levels. It is normal that the large noise influences the accuracy of the nonlinear filter. But the estimated errors are within the allowable range.

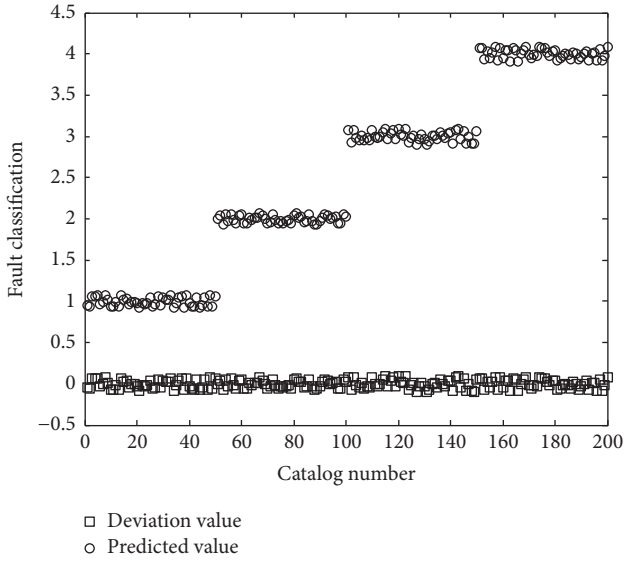


Figure 8: Prediction result of Algo1 for $R1 = 0.1$.

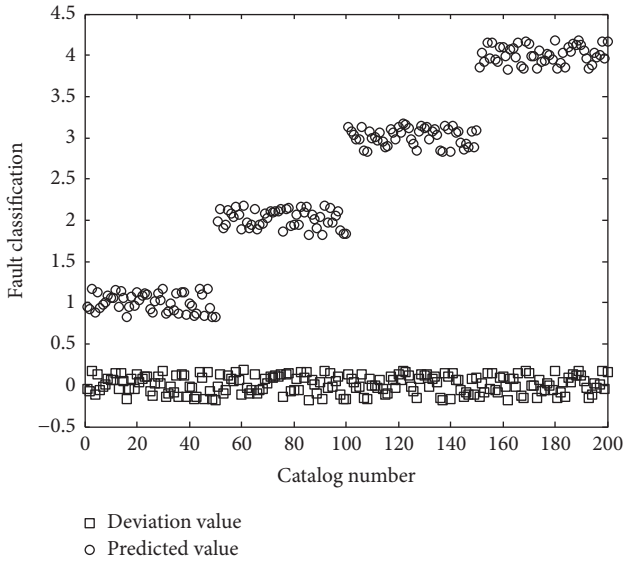


Figure 9: Prediction result of MKF 5th-degree CKF-SVR for $R2 = 0.3$.

7. Conclusions

The proposed algorithm parameter selection method for support vector regression based on adaptive fusion of mixed kernel function combines the mixed kernel function fusion coefficient, kernel function parameter, and regression parameters together as the parameters of the state vector and obtains the predicted output of the original data set based on LIBSVM. The fusion coefficients are adaptively adjusted by the 5th-degree Kalman filter, and the kernel function parameters and the regression parameters are selected by using the estimated parameters values. Finally, the prediction of the bearing fault diagnosis experiment shows that the kernel function and the parameters based on the method proposed in this paper can obtain stronger generalization ability of the support vector regression and higher prediction accuracy.

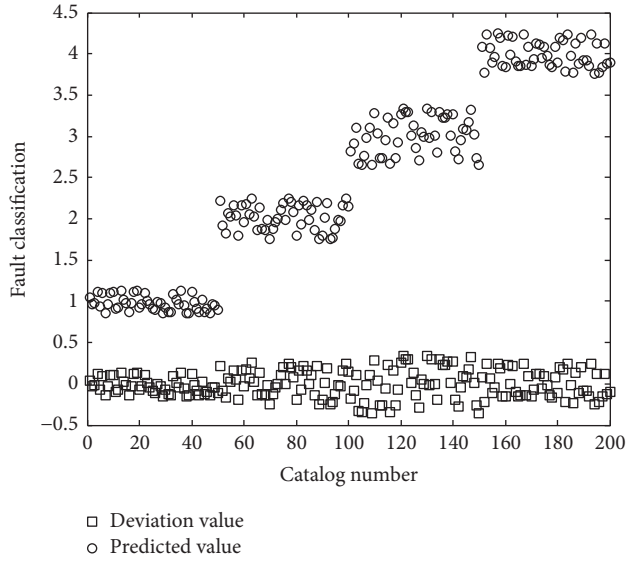


Figure 10: Prediction result of MKF 5th-degree CKF-SVR for $R3 = 0.5$.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61403229) and Natural Science Foundation of Zhejiang Province (LY13F030011).

References

[1] A. Widodo and B. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[2] M. Kang, J. Kim, A. C. Tan, E. Y. Kim, and B. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2786–2797, 2015.

[3] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2013.

[4] V. N. Vapnik, *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, Wiley- Interscience, New York, NY, USA, 1998.

[5] P. Refaeilzadeh, L. Tang, and H. Liu, *Cross-validation*, Springer-US, 2009.

[6] Z. Shao, M. J. Er, and N. Wang, "An effective semi-cross-validation model selection method for extreme learning machine with ridge regression," *Neurocomputing*, vol. 151, no. 2, pp. 933–942, 2015.

[7] K. Sopyła and P. Drozda, "Stochastic gradient descent with Barzilai-Borwein update step for SVM," *Information Sciences*, vol. 316, pp. 218–233, 2015.

[8] T. Villmann, S. Haase, and M. Kaden, "Kernelized vector quantization in gradient-descent learning," *Neurocomputing*, vol. 147, no. 1, pp. 83–95, 2015.

[9] W. Froelich and J. L. Salmeron, "Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series," *International Journal of Approximate Reasoning*, vol. 55, no. 6, pp. 1319–1335, 2014.

[10] O. J. H. Bosch, N. C. Nguyen, T. Maeno, and T. Yasui, "Managing Complex Issues through Evolutionary Learning Laboratories," *Systems Research and Behavioral Science*, vol. 30, no. 2, pp. 116–135, 2013.

[11] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "More efficiency in multiple kernel learning," in *Proceedings of the 24th International Conference on Machine Learning, ICML 2007*, pp. 775–782, June 2007.

[12] J. R. Lee, P. Raghavendra, and D. Steurer, "Lower bounds on the size of semidefinite programming relaxations," in *Proceedings of the 47th Annual ACM Symposium on Theory of Computing, STOC 2015*, pp. 567–576, June 2015.

[13] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," *Bioinformatics*, vol. 1, no. 1, 2003.

[14] D. Whitley, "An executable model of a simple genetic algorithm," *Foundations of Genetic Algorithms*, vol. 1519, pp. 45–62, 2014.

[15] F. Kuang, S. Zhang, Z. Jin, and W. Xu, "A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection," *Soft Computing*, vol. 19, no. 5, pp. 1187–1199, 2015.

[16] T. Mu and A. K. Nandi, "Automatic tuning of L2-SVM parameters employing the extended Kalman filter," *Expert Systems with Applications*, vol. 26, no. 2, pp. 160–175, 2009.

[17] D. Y. Huang and X. Y. Chen, "A novel approach of model selection for SVR," *Journal of Fuzhou University*, vol. 39, no. 4, pp. 527–532, 538, 2011.

[18] H. Wang, M. Lv, and L. Zhang, "Parameter optimization of SVR based on DRVB-ASCKF," in *Proceedings of the International Conference on Estimation, Detection and Information Fusion, ICEDIF 2015*, pp. 141–145, January 2015.

[19] B. Jia, M. Xin, and Y. Cheng, "High-degree cubature Kalman filter," *Automatica*, vol. 49, no. 2, pp. 510–518, 2013.

[20] Y. Li, *Research on Multi-core Learning SVM Algorithm and Identification of Pulmonary Nodules*, College of Communication Engineering, Jilin University, Jilin, China, 2014.

[21] X. Lin, J. Yang, and C. Z. Ye, "Face recognition technology based on support vector machine," *Journal of Infrared and Laser Engineering*, vol. 30, no. 5, pp. 318–322, 2001.

[22] S. G. Pour and F. Girosi, "Joint Prediction of Chronic Conditions Onset: Comparing Multivariate Probits with Multiclass Support Vector Machines," in *Proceedings of the Symposium on Conformal and Probabilistic Prediction with Applications*, Springer International Publishing, 2016.

[23] C.-J. Hsieh, S. Si, and I. S. Dhillon, "A divide-and-conquer solver for kernel support vector machines," in *Proceedings of the 31st International Conference on Machine Learning, ICML 2014*, pp. 855–870, June 2014.

[24] X. Huang, *The study on Kernel in Support Vector Machine [Ph. D. dissertation]*, Soochow University, 2008.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[26] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

[27] "The experimental data of rolling bearings for Electrical Engineering Laboratory of Case Western Reserve University," 2008, https://case.edu/bulletin/.

[28] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression," *IEEE Transactions on Reliability*, vol. 62, no. 4, pp. 821–832, 2013.

[29] J. P. Jacobs, "Bayesian support vector regression with automatic relevance determination kernel for modeling of antenna input characteristics," *Institute of Electrical and Electronics Engineers. Transactions on Antennas and Propagation*, vol. 60, no. 4, pp. 2114–2118, 2012.

[30] M. Ghaedi, M. R. Rahimi, A. M. Ghaedi, I. Tyagi, S. Agarwal, and V. K. Gupta, "Application of least squares support vector regression and linear multiple regression for modeling removal of methyl orange onto tin oxide nanoparticles loaded on activated carbon and activated carbon prepared from Pistacia atlantica wood," *Journal of Colloid and Interface Science*, vol. 461, pp. 425–434, 2016.

[31] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Health assessment and life prediction of cutting tools based on support vector regression," *Journal of Intelligent Manufacturing*, vol. 26, no. 2, pp. 213–223, 2015.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 8168627, 8 pages
https://doi.org/10.1155/2017/8168627

Research Article

Development of Fault Identification System for Electric Servo Actuators of Multilink Manipulators Using Logic-Dynamic Approach

V. Filaretov,[1,2] A. Zuev,[1,2] A. Zhirabok,[3] and A. Protsenko[3]

[1]*Institute of Automation and Control Processes, FEB RAS, Vladivostok, Russia*
[2]*Far Eastern Federal University, Vladivostok, Russia*
[3]*Institute of Marine Technology Problems, Vladivostok, Russia*

Correspondence should be addressed to A. Zuev; alvzuev@gmail.com

Received 28 April 2017; Revised 24 August 2017; Accepted 7 September 2017; Published 1 November 2017

Academic Editor: Chunhui Zhao

This paper presents the method of synthesis of faults identification systems for electric servo actuators of multilink manipulators. These actuators are described by nonlinear equations with significantly changing coefficients. The proposed method is based on logic-dynamic approach for design of diagnostic observers for fault detection and isolation. An advantage of this approach is that it allows studying systems with nonsmooth nonlinearities by linear methods only. For solving the task of faults identification, a residual signal feedback was proposed to be used for observers. The efficiency of the proposed fault identification system was confirmed by results of simulation.

1. Introduction

At present, the use of multilink manipulators (MM) has considerably extended and the operations which are carried out by them become more complicated [1–4]. The occurrence of any faults in actuators of MM may lead to significant aggravation of quality of performance of the MM operations and reduction of accuracy. Therefore, the reliability and safety of such complex equipment at operating in autonomous mode have the significant importance. Today the most typical actuators of MM are electric servo actuators. One way to improve the reliability of electric servo actuators of MM is using real-time fault diagnosis and fault tolerant control [1]. Fault tolerant control is based on forming special control, providing automatic stabilization of the main parameters of actuators when faults occur [2]. For providing such control, all possible faults should be timely detected and their values should be estimated.

The problem of fault detection and isolation (FDI) was extensively investigated for the past 20 years; see, for example, the surveys [2, 5–8] and the books [9, 10]. Many problems have been studied and solved: different methods of residual generation and relationships including robustness and adaptive threshold test, H_∞-approach, and fuzzy logic; many types of systems have been considered: linear, descriptor, and different classes of nonlinear systems. Many practical examples were considered; in particular, special book was devoted to robotic systems [11].

As a rule, the electric servo actuators of MM are nonlinear essentially with such types of no differentiable nonlinearities as saturation, Coulomb friction, backlash, and hysteresis [12]. Therefore, one has to use nonlinear methods of the FDI. However, most of papers dealing with the FDI problem consider the nonlinear systems with differential nonlinearities [13–16]; therefore they cannot be used in our case.

At present, there are several approaches for designing diagnostic observers allowing fault diagnosis for systems with no differentiable nonlinearities. Among them there are methods using the algebra of functions [17], the logic-dynamic approach (LDA) [18–20], and others [21–24]. Whereas the algebra of functions demands complex analytical calculations using symbolic mathematical packages, the LDA allows

using the methods of linear algebra for nonlinear systems and provides relatively simple procedure for synthesis of diagnostic observers. The LDA includes three steps: replacing the original nonlinear system with appropriate linear system, constructing the set of linear diagnostic observers, and adding to these observes the nonlinear parts. But LDA can solve only FDI problem without fault identification. To estimate the value of the faults, one uses in [19, 20] a feedback in diagnostic observers. These methods provide the faults identification for only first-order diagnostic observers. But at case when electric servo actuators of MM have partially measured state vector, then there are no guarantees that observers will have first order.

The present paper suggests a solution to the problem of developing the method of synthesis of fault identification (FI) system for electric servo actuators of MM which are described by mathematical model including differential equations with nonlinearities and partially measured state vector. The proposed method assumes the use of the diagnostic observers based on LDA with special feedback signal dependent on the residual for real-time identification of fault value.

The paper is organized as follows. Section 2 describes the model of electric servo actuators of MM. Section 3 is devoted to the synthesis of diagnostic observers for FDI using logic-dynamic approach. Firstly, the main steps of this approach are given and then the observers are described in detail. In Section 4, the methods of synthesis of FI system for created observers are presented. First, the analysis of mismatch of states of electric servo actuator and the observer is given and then the synthesis of residual signal feedback is shown. An example is considered in Section 5. Section 6 concludes the paper.

2. Description of Dynamics of Electric Servo Actuators of MM

In this paper, the MM with PUMA type kinematic scheme with 6 degrees of freedom is considered (see Figure 1). Today, the MM with such kinematic scheme are mostly used in industry. Each degree has an electric servo actuator that includes DC motor, reducing gear and sensors measuring the electric current, angular speed of DC motor rotor, and angle of rotation of output shaft of reducing gear.

It is expedient to use the second form Lagrange equation or other methods to obtain the moment torques M_i in each degree of freedom of the MM (see Figure 1) while moving its gripper along the complex spatial trajectory, if laws of change of generalized coordinates q_i ($i = \overline{1,6}$) of the MM are known.

The expressions for moments acting in each degree of freedom of MM can be represented in the form [3]

$$M_i = H_{ii}(q)\,\ddot{q}_i + h_i(q,\dot{q})\,\dot{q}_i + M_{Ei}(q,\dot{q},\ddot{q}), \qquad (1)$$

where $H_{ii}(q)$ is the diagonal element of matrix of inertia of MM; $h_i(q,\dot{q})$ is the element of the vector of Coriolis and centrifugal forces; and $M_{Ei}(q,\dot{q},\ddot{q})$ is the torque which represents interactions acting to ith degree of freedom from other MM's links and also action of the gravitational forces. These values with the help of second Lagrange equation are defined. Further for simplicity we will not write the index i.

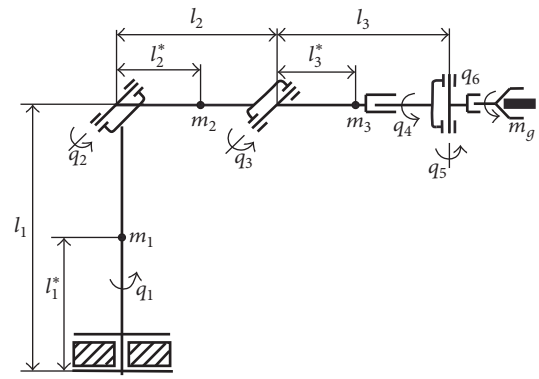


Figure 1: MM with PUMA type kinematic scheme.

It is assumed that the following typical faults are possible in each electric servo actuator of the MM:

(i) The fault $d_1(t)$ caused by change of value of Coulomb friction moment $\widetilde{M}_{\mathrm{cfg0}}$ in reducing gear

(ii) The fault $d_2(t)$ caused by change of value of Coulomb friction moment $\widetilde{M}_{\mathrm{cfm0}}$ in DC motor

(iii) The fault $d_3(t)$ caused by change of value of electrical resistance $\widetilde{R}$.

Considering (1) and these faults, model of electric servo actuator of the MM can be presented by the following differential equations:

$$\begin{aligned}
\dot{x}_1 &= x_2,\\
\dot{x}_2 &= -\frac{c i_g^{\,2}}{H(q)}x_1 - \frac{\left(k_{vg} + h(q,\dot{q})\right)}{H(q)}x_2 - \frac{M_E(q,\dot{q},\ddot{q})}{H(q)}\\
&\quad - \frac{M_{\mathrm{cfg0}}}{H(q)}\operatorname{sign}(x_2) + d_1(t),\\
\dot{x}_3 &= x_4,\\
\dot{x}_4 &= \frac{c i_g}{J}x_1 - \frac{c}{J}x_3 - \frac{k_{vm}}{J}x_4 + \frac{k_m}{J}x_5\\
&\quad - \frac{M_{\mathrm{cfm0}}}{J}\operatorname{sign}(x_4) + d_2(t),\\
\dot{x}_5 &= \frac{-k_\omega}{L}x_4 - \frac{R}{L}x_5 + \frac{k_u}{L}u + d_3(t),
\end{aligned} \qquad (2)$$

where $x_1 = \dot{a}_g$ and $x_2 = a_g$ are speed and angle of output shaft of reducing gear; $x_3 = \dot{a}_m$ and $x_4 = a_m$ are speed and angle of DC motor shaft; R, L, and $x_5 = I$ are resistance, inductance, and current of electric motor rotor circuits, accordingly; k_ω is a coefficient of counter-EMF; k_m is a coefficient of motor moment; J is a rotational inertia of rotating parts of reduction drive and DC motor; c is rigidity of mechanical transmission; i_g is a coefficient of the reducing gear; u is control signal; k_u is an amplifier gain; M_{cfm0} and M_{cfg0} are coefficients of

Coulomb friction of DC motor and reducing gear; and the variables d_j $(j = 1, 2, 3)$ represent the effects of faults on the system and are of the form

$$d_1(t) = -\frac{\widetilde{M}_{\text{cfg0}}}{H(q)}\operatorname{sign}(x_2),$$
$$d_2(t) = -\frac{\widetilde{M}_{\text{cfm0}}}{H(q)}\operatorname{sign}(x_2), \tag{3}$$
$$d_3(t) = -\frac{\widetilde{R}}{L}I.$$

Equation (2) has nonlinearities and significantly changing coefficients because of changes in parameters $H(q)$, $h(q, \dot{q})$, and $M_E(q, \dot{q}, \ddot{q})$.

3. Synthesis of Diagnostic Observers for FDI Using LDI

System (2) can be represented by the differential equation in a matrix form as follows:

$$\dot{x}(t) = Fx(t) + Gu(t) + B(x) + Ed(t),$$
$$y(t) = Cx(t), \tag{4}$$

where $B(x)$ is a vector corresponding to the system nonlinearities, $x \in R^5$, $y \in R^3$, $C \in R^{3\times5}$, $F \in R^{5\times5}$, $G \in R^5$,

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -\dfrac{ci_g^{\,2}}{H(q)} & -\dfrac{\left(k_{vg} + h(q,\dot{q})\right)}{H(q)} & \dfrac{ci_g}{H(q)} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \dfrac{ci_g}{J} & 0 & -\dfrac{c}{J} & -\dfrac{k_{vm}}{J} & \dfrac{k_m}{J} \\ 0 & 0 & 0 & -\dfrac{k_\omega}{L} & -\dfrac{R}{L} \end{bmatrix},$$

$$E = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 \\ \dfrac{\left(-M_E(q,\dot{q},\ddot{q}) - M_{\text{cfg0}}\operatorname{sign}(x_2)\right)}{H(q)} \\ 0 \\ -\dfrac{M_{\text{cfm0}}\operatorname{sign}(x_4)}{J} \\ 0 \end{bmatrix},$$

$$x = \begin{bmatrix} a_g \\ \dot{a}_g \\ a_m \\ \dot{a}_m \\ I \end{bmatrix},$$

$$G = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \dfrac{k_u}{L} \end{bmatrix}. \tag{5}$$

If faults occur during the work of MM, they should be promptly detected and identified. Then the fault tolerant control can be used. The traditional methods of diagnosis that focus on systems described by linear models cannot provide the solution to the problem of synthesis of FI systems for electric servo actuators of the MM described by model with the parameters $H(q)$, $h(q, \dot{q})$, and $M_E(q, \dot{q}, \ddot{q})$ that are significantly variable and under the presence of the nonlinear component $B(x)$.

In this paper, another synthesis method of FI systems is proposed. This method includes two basic steps [25]: (1) using LDA to construct the diagnostic observer [17, 18] which can detect occurrence of specific faults; (2) introducing the specific feedback in observers for identification of the fault values [19]. Below these stages are considered in more detail.

The LDA to solve the FDI problem for system (4) includes the following steps [18]: (1) replacing the initial nonlinear system (4) by linear system; (2) solving the FDI problem for the linear system with some additional restrictions and obtaining the bank of linear observers; (3) transforming the linear observers into the nonlinear ones. Thus, the nonlinear diagnostic observers detecting the faults are of the following general form:

$$\dot{x}_*(t) = F_* x_*(t) + G_* u(t) + B_*(x_*) + M_* y(t),$$
$$y_*(t) = C_* x_*(t), \tag{6}$$

where $x_*(t) \in R^k$ is the state vector of the observer, $y_*(t)$ is the output vector, $F_* \in R^{k\times k}$, $G_* \in R^k$, $B_* \in R^k$, $M_* \in R^{k\times3}$, and $C_* \in R^k$ are constant matrices, and k is dimension of the

observer; it is assumed that the matrices F_* and C_* can be represented in a canonical form

$$F_* = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad C_* = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{7}$$

The observer (6) generates the residual $r(t) = Ny(t) - y_*(t)$ for certain row matrix $N \in R^3$ which has to be determined. If there are no faults and $d(t) = 0$, then $r(t) = 0$; if a fault occurs, $r(t) \neq 0$. It is well-known from the linear FDI theory [2, 9, 26] that, for the observer design, the matrix Φ where $\Phi x(t) = x_*(t)$ in the unfaulty case after the response to unlike conditions has died out plays the main role. In the absence of faults, the following well-known set of equations holds [2, 9, 26]:

$$\begin{aligned} C_*\Phi &= NC, \\ \Phi F &= F_*\Phi + M_*C, \\ G_* &= \Phi G, \\ B_*(x_*) &= \Phi B(x). \end{aligned} \tag{8}$$

For each fault d_i, the diagnostic observer O_i should be constructed. Matrices and vectors Φ, F_*, $B_*(x_*)$, G_*, N, and M_* describing the observer can be found in (8) [17]. As a result, the observers O_1, O_2, and O_3 for the faults d_1, d_2, and d_3 were obtained:

O_1:

$$\Phi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \dfrac{\left(k_{vg} + h(q,\dot{q})\right)}{H(q)} & 1 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{ci_g}{H(q)} & 0 & 0 \end{bmatrix},$$

$$M_* = \begin{bmatrix} -\dfrac{\left(k_{vg} + h(q,\dot{q})\right)}{H(q)} & 0 & 0 \\ -\dfrac{ci_g^2}{H(q)} & 0 & 0 \\ 0 & \dfrac{ci_g}{H(q)} & 0 \end{bmatrix},$$

$$N = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},$$

$$F_* = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$G_* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$B_* = \begin{bmatrix} 0 \\ -\dfrac{M_E(q,\dot{q},\ddot{q}) + M_{\mathrm{cfg0}}\operatorname{sign}(x_2)}{H(q)} \\ 0 \end{bmatrix}. \tag{9}$$

O_2:

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -\dfrac{c}{J} & 0 & 0 \end{bmatrix},$$

$$M_* = \begin{bmatrix} \dfrac{ci_g}{J} & -\dfrac{k_{vm}}{J} & \dfrac{k_m}{J} \\ 0 & -\dfrac{c}{J} & 0 \end{bmatrix},$$

$$N = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix},$$

$$F_* = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$G_* = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$B_* = \begin{bmatrix} -\dfrac{M_{\mathrm{cfm0}}\operatorname{sign}(x_4)}{J} \\ 0 \end{bmatrix}. \tag{10}$$

O_3:

$$\Phi = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M_* = \begin{bmatrix} 0 & -\dfrac{k_\omega}{L} & -\dfrac{R}{L} \end{bmatrix},$$

$$N = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix},$$

$$F_* = 0,$$

$$G_* = \frac{k_u}{L},$$

$$B_* = 0. \tag{11}$$

As a result, the synthesized observers O_1, O_2, and O_3 provide solving the FDI problem for electric servo actuators

of MM. Clearly, the observers O_1 and O_2 have order more than one. This is caused by the fact that the electric servo actuator (2) has the partially measured state vector. Therefore, the known methods [19, 20] for FI cannot be used. Below the new synthesis method of FI systems for observers constructed by using LDA is proposed.

4. Synthesis of FI System for Observers Created by Using LDA

To solve the problem of FI for synthesized observers, consider how residual $r(t)$ changes when the faults $d_j(t)$ occur. After differentiating the equation for residual and taking into account (4), (6), and (8), the following can be obtained:

$$\dot{r}(t) = C_* \left(F_* \left(\Phi x(t) - x_*(t)\right) + \Phi B(x) - B_*\left(x_*\right) + \Phi E d(t)\right). \tag{12}$$

The vector $e(t) \in R^k$ is introduced which is a mismatch of states of electric servo actuator and the observer:

$$e(t) = \Phi x(t) - x_*(t). \tag{13}$$

Taking into account the vector C_*, the residual $r(t)$ is a first element of mismatch vector $e(t)$: $r(t) = e_1(t)$. Therefore, (12) can be rewritten as

$$\dot{e}(t) = F_* e(t) + \Phi B(x) - B_*\left(x_*\right) + \Phi E d(t). \tag{14}$$

Analysis of (14) shows that if order of the obtained observer is one (i.e., $y_*(t) = x_*(t)$), then if faults occur, the following conditions are true: (1) $F_* = 0$, (2) $e(t) = r(t) \in R^1$, and (3) $B_* = \Phi B$. In this case, the feedback is proportional to the residual signal as $Tr(t)$ (T is the feedback coefficient) and allows identifying the faults value [19, 20]. However, in case when order of the observer is more than one, $B_* \neq \Phi B$ due to $x_*(t) \neq \Phi x(t)$. Moreover, in (6) the components $F_* e(t)$ and $\Phi B(x(t))$ are unknown. As a result, it is impossible to make fault identification. To solve this problem, one proposes a feedback of residual signal which provides $e(t) = 0$. In this case the value of this feedback is proportional to the value of fault. This provides a possibility of identifying the fault value.

After the introduction of proposed residual signal feedback, the model of observer will be

$$\dot{x}_*(t) = F_* x_*(t) + G_* u(t) + B_*\left(x_*\right) + M_* y(t) + w(r), \tag{15}$$

$$y_*(t) = C_* x_*(t), \tag{16}$$

where $w(r) \in R^k$ is the vector specifying feedback dependent on the residual signal.

Considering the feedback (15), equation for the mismatch vector is as follows:

$$\dot{e}(t) = F_* e(t) + \Phi B(x) - B_*\left(x_*\right) + \Phi E d(t) - w(r). \tag{17}$$

Equation (17) can be rewritten in the form

$$\begin{aligned}
\dot{e}_1(t) &= e_2(t) - w_1(r),\\
\dot{e}_2(t) &= e_3(t) - w_2(r_1),\\
\dot{e}_3(t) &= e_4(t) - w_3(r),\\
&\vdots\\
\dot{e}_v(t) &= e_{v+1}(t) + f_v d_j(t) + a_v - w_v(r).\\
&\vdots\\
\dot{e}_k(t) &= -w_k(r),
\end{aligned} \tag{18}$$

where a_v is the vth element of the vector $\Phi B(x(t)) - B_*(x_*(t))$, $v = \overline{1,k}$, and f_v is the vth element of vector ΦL.

When $w_{v+1}(r) = 0, \ldots, w_k(r) = 0$ and the initial conditions of the electric servo actuator and the observer are agreed upon, $e_{v+1}(t), \ldots, e_k(t)$ are zero regardless of occurrence of the fault. Taking this into account, it is possible to rewrite (18) in the form of a differential equation:

$$\begin{aligned}
f_v d_j(t) + a_v &= \dot{e}_v(t) + w_v(r)\\
&= \ddot{e}_{v-1}(t) + \dot{w}_{v-1}(r) + w_v(r)\\
&= \dddot{e}_{v-2}(t) + \ddot{w}_{v-2}(r) + \dot{w}_{v-1}(r) + w_v(r)\\
&= e_1^{(v)}(t) + w_1^{(v-1)}(r) + \cdots + w^{(1)}{}_{v-1}(r)\\
&\quad + w_v(r).
\end{aligned} \tag{19}$$

Making elements of the feedback vector, except the vth element, proportional to the residual $w_n(r) = (T_n/T_0)r(t)$ ($n = \overline{1, v-1}$) and $w_v(r) = (T_v/T_0)r(t) + (1/T_0)z(r)$ ($T_0, \ldots, T_v$ are feedback coefficients), one obtains

$$f_v d_j(t) + a_v = r^{(v)}(t) + \frac{T_1}{T_0} r^{(v-1)}(t) + \cdots + \frac{T_{v-1}}{T_0} r^{(1)}(t) + \frac{T_v}{T_0} r(t) + \frac{1}{T_0} z(r). \tag{20}$$

Feedback coefficients $T_0, \ldots, T_v$ are chosen to guarantee the stability of observers [25].

Consider $f_v d_j(t) + a_v$ as input signal and $r(t)$ as output; then (20) can be presented by a transfer function $W = T_0/(T_0 s^v + \cdots + T_v + z(s))$, where s is the symbol of Laplace transformation. Suppose that input $f_v d_j(t) + a_v$ can be presented by a polynomial of degree q : $f_v d_j(t) + a_v = b_0 + b_1 t + \cdots + b_q t^q$. The Laplace transformation of such input is $Q = b_0/s + b_1/s^2 + \cdots + b_q/s^{q+1}$. Then the steady value of $r(t)$ can be found as $r(t) = \lim_{s\to 0} = \lim_{s\to 0} WQs = T_0(b_0 s^{q-1} + \cdots + b_q)s/((T_0 s^v + \cdots + T_v)s^{q+1} + z(s)s^{q+1})$. Assuming $z(s) = T_{v+1}/s + T_{v+2}/s^2 + \cdots + T_{v+q}/s^q + T_{v+q+1}/s^{q+1}$, it is possible to obtain

$r(t) = \lim_{s\to 0}(T_0(b_0 s^{q-1} + \cdots + b_q)s/((T_0 s^v + \cdots + T_v)s^{q+1} + T_{v+1}s^q + \cdots + T_{v+q+1})) = 0$.

Thus, the introduction of such a feedback results in a steady value of $r(t) = 0$. This also results in resetting all elements of the vector $e(t)$. For many real systems, value of the fault is constant or slowly changes and it is sufficient to introduce one integral of residual in feedback to achieve the required accuracy of the faults identification: $z(r) = \int r(t)dt$. Thus, after using this kind of feedback, residual tends to zero, after the completion of the transition process, including the event when the faults occur. This ensures the synchronization of state vectors of the electric servo actuator and the observer. Thus, fault d_j can be found as

$$d_j(t) = \frac{1}{f_v T_0}\int r(t)\,dt. \tag{21}$$

After applying the proposed procedure for observers O_1, O_2, and O_3, we receive the following feedback for fault identification:

O_1:

$$\begin{aligned} w &= \begin{bmatrix} w_{11} & w_{12} & 0 \end{bmatrix}^T, \\ w_{11} &= \frac{T_1}{T_{01}} r(t), \\ w_{12} &= \frac{T_2}{T_{01}} r(t) + \int \frac{1}{T_{01}} r(t)\,dt. \end{aligned} \tag{22}$$

O_2:

$$\begin{aligned} w &= \begin{bmatrix} w_{12} & 0 \end{bmatrix}^T, \\ w_{12} &= \frac{T_3}{T_{02}} r(t) + \int \frac{1}{T_{02}} r(t)\,dt. \end{aligned} \tag{23}$$

O_3:

$$w(r) = T_4 r(t). \tag{24}$$

Finally, considering the expression (21), it is possible to identify the faults value for the electric servo actuators of the MM:

$$\begin{aligned} \widetilde{M}_{\mathrm{cfg0}} &= -\frac{H(q)\,\mathrm{sign}\left(\dot{a}_g\right)}{f_v T_0}\int r(t)\,dt, \\ \widetilde{M}_{\mathrm{cfm0}} &= -\frac{J\,\mathrm{sign}\left(\dot{a}_m\right)}{f_v T_0}\int r(t)\,dt, \\ \widetilde{R} &= -\frac{r(t)\,LT_4}{I}. \end{aligned} \tag{25}$$

Thus, the problem of identification of value of faults has been solved.

5. Simulation of Performance of FI System

Analysis of effectiveness of developed FI system was implemented using PUMA type of the MM. The FI system was synthesized for electric servo actuator of second degree of freedom of MM (see Figure 1). The results for other degrees of freedom of the MM are similar.

The second degree of freedom of MM has the terms H, h, and M_E as follows:

$$\begin{aligned} H &= J_{n2} + J_{n3} + m_2 l_2^{*2} + m_3 l_3^{*2} + \left(m_3 + m_g\right) l_2^2 \\ &\quad + m_g l_3^2 + 2l_2\left(m_3 l_3^* + m_g l_3\right)\cos q_3, \\ h &= -2l_2\left(m_3 l_3^* + m_g l_3\right)\dot{q}_3 \sin q_3, \\ M_E &= \left[J_{n3} + m_3 l_3^{*2} + m_g l_3^2 + \left(m_3 l_3^* + m_g l_3\right) l_2 \cos q_3\right] \\ &\quad \cdot \ddot{q}_3 - l_2\left(m_3 l_3^* + m_g l_3\right)\dot{q}_3^2 \sin q_3 + g\left[m_2 l_2^*\right. \\ &\quad \left. + \left(m_3 + m_g\right) l_2\right]\sin q_2 + g\left(m_3 l_3^* + m_g l_3\right) \\ &\quad \cdot \sin\left(q_2 + q_3\right) \\ &\quad - \frac{1}{2}\left[\left(J_{n2} - J_{s2} + m_2 l_2^{*2} + \left(m_3 + m_g\right) l_2^2\right)\sin 2q_2\right. \\ &\quad + \left(J_{n3} - J_{s3} + m_3 l_3^{*2} + m_g l_3^2\right)\sin 2\left(q_2 + q_3\right) \\ &\quad \left. + 2l_2\left(m_3 l_3^* + m_g l_3\right)\sin\left(2q_2 + q_3\right)\right]\dot{q}_1^2, \end{aligned} \tag{26}$$

where J_{si} and J_{ni} are the moments of inertia of link i along its longitudinal and transverse axis passing through its center of mass, accordingly; m_i is the mass of link i; m_g is the mass of the load; l_i is the length of link i; l_i^* is the length from ith joint to the center of mass of ith link; g is the acceleration of gravity; and q_i, $\dot{q}_i$, $\ddot{q}_i$ are the angle of rotation of the link i and its derivatives.

The following set of parameters of MM was used for simulation: $l_1 = l_2 = l_3 = 0.5$ m, $l_1^* = 0.1$ m; $l_2^* = l_3^* = 0.2$ m, $m_g = 5$ kg; $m_1 = 25$ kg, $m_2 = m_3 = 5$ kg; $J_{s2} = 0.007$ kg·m^2, $J_{s3} = 0.005$ kg·m^2, $J_{n2} = 0.14$ kg·m^2, $J_{n3} = 0.2$ kg·m^2, $R = 0.5\,\Omega$, $L = 0.0005$ H, $k_\omega = 0.04$ V·s, $k_m = 0.04$ N·m/A, $J = 10^{-3}$ kg·m^2, $i_g = 100$; $k_u = 800$, $M_{\mathrm{cfg0}} = M_{\mathrm{cfm0}} = 0.02$ N·m, and $k_{vg} = k_{vm} = 0.005$ N·m·s/rad.

The input signals $q_1 = \sin(0.2t)$, $q_2 = u = \sin(0.5t)$, and $q_3 = \sin(t)$ were applied to the electric servo actuator. The parameters of feedback for observers were chosen as follows:

$$\begin{aligned} T_{01} &= 1.372 \times 10^{-6}, \\ T_{02} &= 2.564 \times 10^{-4}, \\ T_1 &= 3.704 \times 10^{-4}, \\ T_2 &= 0.033, \\ T_3 &= 0.033, \\ T_4 &= 100. \end{aligned} \tag{27}$$

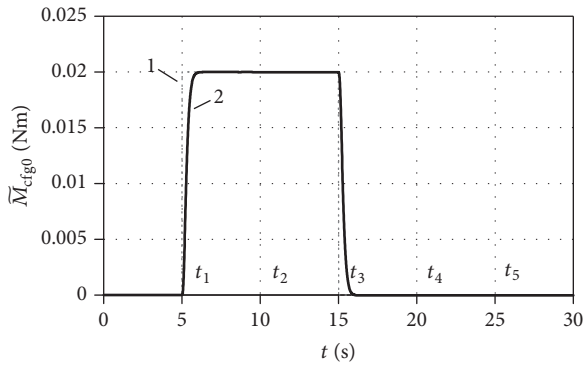


Figure 2: Estimation of constant fault $\widetilde{M}_{cfg0}$ value by proposed FI system.

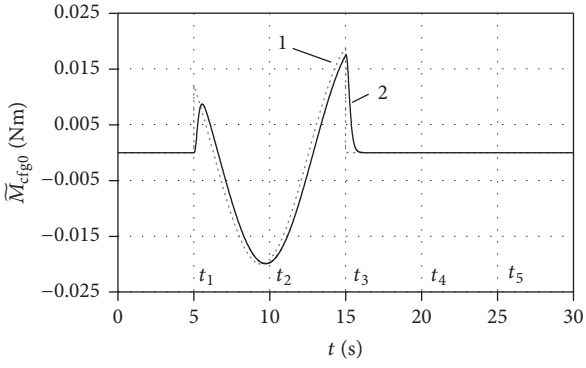


Figure 3: Estimation of variable fault $\widetilde{M}_{cfg0}$ value by proposed FI system.

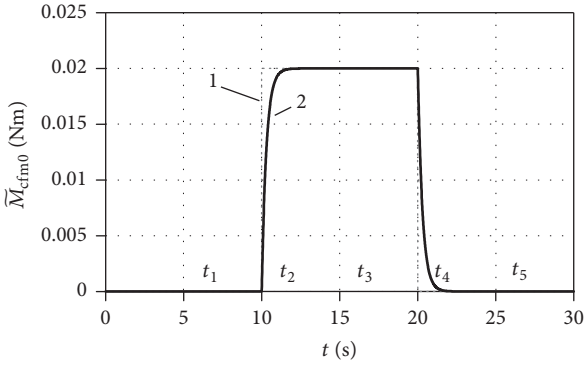


Figure 4: Estimation of constant fault $\widetilde{M}_{cfm0}$ value by proposed FI system.

Faults were presented by changing the following values:

(i) Coulomb friction coefficient in reducing gear by 50% ($\widetilde{M}_{cfg0} = 0.02$ N·m and $\widetilde{M}_{cfg0} = 0.02\sin(0.5t)$ N·m) at the time from $t_1 = 5$ s to $t_3 = 15$ s

(ii) Coulomb friction coefficient in DC motor by 50% ($\widetilde{M}_{cfm0} = 0.02$ N·m and $\widetilde{M}_{cfm0} = 0.02\sin(0.5t)$ N·m) at the time from $t_2 = 10$ s to $t_4 = 20$ s

(iii) Electrical resistance $\widetilde{R}$ by 50% ($\widetilde{R} = 0.2\,\Omega$ and $\widetilde{R} = 0.2\sin(0.5t)\,\Omega$) at the time from $t_3 = 15$ s to $t_5 = 25$ s.

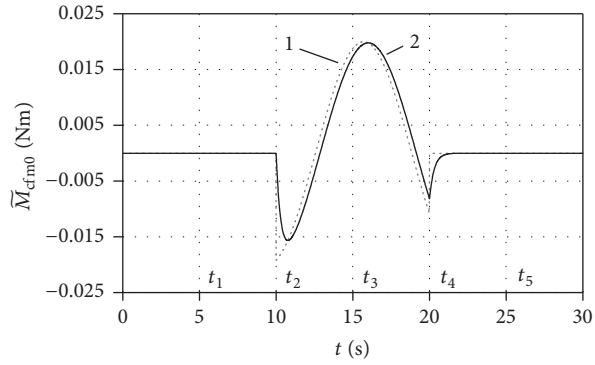


Figure 5: Estimation of variable fault $\widetilde{M}_{cfm0}$ value by proposed FI system.

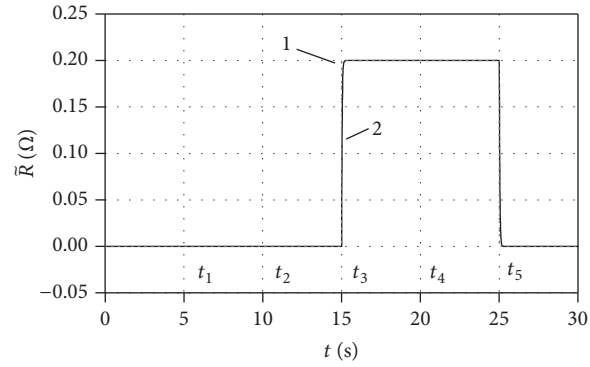


Figure 6: Estimation of constant fault $\widetilde{R}$ value by proposed FI system.

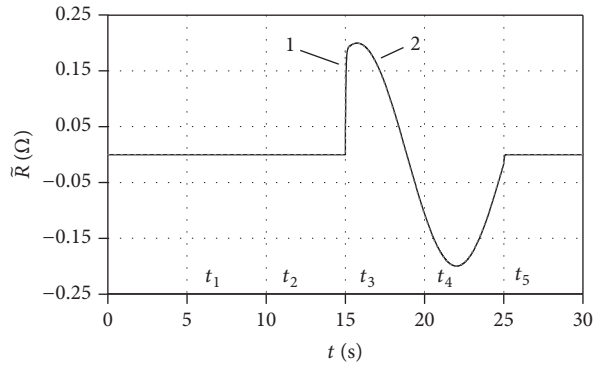


Figure 7: Estimation of variable fault $\widetilde{R}$ value by proposed FI system.

Values of introduced faults (curve 1) and estimated value of fault (curve 2) are shown in Figures 2–7. It follows from these figures that the constructed FI system provides accurate estimation of occurred fault. The results of simulation confirm the efficiency of the proposed method. Realization of the synthesized FI systems does not present any difficulties. These FI systems can be realized with the help of serial microprocessors.

6. Conclusion

In this paper, the method of synthesis of accurate FI systems for electric servo actuators of the MM, which are described by nonlinear differential equations with variable parameters and partially measured state vector, is suggested. This method consists in applying the logic-dynamic approach for synthesis of diagnostic observers guaranteeing detection and localization of possible faults and introducing of special feedback in diagnostic observers, allowing identifying values of faults. Efficiency of the proposed method was confirmed by the results of simulation. The subject of further research is application of the proposed method for synthesis of the fault tolerant control systems for actuators of the MM.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

These researches were supported by projects: Grant of President of Russia (MK-8536.2016.8) and Grants 16-29-04195 and 16-07-00300 of RFBR.

References

[1] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and fault-tolerant control*, Springer, Heidelberg, Third edition, 2016.

[2] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[3] F. L. Lewis, *Robot Manipulator Control Theory and Practice*, CRC Press, 2003.

[4] V. Filaretov, A. Zuev, D. Yukhimets, A. Gubankov, and E. Mursalimov, "The Automatization Method of Processing of Flexible Parts without their Rigid Fixation," *Procedia Engineering*, vol. 100, pp. 4–13, 2015.

[5] R. J. Patton, "Robust Model-Based Fault Diagnosis: The State of the ART," *IFAC Proceedings Volumes*, vol. 27, no. 5, pp. 1–24, 1994.

[6] R. Isermann and P. Balle, "Trends in the application of model based fault detection and diagnosis of technical processes," in *Proc. of the 13th World Congress IFAC*, p. 12, San Francisco, California, CA, USA, 1996.

[7] V. Filaretov, A. Zhirabok, A. Zuev, and A. Protcenko, "The new approach for synthesis of diagnostic system for navigation sensors of underwater vehicles," *Procedia Engineering*, vol. 69, pp. 822–829, 2014.

[8] V. Filaretov, A. Zuev, A. Gubankov, A. Procenko, and D. Yukhimets, "A new approach to synthesis of high speed and high reliable control systems for multilink manipulators," in *Proceedings of the 2016 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pp. 158–162, Tangerang, Indonesia, October 2016.

[9] R. Patton, P. Frank, and R. Clark, *Issue of Fault Diagnosis for Dynamic Systems*, Springer-Verlag, 2000.

[10] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Springer, 2001.

[11] F. Caccavale and L. Villani, *Fault Diagnosis and Fault Tolerance for Mechatronic Systems:Recent Advances*, Springer, Berlin, Heidelberg, 2003.

[12] V. F. Filaretov, "Synthesis of Adaptive Control Systems for Electric Servo Actuators of Manipulators," in *Proceedings of the Proc. of the 18th DAAAM Int. Symp., Intelligent Manufacturing Automation*, pp. 277–279, Zadar, Croatia, 2007.

[13] R. Seliger and P. Frank, "Fault diagnosis by disturbance decoupling nonlinear observers," in *Proceedings of the Proc. of the 30th Conf. on Decision and Control*, pp. 2248–2253, 1991.

[14] G. Schreier, J. Ragot, R. Patton, and P. Frank, "Observer design for class of nonlinear systems," in *Proceedings of the Proc. of the IFAC Symposium SAFEPROCESS'97*, pp. 498–503, Hull, UK, 1997.

[15] M. Staroswiecki and M. Comtet-Varga, "Fault detectability and isolability in algebraic dynamic systems," in *Proceedings of the CD ROM Proc. of the European Control Conference ECC'99*, Germany, Karlsrue, 1999.

[16] C. D. Persis and A. Isidori, "A geometric approach to nonlinear fault detection and isolation," *IEEE Transactions on Automatic Control*, vol. 46, no. 6, pp. 853–865, 2001.

[17] A. Shumsky and A. Zhirabok, "Nonlinear diagnostic filter design: algebraic and geometric points of view," *International Journal of Applied Mathematics and Computer Science*, vol. 16, no. 1, pp. 115–127, 2006.

[18] A. Zhirabok and S. Usoltsev, "Linear diagnosis methods for nonlinear systems," *Automation and Remote Control*, vol. 7, pp. 149–159, 2000.

[19] V. Filaretov, A. Zhirabok, A. Zuev, and A. Procenko, "Synthesis Method of Fault Tolerant Control System for Manipulators," *Advanced Materials Research*, vol. 717, pp. 551–556, 2013.

[20] V. F. Filaretov, A. N. Zirabok, A. V. Zuev, and A. A. Protsenko, "The Development of the Faults Accommodation System for Actuators of Multilink Manipulator," in *Proceedings of the Proc. of the 23rd International DAAAM Symposium*, pp. 575–578, Zadar, Croatia, 2012.

[21] L. Mironovsky, *Functional diagnosis of dynamic systems*, MSU-GRIF, St, Petersburg, Russia, 1998.

[22] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer-Verlag, Berlin, Germany, 2006.

[23] M. Staroswiecki, H. Yang, and B. Jiang, "Progressive accommodation of aircraft actuator faults," in *Proceedings of the IFAC Symposium SAFEPROCESS*, pp. 877–882, Beijing, China, 2006.

[24] A. Siqueira and M. Terra, "A fault tolerant manipulator robot based on h2, h∞ and mixed h2/h∞ markovian controls," in *Proceedings of the IEEE International Conference on Control Applications*, pp. 309–314, Taipei, Taiwan, 2004.

[25] V. Filaretov, A. Zhirabok, A. Zuev, and A. Protcenko, "Fault identification in nonlinear dynamic systems," in *Proceedings of the 2016 5th International Conference on Systems and Control (ICSC)*, pp. 273–277, Marrakesh, Morocco, May 2016.

[26] A. Zhirabok, "Fault detection and isolation: linear and nonlinear systems," in *Proceedings of the Proc. of the IFAC Symposium SAFEPROCESS97*, pp. 903–908, Hall, UK, 1997.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 9517385, 10 pages
https://doi.org/10.1155/2017/9517385

Research Article

A Bayesian Approach to Control Loop Performance Diagnosis Incorporating Background Knowledge of Response Information

Sun Zhou and Yiming Wang

Department of Automation, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Sun Zhou; zhousun@xmu.edu.cn

Received 20 June 2017; Accepted 3 August 2017; Published 28 September 2017

Academic Editor: Chunhui Zhao

To isolate the problem source degrading the control loop performance, this work focuses on how to incorporate background knowledge into Bayesian inference. In an effort to reduce dependence on the amount of historical data available, we consider a general kind of background knowledge which appears in many applications. The knowledge, known as response information, is about what faults can possibly affect each of the monitors. We show how this knowledge can be translated to constraints on the underlying probability distributions and introduced in the Bayesian diagnosis. In this way, the dimensionality of the observation space is reduced and thus the diagnosis can be more reliable. Furthermore, for the judgments to be consistent, the set of posterior probabilities of each possible abnormality that are computed from different observation subspaces is synthesized to obtain the partially ordered posteriors. The eigenvalue formulation is used on the pairwise comparison matrix. The proposed approach is applied to a diagnosis problem on an oil sand solids handling system, where it is shown how the combination of background knowledge and data enhances the control performance diagnosis even when the abnormality data are sparse in the historical database.

1. Introduction

Fault diagnosis is a topic of practical significance in process industries. In complex control loop systems, the control performance could be degraded due to various reasons [1]. Sensors and instrumentation problems are usually revealed as the systematic errors in mass or energy balance equations [2]. Poor control loop performance might also be due to changes in the model or modeling error [3].

One challenge of control loop diagnosis is that some similar evidences could be shared among different faults [4]. In a complex industrial control loop system, there may be lots of observations. For example, a large-scale industrial process includes thousands of process measurements [5]. Many diagnostic algorithms are often designed to identify specific components, while the faults may propagate and influence other components which are not to be detected [6]; thus, the methods could be influenced by possible faults in other components [7]. Moreover, all processes may run subject to uncertainty due to missing information, noise, and so on. Therefore, the occurrence of one fault may lead to flood of abnormal measurements and alarms and make it difficult to distinguish true underlying source.

Fault diagnosis methods for control loop systems may be classified into three categories, qualitative model-based methods, quantitative model-based methods, and data-driven methods [8–13]. However, these methods result in problems when multiple abnormalities have the same influence on the measurements. To deal with these problems, the methods were extended based on qualitative information about signs, magnitudes, and so on, to consider the direction and the magnitude of change [14, 15]. Also, Bayesian method has been proposed; for example, a Bayesian network was constructed from expert knowledge [16]. However, in these methods, the models are assumed to be known. Besides, fuzzy logic methods were proposed [17]. All these previous works rely on prior knowledge only.

To overcome the drawbacks of both quantitative and qualitative model-based diagnosis approaches, data-driven methods are developed [18]. For example, the support

vector machine (SVM) methods were proposed that take the diagnosis problem as a classification one [19]. Besides, multivariate statistical process monitoring methods were suggested [20]. Nevertheless, the problem sources of abnormality may not be explicitly identified by means of the variable contribution methods [21].

Then, a systematic probabilistic approach based on Bayesian inference was proposed that considers all possible abnormal observations. A Bayesian framework for control loop performance diagnosis was developed in [4]. The measurements from plenty of monitors are synthesized to generate a probabilistic result to diagnose the fault. Pernestal [22] proposed using Bayesian approaches to isolate faults in diesel engines. In a similar way, a data-based Bayesian approach is suggested in [23], to diagnose underlying sources of control performance degradation.

The main disadvantage with the data-driven or statistical approaches in diagnosis is that their performance relies heavily on the amount of available historical data. However, the requirement of sufficient training data is hardly met in diagnosis applications since faults are rare in normal processes. On the one hand, in their general form, they require sufficient historical samples from all faulty cases. However, in practical applications, there is only a limited amount of data available. On the other hand, the large number of monitors is a principle challenge for Bayesian diagnosis to be applied in industries. It is required in Bayesian inference to estimate the joint likelihood probability density of the observations from all monitors. In previous works, it is shown that the computational effort in estimating the probabilities grows exponentially with respect to the number of monitors [24, 25] That phenomenon is also regarded as the *curse of dimensionality*. It makes it difficult to correctly estimate the likelihood probability in more than five dimensions with practical sample sizes. These works using Bayesian methods are based on training data only, and no explicit background knowledge is integrated about the process under diagnosis.

There is also a general type of background knowledge available. In this paper, we consider incorporating the background knowledge together with the training data under the Bayesian framework in order to improve the diagnosis even if the historical data are insufficient with respect to the monitors number. Regarding the process knowledge, it is possibly known that one measurement in observation vector is equally distributed given different abnormalities. This type of knowledge is very general and can be formulated as constraints on the underlying likelihood probability distributions [22, 27]. It can express several types of process knowledge and appear in many diagnosis applications naturally.

In this paper, the background information is expressed in terms of response signature matrix (RSM). With a translation from RSM to the constraints of the marginal probability of the likelihoods, the background knowledge is explicitly taken into account in Bayesian control loop diagnosis. Moreover, we also suggest using a moving window method to consider a sequence of observation rather than a single observation in the diagnosis. In order to evaluate the proposed approach, we applied it to an oil sand solid handling system in such a case where only a few samples from abnormalities are available.

The rest of this paper is organized as follows. A description of the Bayesian control performance diagnosis problem is introduced first, and in Section 2 some terminologies are reviewed. In Section 3, the problem studied in this paper is stated formally, and in Section 4 Bayesian diagnosis evaluating multiple consecutive observations is presented. The computations of the posterior probabilities for different modes considering historical data only are presented first, and, the approach is extended to incorporate process knowledge in Section 5. In Section 6, the proposed approach is evaluated on the diagnosis problem on the oil sand solids handling system, using training and background knowledge. The conclusion is given in Section 7.

2. Preliminaries: Bayesian Diagnosis for Control Loop Systems

Before going into the details of the Bayesian diagnosis, some terminologies are introduced based on the definition proposed in [23].

Component. Assume that the process under diagnosis consists of some components which is possible to fail or not fail. In a typical control loop, the components can be sensors, actuators, controllers, process models, and so on [28]. Assume there are P components of interest. Each component may have several different states. For example, a sensor may consist of three states: unbiased, moderate biased, and severe biased.

Mode. A mode of the process is defined as an assignment of the states of all the components. It indicates the state of the system. For example, a mode can be as follows: {controller is normal, activator has severe stiction, process model has no error, and sensor has moderate bias}. Each mode represents the status under which the process is operating. It can be normal state or abnormal state. Assume the total number of the modes is J. If each component has n states, $J = n^P$. The mode can be considered as a random scalar variable described by M with values $\{m_j\}$, $j = 1, \ldots, J$.

Observation. There are some monitors, sensors, or add-on indirect measurements such as "ad hoc tests" conducted by engineers, model-based diagnostic tests, and monitoring algorithms that are designed to measure certain parameters. They can be represented by a general designation, *monitors*. the qth monitor is denoted by π_q. Assume there are L monitors; then $q = 1, \ldots, L$. Each monitor output or measurement signal can be represented by a discrete value; for example, low, medium, and high are 3 possible values. Define that *observation* is a L-dimension vector composed of the discrete values of all monitor outputs. These outputs may be preprocessed, for example, in diagnostic tests. The observation vector $\mathbf{X} = (X_1, \ldots, X_L)$ with the domain $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_L$. Denote an assignment of the observation vector by $\mathbf{X} = x_k$, $k = 1, \ldots, K$, where K is the number of different observations. If the lth monitor output has K_l discrete values, $K = \prod_{l=1}^{L} K_l$. Each value x_k $(k = 1, \ldots, K)$ is an L-dimensional vector, and we write $x_k = (x_k[1], \ldots, x_k[L])$ to explicitly

denote the elements. Consider the observation as a random variable.

Training Data. A training sample at time t consists of simultaneous values of the mode variable M and the observation vector X at that time. The value is denoted by $d^t = (x^t, m^t)$. All training samples collected from different modes of the system form the training dataset. A realization of training data is denoted by D. And D_{m_j} denotes the subset of training data entries where the underlying mode is m_j.

2.1. Bayesian Diagnosis. The Bayesian control loop diagnosis proposed in [23] is briefly reviewed in this section. Each component is possible to suffer from some abnormal operating conditions that may degrade the control performance. Also, any fault in one component may influence the monitors for the other components [4]. Consider there are certain probabilistic interconnections between problem causes and monitor outputs [4]. Bayesian inference is applied to compute the probability of mode variable M given a current observation X and the training observation data set D. The posterior probability of every operating mode can be computed based on Bayes' rule.

$$p(M \mid X, D) \propto p(X \mid M, D)\, p(M), \tag{1}$$

where $p(X \mid M, D)$ is the likelihood probability and $p(M)$ is the probability of mode M which is typically specified by a priori knowledge. The mode with the highest posterior probability can be determined as the underlying mode based on the *maximum a posteriori* (MAP) principle, and the related abnormality is generally regarded as the fault source.

Thus, the main issue to construct a Bayesian diagnostic system is estimating the likelihood in line with the training observation data D. Following the results of [22, 23], a Bayesian algorithm is presented for the likelihood estimation for control loop diagnosis.

3. Formal Problem Formulation

Consider that there is a general type of background information and multiple consecutive observations available. The task is to determine which fault(s) which has caused the measurements, given consecutive observations X, training data D, and background knowledge Q that is described as follows.

Background Knowledge in Terms of Probability Constraints. Background information usually comes from expert or process knowledge. It can be described as Q. It can be considered to consist of two parts of information. One specifies the prior probability for the modes, and the other defines that there are elements, representing the monitor outputs, in the observation vector which are equally distributed under different modes.

In addition, rather than considering a single observation as [23], assume that r consecutive observations are recorded and that the same fault is present during collection of these observations. Now, the fault diagnosis problem studied in this work can be stated formally as to compute

$$p\left(M^{J_r} = m_j \mid X^{J_1} = x_{k_1}, \ldots, X^{J_r} = x_{k_r}, D, Q\right), \tag{2}$$

that is, to compute the probabilities that each mode m_j is present at a time instant J_r, given the training dataset D, the background knowledge Q, and the observation values $X^{J_1} = x_{k_1}, \ldots, X^{J_r} = x_{k_r}$ from the control loop process under diagnosis. The subscripts on J are used to denote observation vectors from consecutive instants and those on k to enumerate the observation values.

In the following, the posterior probabilities of each mode given consecutive observations are calculated with the training data only.

4. Bayesian Inference Using Training Data Only

To solve the stated problem (2), a new method is proposed for learning the likelihood probability distribution. Before going into the details, first let us present a previous result on inference based on training data only.

According to Bayes' rule, to compute (2), the likelihood probability needs to be calculated:

$$\begin{aligned} &p\left(X^{J_1} = x_{k_1}, \ldots, X^{J_r} = x_{k_r} \mid M^{J_r} = m_j, D\right) \\ &\quad = p\left(X^{J_1} = x_{k_1}, \ldots, X^{J_r} = x_{k_r} \mid M^{J_r} = m_j, D_{m_j}\right), \end{aligned} \tag{3}$$

where D_{m_j} denotes the training data under mode $M = m_j$.

Assume that the likelihood of all possible values of observation under mode m_j is parameterized by Θ_{m_j},

$$\begin{aligned} &p\left(x_k \mid m_j, \Theta, Q\right) = \theta_{k|m_j}, \\ &\Theta_{m_j} = \left\{\theta_{1|m_j}, \ldots, \theta_{K|m_j}\right\}, \\ &\theta_{k|m_j} > 0, \ \sum_{k=1}^{K} \theta_{k|m_j} = 1. \end{aligned} \tag{4}$$

Let Ω_{m_j} denote the space of all the likelihood parameters when the mode $M = m_j$. Also, the prior probability of these parameters is Dirichlet distributed

$$f\left(\Theta_{m_j} \mid Q\right) = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_{k|m_j}\right)}{\prod_{k=1}^{K} \Gamma\left(\alpha_{k|m_j}\right)} \prod_{k=1}^{K} \theta_{k|m_j}^{\alpha_{k|m_j}-1}, \quad \alpha_{k|m_j} > 0. \tag{5}$$

It can be shown that the Dirichlet distribution is the only possible choice for $f(\Theta_{m_j} \mid Q)$ under certain, not very restrictive assumptions [29]. One attractive property of Dirichlet distribution is that it is conjugate to the multinomial distribution [30], and the distribution for the training samples is proportional to the multinomial distribution. This makes the computations particularly simple. Further, the parameters $\alpha_{1|m_j}, \ldots, \alpha_{K|m_j}$ of the distribution are required. $\Gamma(\cdot)$ is the gamma function. For real number x, $\Gamma(x) = (x-1)!$.

By marginalizing over all the likelihood parameters, we have

$$p\left(X^{J_1}=x_{k_1},\ldots,X^{J_r}=x_{k_r}\mid M^{J_r}=m_j,D_{m_j}\right)=\int_{\Omega_{m_j}}p\left(X^{J_1}=x_{k_1},\ldots,X^{J_r}=x_{k_r}\mid M^{J_r}=m_j,\Theta_{m_j},D_{m_j}\right)f\left(\Theta_{m_j}\mid m_j,D_{m_j}\right)d\theta_{m_j}. \tag{6}$$

For the first factor of the integral (6), given the likelihood parameters Θ_{m_j} and assuming that these observations from mode m_j are independent,

$$p\left(X^{J_1}=x_{k_1},\ldots,X^{J_r}=x_{k_r}\mid M^{J_r}=m_j,\Theta_{m_j},D_{m_j}\right)=p\left(x_{k_1}\mid m_j,\Theta_{m_j}\right)\cdot\ldots\cdot p\left(x_{k_r}\mid m_j,\Theta_{m_j}\right)=\theta_{k_1|m_j}\times\cdots\times\theta_{k_r|m_j}. \tag{7}$$

And for the second factor, following the derivation of [23], we can write

$$f\left(\Theta_{m_j}\mid m_j,D_{m_j}\right)=\frac{p\left(D_{m_j}\mid\Theta_{m_j},m_j\right)f\left(\Theta_{m_j}\mid m_j\right)}{p\left(D_{m_j}\mid m_j\right)}. \tag{8}$$

Further, the likelihood of training data subset D_{m_j} related to the operating mode m_j can be calculated as

$$p\left(D_{m_j}\mid\Theta_{m_j},m_j\right)=\prod_{k=1}^{K}\theta_{k|m_j}^{n_{k|m_j}}, \tag{9}$$

where $n_{k|m_j}$ is the number of training data samples with the observation x_k from the mode m_j.

Then combining (7)–(9) and substituting in (6), likelihood (3) can be obtained.

To introduce the consecutive observations, first some notations are needed. Let $\varepsilon_{\text{obs}}\subset\mathbb{X}$ denote the set of distinct values present in consecutive observations $\mathbf{X}^{J_{1:r}}=(X^{J_1},\ldots,X^{J_r})$, and let N_{x_k} be the total number of observations in $\mathbf{X}^{J_{1:r}}$ with the value x_k. Following [31], the likelihood probability is given by the expression

$$p\left(X^{J_1}=x_{k_1},\ldots,X^{J_r}=x_{k_r}\mid M^{J_r}=m_j,D,Q\right)=\frac{\prod_{x_k\in\varepsilon_{\text{obs}}}\prod_{m=0}^{N_{x_k}-1}\left(n_{k|m_j}+\alpha_{k|m_j}+m\right)}{\prod_{m=0}^{r-1}\left(N_{m_j}+A_{m_j}+m\right)}, \tag{10}$$

where $N_{m_j}=\sum_{k=1}^{K}n_{k|m_j}$ is the count of the hypothetical samples and $A_{m_j}=\sum_{k=1}^{K}\alpha_{k|m_j}$ is the count of training samples. Theorem A.1 in the appendix can be referred to for the derivation of (10).

Table 1: A priori response information.

Monitor	$M=m_0$ (no fault)	$M=m_1$	$M=m_2$	$M=m_3$	$M=m_4$	$\cdots$
π_1	—	0	0	r	0	
π_2	—	r	0	0	r	
π_3	—	0	r	0	0	

5. Bayesian Diagnosis Incorporating RSM and Data

To combine the background knowledge with training data, first, the problem dimensionality needs to be reduced utilizing the probability constraints implied in the available background information, and in the dimension-reduced subspaces, estimate the likelihoods with Bayesian inference. In this way, the estimation accuracy can be improved in the case of small amount of available historical samples that is often encountered in real applications since abnormalities are rare in normal process operations. Then, from the set of posterior probabilities that might be inconsistent as they are computed from different subspaces, derive the partially ordered posteriors that are consistent in the original probability space.

5.1. Background Knowledge Expressed as RSM. In many applications, there are only a few historical samples available. Therefore, the process knowledge must be explicitly handled. We consider a general type of process knowledge about what abnormalities can possibly affect each of the monitors. It can be expressed in terms of the following: "observation X_k has the same but maybe unknown probability distribution under $M=m_s$ and $M=m_t$."

Table 1 gives an example of such knowledge. "r" at the lth row and the jth column represents a response signature meaning that the lth element of the observation that is from the lth monitor is affected under abnormal mode m_j, compared with that under the normal mode. The likelihood distribution of the lth observation element given $M=m_j$ is different from that under the normal mode $M=m_j$. In other words, the lth monitor output would respond when the operating mode turns into the jth abnormal operating mode. And "0" in the table indicates that the likelihood probability distribution is the same as that under the fault free mode. Or to say, the lth monitor measurement shows zero response to the jth abnormal mode.

The matrix corresponding to the response information table is the Response Signature Matrix (RSM), denoted by $R=(r_{lj})_{L\times J}$, and use "1" for "r" in the matrix. For example, the RSM according to Table 1 is

$$R=\begin{pmatrix}0&0&0&1&0&\cdots\\0&1&0&0&1&\cdots\\0&0&1&0&0&\cdots\end{pmatrix}. \tag{11}$$

5.2. Dimensionality Reduction. Using training data only, the Bayesian diagnosis would suffer from the curse of dimensionality. In statistics, the phrase reflects the sparsity of data in multiple dimensions. That phenomenon is an inevitable problem in Bayesian diagnosis. For instance, when a process employs 20 monitors, each one having the same three states that are low, medium, and high, the total number of possible observation values is $3^{20} = 3.497 \times 10^9$. This large observation space requires substantially more data to estimate.

Consider any two operating modes m_s and m_t, and the domain of discourse consists only of these two modes.

$$p(m_s) + p(m_t) = 1. \tag{12}$$

For the lth monitor π_l, $r_{ls} = r_{lt} = 0$ indicates that the marginal probability distributions of the lth monitor output under the two modes are equal; that is,

$$p(X[l] = x_k[l] \mid m_s, Q) = p(X[l] = x_k[l] \mid m_t, Q). \tag{13}$$

Therefore, the lth monitor readings can be ignored, and it is possible to reduce the dimension by one. Whereas if $r_{ls} = 1$ or $r_{lt} = 1$, (13) does not hold. Thus, this measurement must be taken into account in the probability computation. Define

$$\gamma_{st} = \{l \in \{1, \ldots, L\} : r_{ls} = 1 \text{ or } r_{lt} = 1\} \tag{14}$$

as a set of numbers of monitors whose readings are affected by the sth or the tth abnormality, and d_{st} is the dimension of γ_{st}, or to say, the number of the elements of set γ_{st}. Given the background knowledge in terms of response information, d_{st} is usually smaller than L. Take the response information given in Table 1 as an example, we can obtain $\gamma_{12} = \{2\}$, $\gamma_{13} = \{3\}$, $\gamma_{14} = \{1\}$, $\gamma_{15} = \{2\}$, $\gamma_{23} = \{2, 3\}$, $\gamma_{24} = \{1, 2\}$, $\gamma_{25} = \{2\}$, $\gamma_{34} = \{1, 3\}$, $\gamma_{35} = \{2, 3\}$, and $\gamma_{45} = \{1, 2\}$.

Define $Z_r^{st} \triangleq (Z[l_1], \ldots, Z[l_{d_{st}}])$, where $l_1, \ldots, l_{d_{st}} \in \gamma_{st}$ is the observation element vector, each element of which represents a monitor output whose distribution is affected under mode m_s or m_t. Z_r^{st} is with the domain $\mathbb{Z}_{st} = \mathbb{X}_{l_1} \times \cdots \times \mathbb{X}_{l_{d_{st}}}$ that is a d_{st}-dimensional observation space, a subspace of $\mathbb{X}$. Also, define Z_{-r}^{st} as the observation vector whose probability is unaffected. For instance, $Z_r^{12} = (Z[2])$ and $Z_{-r}^{12} = (Z[1], Z[3])$. From (13), we have

$$p\left(Z_{-r}^{st} \mid m_s, Q\right) = p\left(Z_{-r}^{st} \mid m_t, Q\right). \tag{15}$$

Combining (12), it can be obtained that

$$p\left(Z_{-r}^{st} \mid Q\right) p\left(m_j \mid Q\right) = p\left(Z_{-r}^{st}, m_j \mid Q\right), \quad j = s, t. \tag{16}$$

(16) indicates that when only two modes instead of all J modes are considered, Z_{-r}^{st} is independent of mode variable. Then, it is easy to prove

$$\frac{p(X \mid m_s, Q)}{p(X \mid m_t, Q)} = \frac{p(Z_r^{st} \mid m_s, Q)}{p(Z_r^{st} \mid m_t, Q)}. \tag{17}$$

Therefore, while comparing the likelihoods of the observation under two modes, the monitor outputs corresponding to Z_{-r}^{st} can be ignored, and only those related to Z_r^{st} are needed to be taken into the probability computation. In this way, given the background information, the dimension of observation space can be reduced from L to d_{st}. Let K_{st} denote the total count of different observations in $\mathbb{Z}_{st}$. Note that in the following, Z_r^{st} is written as Z for simplicity.

In the d_{st}-dimensional subspace $\mathbb{Z}_{st}$, the likelihood can also be obtained applying Bayesian inference as follows.

First, consider the likelihood estimation given one observation. We want to compute

$$\begin{aligned} &p\left(Z = z_k \mid m_j, D, Q\right), \\ &z_k = \left(z_k[l_1], \ldots, z_k\left[l_{d_{st}}\right]\right) \in \mathbb{Z}_{st}, \quad j = s, t. \end{aligned} \tag{18}$$

Assume that the likelihood of all possible values of d_{st}-dimensional observation under mode m_j is parametrized by a set of parameters Θ_{m_j},

$$\begin{aligned} &p\left(z_k \mid m_j, \Theta, Q\right) = \theta_{k|m_j}, \\ &\Theta_{m_j} = \left\{\theta_{1|m_j}, \ldots, \theta_{K_{st}|m_j}\right\}, \\ &z_k \in \mathbb{Z}_{st}, \ \theta_{k|m_j} > 0, \ \sum_{k=1}^{K_{st}} \theta_{k|m_j} = 1, \ j = s, t. \end{aligned} \tag{19}$$

Also, the prior probability of these parameters is assumed to be Dirichlet distributed.

$$\begin{aligned} &f\left(\Theta_{m_j} \mid Q\right) = \frac{\Gamma\left(\sum_{k=1}^{K_{st}} \alpha_{k|m_j}\right)}{\prod_{k=1}^{K_{st}} \Gamma\left(\alpha_{k|m_j}\right)} \prod_{k=1}^{K_{st}} \theta_{k|m_j}^{\alpha_{k|m_j} - 1}, \\ &\alpha_{k|m_j} > 0, \ z_k \in \mathbb{Z}_{st}, \ j = s, t. \end{aligned} \tag{20}$$

Then, applying Bayesian inference, the likelihood (18) can be obtained as this expression

$$p\left(Z = z_k \mid m_j, D, Q\right) = \frac{n_{k|m_j} + \alpha_{k|m_j}}{N_{m_j} + A_{m_j}}, \quad j = s, t, \tag{21}$$

where $N_{m_j} = \sum_{k=1}^{K_{st}} n_{k|m_j}$ and $A_{m_j} = \sum_{k=1}^{K_{st}} \alpha_{k|m_j}$ are the count of hypothetical samples and training samples, respectively.

Now consider we have r consecutive observations $z_{k_1}, \ldots, z_{k_r} \in \mathbb{Z}_{st}$, define $\omega_{\text{obs}} \subset \mathbb{Z}_{st}$ as the set of distinct values presenting in the consecutive d_{st}-dimensional observations $\mathbf{Z}^{J_{1:r}} = (Z^{J_1}, \ldots, Z^{J_r})$, and N_{z_k} is the total number of observations in $\mathbf{Z}^{J_{1:r}}$ with value z_k. The sought likelihood can be obtained as

$$\begin{aligned} &p\left(Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r} \mid M^{J_r} = m_j, D, Q\right) \\ &= \frac{\prod_{z_k \in \omega_{\text{obs}}} \prod_{m=0}^{N_{z_k}-1} \left(n_{k|m_j} + \alpha_{k|m_j} + m\right)}{\prod_{m=0}^{r-1} \left(N_{m_j} + A_{m_j} + m\right)}, \quad j = s, t. \end{aligned} \tag{22}$$

Given the prior probabilities for the modes, the posterior probability can also be computed.

$$p\left(m_j \mid Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r}, D, Q\right) = \frac{p\left(Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r} \mid M^{J_r} = m_j, D, Q\right) p\left(m_j \mid Q\right)}{p\left(Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r} \mid D, Q\right)}, \quad j = s, t. \tag{23}$$

5.3. Consistent Partially Ordered Posteriors. For J modes, there are totally $(1/2)J(J-1)$ pairs of modes. As discussed in the last section, for each pair, the posteriors of each of the two modes can be obtained. However, the probability space contains only those two modes, not all the J modes; thus, these pairs of computed posteriors might be inconsistent.

Construct a pairwise comparison matrix **C**. To simplify the notation, write the posterior probability of mode m_j; that is, $p(m_j \mid Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r}, D, Q)$ as p_j. Let **C** be the set of $(J \times J)$ positive reciprocal matrix

$$\mathbf{C} = \left(c_{st}\right)_{J\times J} = \left(\frac{p_s}{p_t}\right)_{J\times J}, \tag{24}$$

where each entry c_{st} is the ratio of the posteriors of modes m_s and m_t. Therefore, the matrix is of the form

$$\mathbf{C} = \begin{pmatrix} 1 & \frac{p_1}{p_2} & \cdots & \frac{p_1}{p_J} \\ \frac{p_2}{p_1} & 1 & \cdots & \frac{p_2}{p_J} \\ \vdots & & \ddots & \\ \frac{p_J}{p_1} & \frac{p_J}{p_2} & \cdots & 1 \end{pmatrix}, \tag{25}$$

where $c_{st} = 1/c_{ts}$, $c_{ss} = 1$, $c_{st} > 0$, $s, t \in \{1, \ldots, J\}$. This comparison matrix consists of paired reciprocal comparisons based on (17). By definition (25), **C** is a positive reciprocal matrix.

Combining (22) and (23), we have

$$c_{st} = \frac{p_s}{p_t} = \frac{p\left(Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r} \mid M^{J_r} = m_s, D, Q\right) \times p\left(m_s \mid Q\right)}{p\left(Z^{J_1} = z_{k_1}, \ldots, Z^{J_r} = z_{k_r} \mid M^{J_r} = m_t, D, Q\right) \times p\left(m_t \mid Q\right)} \tag{26}$$

or in an equivalent form

$$c_{st} = \frac{\prod_{z_k \in \omega_{\text{obs}}} \prod_{m=0}^{N_{z_k}-1} \left(n_{k|m_s} + \alpha_{k|m_s} + m\right)}{\prod_{z_k \in \omega_{\text{obs}}} \prod_{m=0}^{N_{z_k}-1} \left(n_{k|m_t} + \alpha_{k|m_t} + m\right)} \times \frac{\prod_{m=0}^{r-1} \left(N_{m_t} + A_{m_t} + m\right)}{\prod_{m=0}^{r-1} \left(N_{m_s} + A_{m_s} + m\right)} \times \frac{p\left(m_s \mid Q\right)}{p\left(m_t \mid Q\right)}. \tag{27}$$

What are the priorities of the modes with respect to the posterior probability? Consider the consistency of the matrix **C**. **C** is consistent if $c_{ij}c_{jk} = c_{ik}$, $\forall i, j, k$. The original matrix **C** itself may be inconsistent. In order to determine which mode has the maximum probability, we need to derive a consistent partially ordered relationship set of all the modes $\{m_1, \ldots, m_J\}$ from the paired comparisons of the posteriors (maybe inconsistent) given in **C**.

There is a number of ways to obtain the vector of priorities. With emphasis on consistency, we suggest adopting an eigenvalue formulation [32]. Using this formulation, our problem becomes

$$\mathbf{C}\mathbf{w} = \lambda_{\max}\mathbf{w}, \tag{28}$$

where $\lambda_{\max}$ is the principal or largest eigenvalue of **C**. The principal eigenvector $\mathbf{w} = (w_1, \ldots, w_J)$; $w_1 > \cdots > w_J$ is the partially ordered vector for all J modes with respect to their posterior probabilities. It is easy to prove that, for any estimate $\mathbf{x}$,

$$\lim_{k\to\infty} \frac{1}{\lambda_{\max}^k} \mathbf{C}^k \mathbf{x} = a\mathbf{w}, \tag{29}$$

where $a > 0$ is a constant and $\mathbf{w}$ is principal eigenvector of **C**. The formula can be interpreted roughly as follows: "if we begin with an estimate and operate on it successively by $\mathbf{C}/\lambda_{\max}$ to get new estimates, the result converges to a constant multiple of the principal eigenvector."

Therefore, the mode corresponding to the largest element w_1 is the sought operating mode based on the MAP principle.

To sum up, following is the algorithm of the proposed diagnosis method for complex control loops that incorporates training data and background knowledge of response information.

(a) Based on process knowledge expressed as RSM, for each pair of modes m_s and m_t, obtain γ_{st} according to (14).

(b) For each pair of modes, in an observation subspace $\mathbb{Z}_{st}$ with respect to γ_{st}, compute the likelihood of each possible observation z_k under m_s and m_t, respectively, with (22).

(c) Construct the pairwise comparison matrix **C** with (25), (26), or (27).

(d) Compute the eigenvector $\mathbf{w}$ using (28), and the mode corresponding to the largest element w_1 is the sought operating mode.

6. Evaluation Results for Oil Sand Solids Handling System Diagnosis

6.1. Diagnostic Settings. We now consider solids handling system for evaluation. This system is the first stage of the oil sands process, which is a typical setup used for oil sands mining operations. The flowchart is presented in Figure 1 that is based on the industrial application [26]. From the flowchart, the mass of each truckload of the oil sand solid and the time when it was dumped into the dump hopper are available in a database. After being crushed, the solid is transported into the surge pile through the conveyor belt.

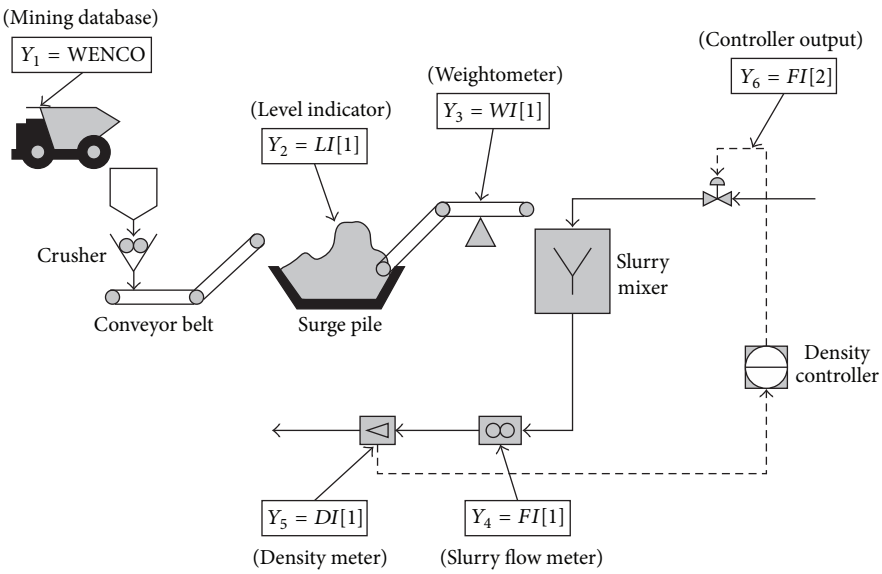


Figure 1: Oil sand solid slurry preparing system [26].

Table 2: Modes considered.

Mode	Mix box density model error	WENCO system instrument bias	Weightometer bias	Slurry flow meter bias	Density meter bias	Water valve stiction
m_1	—	—	—	—	—	—
m_2	*	—	—	—	—	—
m_3	—	*	—	—	—	—
m_4	—	—	*	—	—	—
m_5	—	—	—	*	—	—
m_6	—	—	—	—	*	—
m_7	—	—	—	—	—	*

A level indicator gives a reading, from 0% to 100%, of the relative level of the surge pile. A weightometer is on the mixer feed conveyor which feed oil sand from the surge pile to the slurry mixer. The slurry is prepared in this mixer by adding water to the oil sand. The amount of water is controlled by a slurry density controller. The controller output is the volumetric flow rate of water. A slurry flow meter and a density meter give the readings of the volumetric flow rate and the density of the effluent slurry, respectively.

In our simulation, four instruments Y_1 (database), Y_3 (weightometer), Y_4 (slurry flow meter), and Y_5 (density meter) are subject to possible bias. The control valve that is used to manipulate the water flow may suffer from stiction, and due to linearization, the model for slurry density controller is subject to error.

The system is designed to run under seven modes as shown in Table 2. The first mode m_1 is the *No Fault* mode. Each of the other six modes is under a fault of one component. Mode m_2 represents the density model error due to the linearization, m_3, m_4, m_5, and m_6 consider bias in each of the four instruments Y_1, Y_3, Y_4, and Y_5, and m_7 considers stiction of the water valve. In this table, a "—" denotes that the corresponding component is fault free and a "$*$" represents that the component has fault. Nine monitors are available for diagnosis, as shown in Table 3. 200 simulation runs were performed for each case, as well as 60 runs used for validation. As there are totally 9 monitors, the generic Bayesian diagnosis using training data only is a 9-dimensional problem. In other words, each likelihood probability given each underlying mode that is needed in the inference is a 9-dimensional joint probability. It is obvious that the available historical samples are very far from sufficient to generate accurate likelihood estimation.

The process knowledge of response information is given (Table 3). From this table, the corresponding RSM can be written that represents the implied probability constraints. According to (14), all γ_{st} sets for each value of s and t ($s, t \in \{1, 2, 3, 4, 5, 6, 7\}$) can be obtained. For instance, $\gamma_{23} = \{2, 4, 5, 6, 7\}$ with a reduced dimension $d_{st} = 5$. Then, for each pair of modes, the likelihood of each possible observation under each mode is computed, respectively, in a subspace with respect to γ_{st}. Finally, through the pairwise comparison matrix $\mathbf{C}$, the underlying operating mode can be determined using the eigenvalue formulation.

Table 3: Monitors and a priori response information.

Monitor	Description	m_1	m_2	m_3	m_4	m_5	m_6	m_7
π_1	Slurry model validation	—	0	0	r	r	0	r
π_2	Density flow model validation	—	r	0	r	0	r	0
π_3	Slurry density disturbance model validation	—	0	0	r	r	r	r
π_4	Data reconciliation residuals for Y_1	—	0	r	r	r	0	0
π_5	Data reconciliation residuals for Y_3	—	0	r	r	r	0	0
π_6	Data reconciliation residuals for Y_4	—	r	0	r	0	r	0
π_7	Data reconciliation residuals for Y_5	—	0	r	r	0	0	0
π_8	Data reconciliation residuals for the control valve input	—	0	0	r	r	r	r
π_9	Valve stiction monitor	—	0	0	0	0	r	0

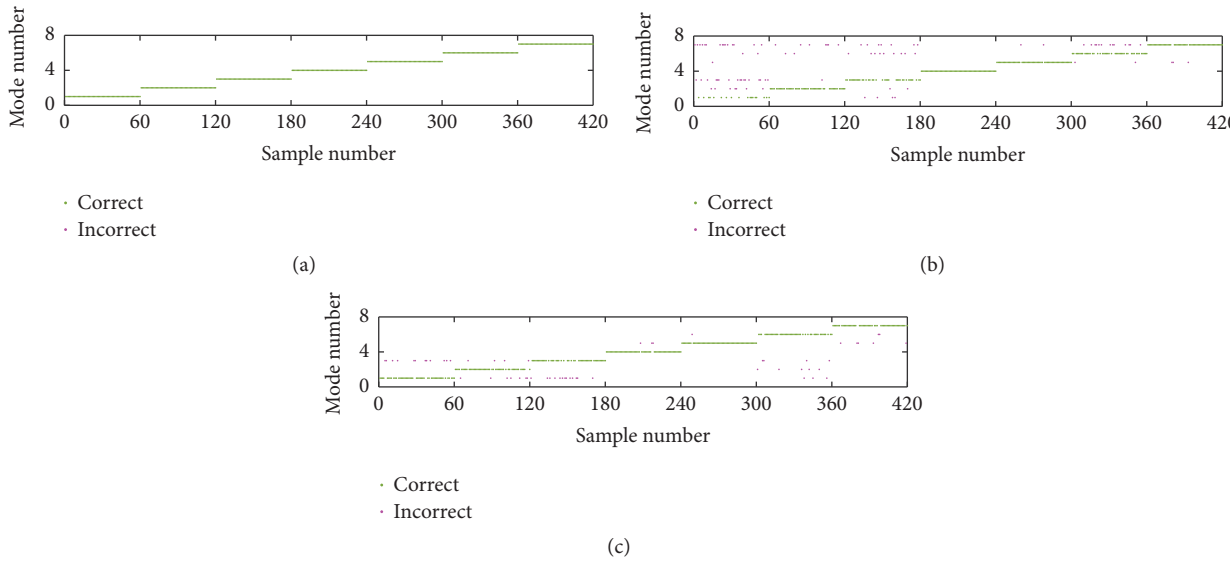


Figure 2: The underlying and diagnosed modes for each validating sample. (a) Underlying mode; (b) modes diagnosed using training data only; (c) modes diagnosed incorporating training data and response information.

6.2. Bayesian Diagnosis Using RSM and Historical Data. In order to evaluate diagnosis performance, two criteria are used. One is the false negative rate, which can be obtained using the simple quotient

$$F_N = \frac{n_{\text{inc}}}{n_{\text{inc}} + n_{\text{cor}}}, \tag{30}$$

where n_{inc} is the number of validating samples that are incorrectly diagnosed and n_{cor} is the number of samples that are correctly diagnosed. The misdiagnosis rate is related to mode number. In order to exclude the influence of the mode number, we define a *relative misdiagnosis rate* (RMR), and the aforementioned misdiagnosis rate will be referred to as *absolute misdiagnosis rate* (AMR). Assume that the underlying mode of a validating sample is m_j; calculate posteriors of all J modes. If there are t posteriors that are less than the posterior of the underlying mode $p(m_j \mid e_k)$, the correct diagnosis number for this sample will be $t/(J-1)$, and the incorrect diagnosis number be $1 - t/(J-1)$. Then n_{cor} and n_{inc} are obtained by adding up the correct and incorrect number of all validating samples, respectively. Finally, the RMR is obtained using the same quotient in (30). By such definition, when $p(m_j \mid x_k)$ is larger than posteriors of all other modes, it will be counted as one correct diagnosis; also, when $p(m_j \mid x_k)$ is larger than some other posteriors, still a positive fraction will be added to n_{cor}.

In order to mimic the background knowledge of response information, we obtained 10000 samples through 10000 runs of simulation and established the response information table based on the distribution of these samples.

The diagnosis performance of the proposed approach is evaluated in comparison with diagnosis using training data only.

In Figure 2, the horizontal axis represents the sample number. The underlying mode of the 1–60, 61–120, 121–180, 181–240, 241–300, 301–360, and 361–420 sample is m_1, m_2, m_3, m_4, m_5, m_6, and m_7, respectively. In Figure 2(a), the true underlying modes for each validating sample are shown, while in Figures 2(b) and 2(c), the modes diagnosed without and with incorporating of background knowledge are shown,

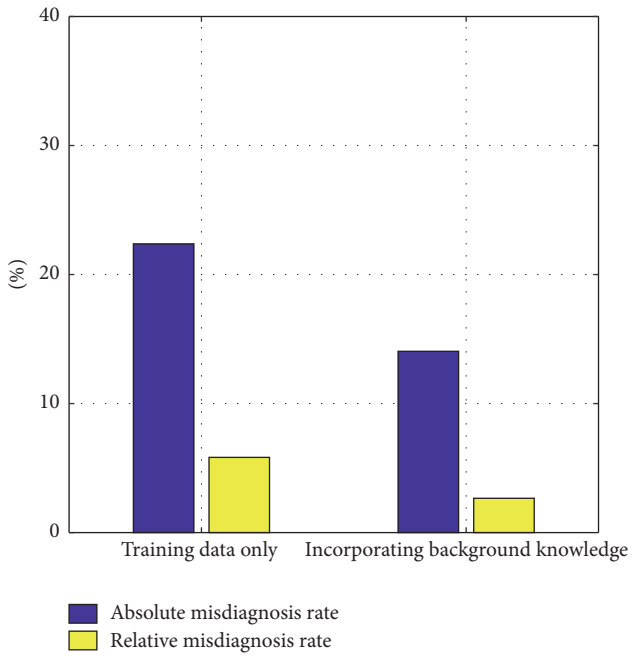


Figure 3: Average AMR and RMR from diagnosis with or without incorporating background knowledge.

respectively. Green points represent the samples that are correctly diagnosed, and pink points represent those incorrectly diagnosed. Not surprisingly, the proposed diagnosis approach incorporating background knowledge expressed as response information results in the better performance, while without use of the background knowledge, the diagnosis performs worse in the percentage of mode diagnosed.

The average AMR and the RMR from diagnosis with or without incorporating response information are shown in Figure 3. It is clearly observed that when combining the background knowledge together with the training data, diagnosis results are much better than when only training data is incorporated.

7. Conclusions

The objective is to isolate the problem source that is degrading the control performance. In order to reduce dependence on the amount of data available, our approach is to emphasize the use of background information and incorporate the background knowledge of response information into the diagnosis. The knowledge in general terms of RSM can be translated to constraints on the underlying probability distribution. We introduce the constraints in the Bayesian inference such that the dimensionality of the observation space is reduced, and thus the diagnosis can be enhanced. Moreover, for the comparative judgments to be consistent, the set of posterior probabilities computed from different observation subspaces is synthesized by using the eigenvalue formulation on pairwise comparison matrix; therefore, we can obtain the partially ordered posteriors and then determine the state of the process under diagnosis. The approach is applied to a diagnosis problem on an oil sand solids handling system. The advantage of combining background knowledge and data is achieved even when the amount of training data is limited. To sum up, training data and background knowledge are used for solving different parts of the control performance diagnosis problems. When both are used, the optimal diagnosis is achieved.

Appendix

Theorem A.1 (see [31]). *Let ε_{obs} and N_{x_k} be defined as in Section 4. Let M^{J_d} and X^{J_l}, $l = 1, \ldots, r$, be discrete variable, and let $\{1, \ldots, K\}$ be the domain of X^{J_l}. Let D denote training data. Introduce parameters Θ according to (4), and let the density $f(\Theta)$ be given by (5). Then it holds that*

$$p\left(X^{J_1} = x_{k_1}, \ldots, X^{J_r} = x_{k_r} \mid M^{J_r} = m_j, D, Q\right) = \frac{\prod_{x_k \in \varepsilon_{obs}} \prod_{m=0}^{N_{x_k}-1} \left(n_{k|m_j} + \alpha_{k|m_j} + m\right)}{\prod_{m=0}^{r-1} \left(N_{m_j} + A_{m_j} + m\right)}, \tag{A.1}$$

where $n_{k|m_j}$ is the number of samples in training data where the observation is $X^j = x_k$ when $M^j = m_j$, $N_{m_j} = \sum_{k=1}^{K} n_{k|m_j}$, and $A_{m_j} = \sum_{k=1}^{K} \alpha_{k|m_j}$.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was financially supported in part by the National Natural Science Foundation of China, nos. 61304141 and 61573296, the Natural Science Foundation of Fujian Province of China, no. 2014J01252, the Open Project of Key Laboratory of Advanced Process Control in Light Industry by the Ministry of Education of China, no. APCLI1602, and the Specialized Research Fund for the Doctoral Program of Higher Education of China, no. 20130121130004.

References

[1] K. Zhang, R. Gonzalez, B. Huang, and G. Ji, "Expectation-maximization approach to fault diagnosis with missing data," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 2, pp. 1231–1240, 2015.

[2] R. Gonzalez, B. Huang, F. Xu, and A. Espejo, "Dynamic bayesian approach to gross error detection and compensation with application toward an oil sands process," *Chemical Engineering Science*, vol. 67, no. 1, pp. 44–56, 2012.

[3] S. Ahmed, B. Huang, and S. L. Shah, "Validation of continuous-time models with delay," *Chemical Engineering Science*, vol. 64, no. 3, pp. 443–454, 2009.

[4] B. Huang, "Bayesian methods for control loop monitoring and diagnosis," *Journal of Process Control*, vol. 18, no. 9, pp. 829–838, 2008.

[5] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and

process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.

[6] H. Jiang, R. Patwardhan, and S. L. Shah, "Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix," *Journal of Process Control*, vol. 19, no. 8, pp. 1347–1354, 2009.

[7] F. Qi, *Bayesian approach for control loop diagnosis [Ph.D. dissertation]*, University of Alberta, Alberta, Canada, 2011.

[8] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis part I: quantitative model-based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.

[9] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, "A review of process fault detection and diagnosis: part II: qualitative models and search strategies," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 313–326, 2003.

[10] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis part III: process history based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003.

[11] K. Peng, B. Wang, and J. Dong, "An efficient quality-related fault diagnosis method for real-time multimode industrial process," *Journal of Control Science and Engineering*, vol. 2017, Article ID 9560206, 13 pages, 2017.

[12] Y. Sun and Z. Liang, "Fault diagnosis and fault tolerant control for non-Gaussian singular time-delayed stochastic distribution systems with disturbance based on the rational square-root model," *Journal of Control Science and Engineering*, Article ID 9563481, Art. ID 9563481, 11 pages, 2016.

[13] G. Mani and J. Jerome, "Intuitionistic fuzzy expert system based fault diagnosis using dissolved gas analysis for power transformer," *Journal of Electrical Engineering and Technology*, vol. 9, no. 6, pp. 2058–2064, 2014.

[14] B. Pulido, V. Puig, T. Escobet, and J. Quevedo, "A new fault localization algorithm that improves the integration between fault detection and localization in dynamic systems," in *Proceedings of 16th International Workshop on Principles of Diagnosis*, 2005.

[15] M. Daigle, X. Koutsoukos, and G. Biswas, "Multiple fault diagnosis in complex systems," in *Proceedings of 17th International Workshop on Principles of Diagnosis*, 2006.

[16] M. Askarian, R. Zarghami, F. Jalali-Farahani, and N. Mostoufi, "Fault diagnosis of chemical processes considering fault frequency via Bayesian network," *Canadian Journal of Chemical Engineering*, vol. 94, no. 12, pp. 2315–2325, 2016.

[17] X.-X. He, L.-T. Chen, and H.-T. Jia, "Building cognizance rule knowledge for fault diagnosis based on fuzzy rough sets," *Journal of Intelligent and Fuzzy Systems*, vol. 29, no. 6, pp. 2327–2333, 2015.

[18] F. Asghar, M. Talha, and S. H. Kim, "Neural network based fault detection and diagnosis system for three-phase inverter in variable speed drive with induction motor," *Journal of Control Science and Engineering*, vol. 2016, Article ID 1286318, 12 pages, 2016.

[19] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang, and C. Yan, "A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers," *Journal of Electrical Engineering and Technology*, vol. 12, no. 2, pp. 830–839, 2017.

[20] R. F. Sales, R. Vitale, S. M. de Lima, M. F. Pimentel, L. Stragevitch, and A. Ferrer, "Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy," *Computers and Chemical Engineering*, vol. 94, pp. 343–353, 2016.

[21] J. Yu and S. J. Qin, "Statistical MIMO controller performance monitoring. Part II: Performance diagnosis," *Journal of Process Control*, vol. 18, no. 3-4, pp. 297–319, 2008.

[22] A. Pernestal, *A Bayesian Approach to Fault Isolation with Application to Diesel Engines [M.S. thesis]*, School of Electrical Engineering, KTH, Stockholm, Sweden, 2007.

[23] F. Qi, B. Huang, and E. C. Tamayo, "A Bayesian approach for control loop diagnosis with missing data," *AIChE Journal*, vol. 56, no. 1, pp. 179–195, 2010.

[24] R. Gonzalez and B. Huang, "Data-driven diagnosis with ambiguous hypotheses in historical data: a generalized Dempter-Shafer approach," in *Proceedings of the 16th International Conference of Information Fusion (FUSION '13)*, pp. 2139–2144, July 2013.

[25] R. Gonzalez and B. Huang, "Control loop diagnosis with ambiguous historical operating modes: Part 2, information synthesis based on proportional parametrization," *Journal of Process Control*, vol. 23, no. 10, pp. 1441–1454, 2013.

[26] R. Gonzalez and B. Huang, "Control-loop diagnosis using continuous evidence through kernel density estimation," *Journal of Process Control*, vol. 24, no. 5, pp. 640–651, 2014.

[27] R. S. Niculescu, T. M. Mitchell, and R. . Rao, "Bayesian network learning with parameter constraints," *Journal of Machine Learning Research*, vol. 7, pp. 1357–1383, 2006.

[28] O. Nasri, I. Gueddi, P. Dague, and K. Benothman, "Spacecraft actuator diagnosis with principal component analysis: application to the rendez-vous phase of the mars sample return mission," *Journal of Control Science and Engineering*, vol. 2015, Article ID 204918, 11 pages, 2015.

[29] D. Geiger and D. Heckerman, "A characterization of the Dirichlet distribution through global and local parameter independence," *The Annals of Statistics*, vol. 25, no. 3, pp. 1344–1369, 1997.

[30] A. O'Hagan and J. Forster, *Kendall's Advanced Theory of Statistics*, Arnold, London, UK, 2004.

[31] A. Pernestal, *Probabilistic fault diagnosis-with automotive applications [Ph.D. dissertation]*, Linköping University, Linköping, Sweden, 2009.

[32] L. Wang, S. Sharkh, A. Chipperfield, and A. Cruden, "Dispatch of vehicle-to-grid battery storage using an analytic hierarchy process," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 2952–2965, 2017.

Hindawi
Journal of Control Science and Engineering
Volume 2017, Article ID 6354208, 9 pages
https://doi.org/10.1155/2017/6354208

Research Article

Fault Diagnosis of Nonlinear Uncertain Systems with Triangular Form

Qi Ding, Xiafu Peng, Xunyu Zhong, and Xiaoqiang Hu

Department of Automation, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Xiafu Peng; pengxiafu@sina.cn

Received 31 March 2017; Revised 14 June 2017; Accepted 28 June 2017; Published 1 August 2017

Academic Editor: Chunhui Zhao

A novel approach to fault diagnosis for a class of nonlinear uncertain systems with triangular form is proposed in this paper. It is based on the extended state observer (ESO) of the active disturbance rejection controller and linearization of dynamic compensation. Firstly, an ESO is designed to jointly estimate the states and the combination of uncertainty, faults, and nonlinear function of nonlinear uncertain systems. It can derive the estimation of nonlinear function via the state estimations and system model. Then, linearization of dynamic compensation is employed to linearize the system by offsetting nonlinear function mandatorily using its estimation. An observer-based residual generator is designed on the basis of the prior linearized model for fault diagnosis. Moreover, threshold treatment technique is adopted to improve the robustness of fault diagnosis. This method is utilizable and simple in construction and parameter tuning. And also we show the construction of ESO and give the corresponding convergence proof succinctly. Finally, a numerical example is presented to illustrate the validity of the proposed fault diagnosis scheme.

1. Introduction

In the past few decades, fault diagnosis technology has been advanced by leaps and bounds, and there has been significant research activity in the design and analysis of fault diagnosis [1–4]. Furthermore, the method based on analytic models is one of the most pioneer and in-depth approaches in fault diagnosis area (see, e.g., [5–14]). Model-based diagnosis can make great use of the system information and has a high diagnosis precision for fault location, fault type judgement, and amplitude estimation of fault. It is mainly applicable for linear systems, whereas in many practical situations, nonlinear properties of the control systems cannot be ignored in dealing with fault diagnosis problem. Therefore, it has practical significance in research fault diagnosis method for nonlinear systems based on analytic models [15].

At present, model-based fault diagnosis for nonlinear systems is generally based upon two kinds of basic theory: One way is to linearize the original nonlinear system near its working point using state feedback linearization technique to obtain the linear model. And the residual is designed based on unknown input decoupled technique to minimize the influence of the modeling error which is treated as unknown input. The other is that nonlinear observer method or nonlinear parameter estimation method is applied to fault diagnosis for nonlinear systems directly, without resort to any linearization. In a word, model-based fault diagnosis can be defined as the determination of faults of a system from the comparison of available system measurements with a prior information represented by the system's mathematical model through generation of residual quantities and their analysis [16]. A back-propagation neural network fault detection method was employed in [17] to detect the proton exchange membrane fuel cell (PEMFC) current fault type. According to the diagnosis results, the reconfiguration mechanism determined three nonlinear controllers based on feedback linearization approaches which are, respectively, built to adjust the voltage and pressure difference in the case of normal to be selected. The sliding mode observer for the nonlinear fault diagnosis was designed to observe the whole nonlinear system and actuator in [18]. Then the faults are diagnosed according to the residual value changes of two observers. In [15], the system model was decomposed by (B, K, ϕ) and the observer with high sensitivity for faults and strong robustness for unknown input disturbances was designed by using the decoupling technology. Reference [19]

proposed a new approach based on adaptive sliding mode observer to detect, isolate, and identify faults to deal with the shortcomings that existing model-based fault diagnosis methods of aircraft engines require high precision engine model. Reference [20] established a time-varying failure model of aeroengine control and designed an adaptive observer to estimate the state, actuator, and sensor faults of the system simultaneously. In recent years, various diagnosis methods based on the knowledge have been proposed, including neural network method, rough set-based method, genetic algorithm-based method, expert system method, and wavelet transform method. But all these methods have well-defined control limits [21].

The novel method broke through some limitations in this paper, compared with the above-mentioned methods, as assumption of measurable system variables and amplitude and type limitation about faults. It can be concluded from theoretical analysis and simulation results that the proposed method is easily used to identify the fault and get its approximate function, which differs from traditional methods. Reference [22] investigated fault diagnosis method based on ESO which estimated several faults directly. References [23, 24] also estimated fault directly using ESO to get its approximate function. Without doubt, this method took uncertainty as part of the fault in uncertain systems, which led to large error of estimated fault and actual fault. Reference [25] used the same technique, but here too there was the same issue about misinformation and inaccurate fault estimation. In short, this method has poor robustness.

In this paper, system variables are estimated to derive the estimation of nonlinear function using ESO with high tracking accuracy, high estimation precision, and less sensitivity to model errors and external disturbances. Then, linearization of dynamic compensation is employed to linearize the system via estimated value. And the residual system is constructed based on full-order state observer for the prior linearized model to get approximate function of fault. Moreover, in order to strengthen the robustness of fault diagnosis, this paper adopts threshold value treatment technique. Eventually the complex fault diagnosis problem of nonlinear system is simplified as one of linear system to perform the task of fault diagnosis, and the method is robust.

The rest of this paper is organized as follows. In Section 2, the problem of nonlinear fault diagnosis is formulated. The proposed fault diagnosis scheme is presented in Section 3. In Section 4, the derivation of threshold for robust fault detection is given. Section 5 describes a simulation example to illustrate the effectiveness of the robust fault diagnosis method. Finally, Section 6 presents some concluding remarks.

2. Problem Statement

Consider a class of uncertain nonlinear systems described by

$$\begin{aligned} \dot{\overline{x}}(t) &= \psi\left(\overline{x}(t), \omega(t)\right) + Bu(t) + B_f f(t), \\ y(t) &= C\overline{x}(t), \end{aligned} \tag{1}$$

where $\overline{x}(t) \in R^n$, $u(t) \in R$, and $y(t) \in R$ are the system state vector, input vector, and output vector, respectively. $f(t) \in R$ is the dynamic process fault of nonlinear system, and $\omega(t) \in R$ is the disturbance or uncertainty of the system. $\psi : R^n \to R^n$ is a real nonlinear vector function and has the form as follows:

$$\psi\left(\overline{x}(t), \omega(t)\right) = \begin{bmatrix} \psi_1\left(\overline{x}_1, \overline{x}_2\right) \\ \vdots \\ \psi_i\left(\overline{x}_1, \ldots, \overline{x}_{i+1}\right) \\ \vdots \\ \psi_{n-1}\left(\overline{x}_1, \ldots, \overline{x}_{n-1}\right) + \overline{x}_n \\ \psi_n\left(\overline{x}_1, \ldots, \overline{x}_n\right) \end{bmatrix}. \tag{2}$$

$B = [0, 0, 0, \ldots, b]^T_{1\times n}$ is the input matrix. $C = [1, 0, 0, \ldots, 0]_{1\times n}$ is the output matrix. In addition, $B_f = [0, 0, 0, \ldots, 1]^T_{1\times n}$ is the fault coefficient matrix.

There are many practical applications of such systems, such as Vander Pol oscillator system and robot arm system. Research on fault diagnosis for such systems has both theoretical and practical importance. It is shown in [26] that the coordinate change

$$\begin{aligned} x_1 &= \overline{x}_1, \\ \dot{x}_i &= x_{i+1}, \quad i = 1, 2, \ldots, n-1, \\ \dot{x}_n &= \phi\left(x_1, x_2, \ldots, x_n, t\right) + \omega(t) + bu(t) + f(t). \end{aligned} \tag{3}$$

transforms the system into the triangular form

$$\begin{aligned} \dot{x}_i &= x_{i+1}, \quad i = 1, 2, \ldots, n-1, \\ \dot{x}_n &= \phi\left(x_1, x_2, \ldots, x_n, t\right) + \omega(t) + bu(t) + f(t), \\ y &= x_1, \end{aligned} \tag{4}$$

where ϕ represents nonlinear function of the transformed system (4).

2.1. System Variables Estimation. In this paper, we estimate the system variables by designing ESO with high tracking accuracy and high estimation precision.

By defining a new system state in system (4) as $x_{n+1} = \phi(x_1, x_2, \ldots, x_n, t) + \omega(t) + f(t)$, the augmented system can be obtained as

$$\begin{aligned} \dot{x}_i &= x_{i+1}, \quad i = 1, 2, \ldots, n-1, \\ \dot{x}_n &= x_{n+1} + bu_0(t), \\ \dot{x}_{n+1} &= \dot{\phi} + \dot{\omega} + \dot{f} = \varphi(x, t), \\ y &= x_1. \end{aligned} \tag{5}$$

In order to track the system states $x_1, \ldots, x_n$ efficiently, in this work it is proposed to extend uncertainty, faults, and

nonlinear function as a new system state x_{n+1} and design an ESO as follows:

$$\begin{aligned}
\widetilde{x}_1 &= z_1(t) - y(t),\\
\dot{z}_i &= z_{i+1}(t) - \beta_i g_i(\widetilde{x}_1), \quad i = 1, 2, \ldots, n-1,\\
&\vdots\\
\dot{z}_n &= z_{n+1}(t) - \beta_n g_n(\widetilde{x}_1) + bu(t),\\
\dot{z}_{n+1} &= -\beta_{n+1} g_{n+1}(\widetilde{x}_1),
\end{aligned} \tag{6}$$

where $z(t) \in R^{n+1}$ and $u(t) \in R$ are ESO state and input. β_i $(i = 1, \ldots, n+1)$ are observer parameters, and $\beta_1 = 1/h, \ldots, \beta_i = (0.2 + 10^{2-i})/h^i$ $(i = 2, \ldots, n+1)$. Based on the experience, the observer parameter can be roughly selected and adjusted in the light of Fibonacci sequence. h is sampling period. Also the function $g(\widetilde{x}_1)$ can be expressed as

$$g_i(\widetilde{x}_1) = \begin{cases} |\widetilde{x}_1|^a \operatorname{sign}(\widetilde{x}_1), & |\widetilde{x}_1| > \delta\\ \dfrac{\widetilde{x}_1}{\delta^{1-a}}, & |\widetilde{x}_1| \le \delta, \end{cases} \tag{7}$$

$$(i = 1, 2, \ldots, n+1),$$

where $a \in (0, 1)$ and δ is range length of linear interval which is of a minor amount.

Let $\widetilde{x}_i = x_i - z_i$, $i = 1, 2, \ldots, n+1$. From (5) and (6), the observer estimation error can be shown as

$$\begin{aligned}
\dot{\widetilde{x}}_1 &= \widetilde{x}_2 + \beta_1 g_1(\widetilde{x}_1),\\
&\vdots\\
\dot{\widetilde{x}}_{n-1} &= \widetilde{x}_n + \beta_{n-1} g_{n-1}(\widetilde{x}_1),\\
\dot{\widetilde{x}}_n &= \widetilde{x}_{n+1} + \beta_n g_n(\widetilde{x}_1),\\
\dot{\widetilde{x}}_{n+1} &= \varphi(x, t) + \beta_{n+1} g_{n+1}(\widetilde{x}_1).
\end{aligned} \tag{8}$$

Theorem 1. *Denoting the Jacobian matrix of system (8) as $A_1(\widetilde{x})$, let $F(\widetilde{x}) = \dot{\widetilde{x}}$; that is, $A_1(\widetilde{x}) = \partial F/\partial \widetilde{x}$. If there exist two symmetric positive definite matrices P_1, Q_1, such that $\forall \widetilde{x} \neq 0$, the matrix $S(\widetilde{x}) = A_1^T P_1 + P_1 A_1 + Q_1$ is negative semidefinite in some neighborhood Ω of the origin. Then system (8) is asymptotically stable.*

Proof. For the symmetric positive definite matrices P_1, Q_1, let us consider a Lyapunov function candidate

$$V_1(\widetilde{x}) = F^T P_1 F. \tag{9}$$

Assuming that $A_1(\widetilde{x})$ is singular, then one can find a nonzero vector y_0 such that $A_1(\widetilde{x}) y_0 = 0$. Since

$$y_0^T S y_0 = y_0^T A_1^T P_1 y_0 + y_0^T P_1 A_1 y_0 + y_0^T Q_1 y_0 \tag{10}$$

the singularity of A_1 implies that $y_0^T A_1^T P_1 y_0 + y_0^T P_1 A_1 y_0 = 0$; we can get

$$y_0^T S y_0 = y_0^T Q_1 y_0 \tag{11}$$

which contradicts the assumed negative definiteness of S.

The invertibility and continuity of A_1 guarantee that the function $F(\widetilde{x})$ can be uniquely inverted. This implies that $F(\widetilde{x}) \neq 0$ for $\widetilde{x} \neq 0$. Given the above result, the function $V_1(\widetilde{x}) = F^T P_1 F$ is positive definite.

The derivative of V_1 can be written

$$\dot{V}_1(\widetilde{x}) = \dot{F}^T P_1 F + F^T P_1 \dot{F}. \tag{12}$$

Using the fact that $\dot{F} = A_1 F$, we can get

$$\dot{V}_1(\widetilde{x}) = F^T A_1^T P_1 F + F^T P_1 A_1 \dot{F} = F^T S F - F^T Q_1 F. \tag{13}$$

Because S is negative semidefinite and Q_1 is positive definite, $\dot{V}_1$ is negative definite. If $V_1(\widetilde{x}) \to \infty$ as $\|\widetilde{x}\| \to \infty$, system (8) is globally asymptotically stable based on Lyapunov theorem.

Note that as long as we select appropriately parameters $\beta_1, \ldots, \beta_{n+1}$, ESO can estimate both states $x_1(t), \ldots, x_n(t)$ of system (4) and the extended state $x_{n+1}(t) = \phi + \omega + f$ very well; that is,

$$\begin{aligned}
&z_1(t) \longrightarrow x_1(t), \ldots, z_n(t) \longrightarrow x_n(t),\\
&z_{n+1}(t) \longrightarrow x_{n+1}(t) = \phi(x_1, x_2, \ldots, x_n, t) + \omega(t) + f(t).
\end{aligned} \tag{14}$$

Also ESO is evolved by the optimal control, and large quantities engineering experiences from practice show that ESO is convergent. □

2.2. Linearization Process. The estimation of nonlinear function $\widehat{\phi}(z_1, z_2, \ldots, z_n, t)$ can be derived and the extended state estimation $z_{n+1}(t)$ can be got with n system state estimations and system model. On this basis we substitute the estimation of nonlinear function for its value and then select appropriately control variable to achieve linearization of dynamic compensation. In this work, the control variable is given by

$$u(t) = u_0 - \frac{\widehat{\phi}}{b}; \tag{15}$$

then nonlinear system (4) is turned into linear integrator series control system (16); that is,

$$\begin{aligned}
\dot{x}_i &= x_{i+1}(t), \quad i = 1, 2, \ldots, n-1,\\
\dot{x}_n &= bu_0(t) + \varpi(t) + f(t),\\
y &= x_1,
\end{aligned} \tag{16}$$

where $\varpi(t)$ is the sum of not eliminated uncertainty and estimation errors.

It is easy to know that system state estimations contain disturbing information to make uncertainty weakened by linearization of dynamic compensation.

3. Fault Diagnosis

As mentioned above, nonlinear system (4) is turned into linear integrator series control system (16) by linearization of dynamic compensation which simplifies fault diagnosis of nonlinear system into that of linear system, whereafter a full-order fault detection filter is designed for linearized system,

achieving convergence of observation error as well as fault diagnosis based on residual error system. Linear integrator series control system (16) can be uniformly expressed in state equations of general forms as, distinctly,

$$\begin{aligned} \dot{x} &= Ax(t) + Bu_0(t) + B_\varpi \varpi(t) + B_f f(t), \\ y &= Cx, \end{aligned} \tag{17}$$

with

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{n\times n},$$

$$B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ b \end{bmatrix}_{n\times 1}, \tag{18}$$

$$C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_{1\times n},$$

$$B_\varpi = B_f = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1},$$

and (A, C) is an observable pair which is verifiable by $\operatorname{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}_{n\times n} = n$.

Based on (17), an observer-based residual generator can be designed as

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x}(t) + Bu_0(t) + L\left(y(t) - \hat{y}(t)\right), \\ \hat{y} &= C\hat{x}(t), \end{aligned} \tag{19}$$

where L is the gain matrix which needs to be set.

Defining the state estimation and the residual, $e(t) = x(t) - \hat{x}(t)$ and $\varepsilon(t) = y(t) - \hat{y}(t)$, respectively, the dynamics of the residual generator are governed by

$$\begin{aligned} \dot{e}(t) &= (A - LC)e(t) + B_\varpi \varpi(t) + B_f f(t), \\ \varepsilon(t) &= Ce(t). \end{aligned} \tag{20}$$

Denote the sum of fault and error by $f_\varpi(t) = f(t) + \varpi(t)$. Note that $f_\varpi(t)$ is input of the residual generator, so the relation between residual $\varepsilon(t)$ and input $f_\varpi(t)$ can be definitely obtained to perform the task of fault diagnosis. In addition, the design and analysis in this paper are based on the assumptions as follows:

(A1) The real time action $\|\phi\|$ is bounded.

(A2) b is a known constant.

(A3) $\varpi(t)$ remains bounded before and after the occurrence of a fault; that is, $\|\varpi(t)\| \le \lambda_\varpi$.

With this method, the fault diagnosis problem can be transformed into the problem of pole placement. Now the object is to compute the gain L to make residual generator asymptotically stable and fulfill the following performance indicators:

(A1) Residual generator (20) is asymptotically stable when no fault occurs.

(A2) For the required property $\tau > 0$, transfer function $G(s)$ between the sum $f_\varpi(t)$ and residual signal $\varepsilon(t)$ satisfies inequality $\|G(s)\| > \tau$ when a fault occurs.

Theorem 2. *Consider the dynamics of residual generator (20) rewritten by*

$$\begin{aligned} \dot{e}(t) &= (A - LC)e(t) + B_\varpi \varpi(t) + B_f f(t), \\ \varepsilon(t) &= Ce(t); \end{aligned} \tag{21}$$

under the assumption given above, suppose the expected eigenvalues of filter (19) are λ_i^* $(i = 1, 2, \ldots, n)$. If there exist two symmetric positive definite matrices P, Q and a state feedback matrix $\overline{K}$ verifying Lyapunov equation $(A - LC)^T P + P(A - LC) = -Q$ and $\lambda_i(A - C^T\overline{K}) = \lambda_i^*$, respectively, then residual generator (19) is asymptotically stable, and also $L = \overline{K}^T$.

Proof. For the symmetric positive definite matrices P, Q, let us consider a quadratic Lyapunov function candidate

$$V(t) = e^T P e. \tag{22}$$

If there are no fault and uncertainty, that is, $f(t) = 0$ and $\varpi(t) = 0$, we can get $\dot{V}(t)$ based on (20)

$$\begin{aligned} \dot{V}(t) &= \dot{e}^T P e + e^T P \dot{e} \\ &= e^T (A - LC)^T P e + e^T P (A - LC) e \\ &= e^T \left[(A - LC)^T P + P(A - LC)\right] e = -e^T Q e. \end{aligned} \tag{23}$$

Then we get $\dot{V}(t) < 0$. Since $V(t) \to \infty$ as $\|e(t)\| \to \infty$, the residual generator (19) is globally asymptotically stable based on Lyapunov theorem.

As previously stated, (A, C) is an observable pair. Thus, integrator series control system (16) is completely observable. And all the eigenvalues of residual generator (19) can be configured arbitrarily based on pole placement theorem; that is, there exists a state feedback matrix $\overline{K}$ such that $\lambda_i(A - C^T\overline{K}) = \lambda_i^*$.

Based on duality principle, the observability of (A, C) is equivalent to the controllability of (A^T, C^T). Computing dual coefficient matrices $\overline{A} = A^T$, $\overline{B} = C^T$, pole assignment method is employed to calculate state feedback matrix $\overline{K}$ with $(\overline{A}, \overline{B})$ and the expected eigenvalues λ_i^* $(i = 1, 2, \ldots, n)$.

Therefore, the gain matrix is determined according to the conditions of convergence of the monitor and pole assignment method; that is, $L = \overline{K}^T$. □

Remark 3. Let us denote the eigenvalues of the matrix $A - C^T\overline{K}$ by $\lambda_i(A - C^T\overline{K})$. The equality $\lambda_i(A - C^T\overline{K}) = \lambda_i^*$ means that all the eigenvalues of filter (19) can be configured arbitrarily.

From the dynamics of the residual generator (20), we can transfer function $G(s)$ between the sum $f_\varpi(t)$ and residual signal $\varepsilon(t)$ as follows:

$$\begin{aligned} G(s) &= \frac{\varepsilon(s)}{F_\varpi(s)} = C(sI - A + LC)^{-1} B_f \\ &= \frac{E_{n-1}s^{n-1} + E_{n-2}s^{n-2} + \cdots + E_0}{\alpha(s)}, \end{aligned} \tag{24}$$

where

$$\begin{aligned} &\alpha(s) = \det(sI - A + LC) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s \\ &\quad + \alpha_0, \\ &E_{n-1} = CB_f, \\ &E_{n-2} = CAB_f + \alpha_{n-1}CB_f, \\ &\qquad\vdots \\ &E_1 = CA^{n-2}B_f + \alpha_{n-1}CA^{n-3}B_f + \cdots + \alpha_2 CB_f, \\ &E_0 = CA^{n-1}B_f + \alpha_{n-1}CA^{n-2}B_f + \cdots + \alpha_1 CB_f. \end{aligned} \tag{25}$$

Then we have

$$\begin{aligned} F_\varpi(s) &= \frac{\varepsilon(s)}{C(sI - A + LC)^{-1} B_f} \\ &= \frac{\varepsilon(s)\cdot\alpha(s)}{E_{n-1}s^{n-1} + E_{n-2}s^{n-2} + \cdots + E_0}. \end{aligned} \tag{26}$$

By applying the inverse Laplace transformation to (26), we have

$$\begin{aligned} f_\varpi(t) &= L^{-1}\left[\frac{\varepsilon(s)}{C(sI - A + LC)^{-1} B_f}\right] \\ &= L^{-1}\left[\frac{\varepsilon(s)\cdot\alpha(s)}{E_{n-1}s^{n-1} + E_{n-2}s^{n-2} + \cdots + E_0}\right]. \end{aligned} \tag{27}$$

The above equation shows the definite relationship between the sum $f_\varpi(t)$ and residual signal $\varepsilon(t)$. Also Theorem 2 indicates that if there is no fault, filter (19) can accurately track the state and output of system (17). That is to say, the residual signal is asymptotically stable at zero. When a fault occurs, there is a sudden change in the state and output of system (17), but filter (19) cannot track them timely. Thus, the residual signal is large at first which can be used to detect fault. From (27), we can approximately estimate the fault $f(t)$ based residual signal $\varepsilon(t)$ to identify fault. Now the comprehensive analysis proves that the proposed scheme is effective for fault diagnosis of system (1).

In addition, linearization of dynamic compensation ensures that the uncertainties would not interfere with fault diagnosis, because the uncertainties are offset by its estimation. References [23, 24] estimated fault directly using ESO to get its approximate function. Without doubt, this method took uncertainty as part of the fault in uncertain systems, which led to large error of estimated fault and actual fault such that the result of fault diagnosis is inaccurate. Compared with [23, 24], the method in this paper is complex to some extent. But the proposed fault diagnosis method based on linear system is mature and applicable. The most important thing is that the proposed method is more robust.

4. Threshold Analysis

Robust technique is employed in this paper to enhance the reliability of fault detection and reduce the false alarm rate. There are two kinds of robust techniques: the unknown input decoupled observer technique and the limited threshold technique. For uncertainties of the system, threshold treatment technique is adopted to realize robust fault evaluation in the algorithm. When the residual signal exceeds the defined threshold, we can identify that there exists fault based on the mechanism. Otherwise, there is no fault.

Theorem 4. *For system (4), ESO (6) is designed to estimate the states, and also the observer-based residual generator (19) is designed to diagnose fault. If there is no fault, the residual signal satisfies the following inequality for the time domain* $t \in [t_1, t_2]$*:*

$$\|\varepsilon(t)\| \le \lambda_1 T \lambda_\varpi. \tag{28}$$

Proof. By defining the state transfer matrix $\Theta(t)$, then we can get

$$\Theta(t,\tau) = e^{(A-LC)(t-\tau)}. \tag{29}$$

For $t \in [t_1, t_2]$, we have

$$\begin{aligned} e(t) &= \Theta(t,t_1)\, e(t_1) + \int_{t_1}^{t} \Theta(t,\tau) B_f f(\tau)\, d_\tau \\ &\quad + \int_{t_1}^{t} \Theta(t,\tau) B_\varpi \varpi(\tau)\, d_\tau. \end{aligned} \tag{30}$$

At the initial time t_1, state estimations are the same as true values; that is, $\widehat{x}(t_1) = x(t_1)$; that is, $e(t_1) = 0$.

Based on norm theory, we obtain

$$\begin{aligned} \|e(t)\| \le &\left\| \int_{t_1}^{t} \Theta(t,\tau) B_f f(\tau)\, d_\tau \right\| \\ &+ \left\| \int_{t_1}^{t} \Theta(t,\tau) B_\varpi \varpi(\tau)\, d_\tau \right\| \end{aligned}$$

$$\leq \int_{t_1}^{t} \left\| \Theta(t,\tau) B_f f(\tau) \right\| d_\tau + \int_{t_1}^{t} \left\| \Theta(t,\tau) B_\varpi \varpi(\tau) \right\| d_\tau. \tag{31}$$

By substituting the above inequality into (20), then we have

$$\begin{aligned} \|\varepsilon(t)\| &= \|Ce(t)\| \\ &\leq \int_{t_1}^{t} \left\| C\Theta(t,\tau) B_f f(\tau) \right\| d_\tau \\ &\quad + \int_{t_1}^{t} \left\| C\Theta(t,\tau) B_\varpi \varpi(\tau) \right\| d_\tau \\ &= \int_{t_1}^{t} \left\| C\Theta(t,\tau) B_\varpi \varpi(\tau) \right\| d_\tau \leq \lambda_1 T \lambda_\varpi, \end{aligned} \tag{32}$$

where T is the length of the time domain; that is, $T = t_2 - t_1$. And

$$\lambda_1 = \sup_{t,\tau \in [t_1,t_2]} \left\| C\Theta(t,\tau) B_\varpi \right\|. \tag{33}$$

Therefore the threshold Γ can be set as $\lambda_1 T \lambda_\varpi$; that is, $\Gamma = \lambda_1 T \lambda_\varpi$. Then the fault detection can be performed using the following mechanism:

$$\begin{aligned} &\|\varepsilon(t)\| \leq \Gamma, \quad \text{there is no fault,} \\ &\|\varepsilon(t)\| > \Gamma, \quad \text{there is a fault.} \end{aligned} \tag{34}$$

□

5. Example and Simulation

Consider the following nonlinear system with a fault:

$$\begin{aligned} \dot{x}_1 &= x_2(t), \\ \dot{x}_2 &= \cos 0.6t \cdot x_1(t) + \cos 0.7t \cdot x_2(t) \\ &\quad + 0.5 \operatorname{sign}(\sin t) + u(t) + \omega(t) + f(t), \\ y &= x_1. \end{aligned} \tag{35}$$

From (35), we can know that nonlinear function is

$$\phi(x_1, x_2, t) = \cos 0.6t \cdot x_1(t) + \cos 0.7t \cdot x_2(t) + 0.5 \operatorname{sign}(\sin t). \tag{36}$$

The combination of uncertainty, fault, and nonlinear function $\Phi(x_1, x_2, t) = \phi(x_1, x_2, t) + \omega(t) + f(t)$ is tracked by ESO which can be designed based on system (6). And so are the states x_1, x_2. According to the estimations of states and system structure, the estimation of nonlinear function $\widehat{\phi}(\widehat{x}_1, \widehat{x}_2, t)$ can be obtained.

In the simulation example, $\omega(t)$ is white noise with an energy of 0.01, permissible estimating error $\gamma = 0.8$, time-domain length $T = 5$, and the fault detection threshold $\Gamma = 5$.

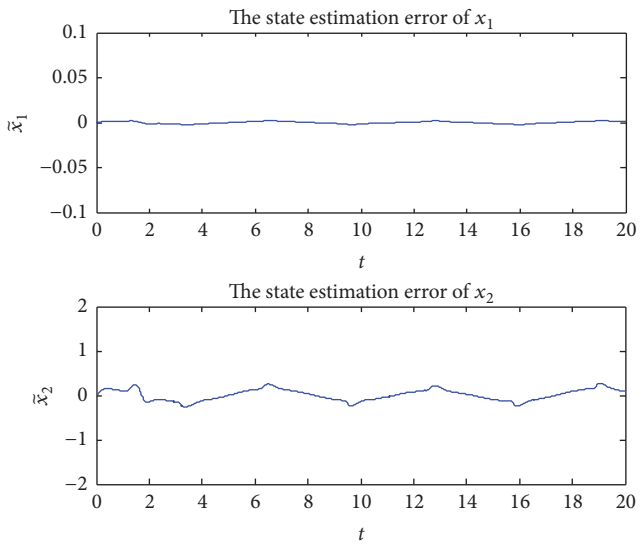


Figure 1: State estimation errors for $f_1(t)$.

The parameters of ESO are adopted as $a = 0.01, \delta = 0.01, \beta_1 = 100, \beta_2 = 65, \beta_3 = 80$, and step length $h = 0.01$. And the faulty signal is nominated as having two types:

$$\begin{aligned} f_1(t) &= \begin{cases} 0, & t < 0, \\ 3.5 \sin t, & t \geq 0, \end{cases} \\ f_2(t) &= \begin{cases} 0, & t < 0, \\ 3.5t, & t \geq 0. \end{cases} \end{aligned} \tag{37}$$

The simulation considers the effect of the two fault signals separately. When sine function fault $f_1(t)$ occurs, the state estimation errors are given in Figures 1 and 2 showing the combination of uncertainty, fault, and nonlinear function $\Phi(x_1, x_2, t) = \phi(x_1, x_2, t) + \omega(t) + f(t)$ tracking trajectory and it can be seen from this figure that a good tracking performance of the proposed ESO is achieved. Figure 3 shows the capability of the proposed method to track sine function fault $f_1(t)$ and diagnose the fault. When ramp function fault $f_2(t)$ occurs, the state estimation errors given by ESO are illustrated in Figure 4, and the tracking effects of ESO on the combination $\Phi(x_1, x_2, t)$ can be seen in Figure 5. Figure 6 shows the ramp function fault signal $f_2(t)$ and its estimation. From Figures 1 and 4 and Figures 2 and 5, it is seen that all the true values are approached by the estimated ones fleetly on the system states and the combination which is estimated by the ESO. The existence of band-limited white noise influences the accuracy of estimation, but the fault estimations can also track them timely which we can know from Figures 3 and 6.

On the other hand, it can be shown from Figures 1–6 that system states x_1, x_2 themselves have sharply periodic oscillation and their oscillatory period is 6 s. Although the oscillation has a great effect on fault diagnosis of the system and tracking performance, the tracking errors are converged and the fault estimations are accurate. That is, the variation of tracking error is caused by the periodic oscillation of states

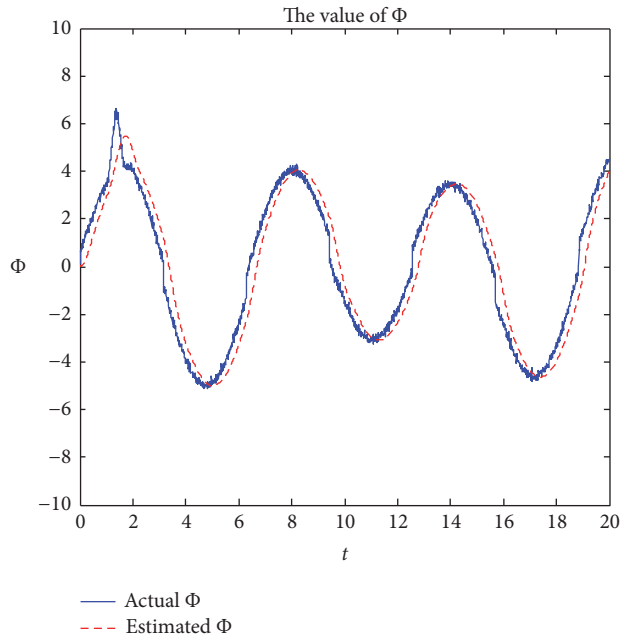


Figure 2: Combination $\Phi(x_1, x_2, t)$ estimate and true value for $f_1(t)$.

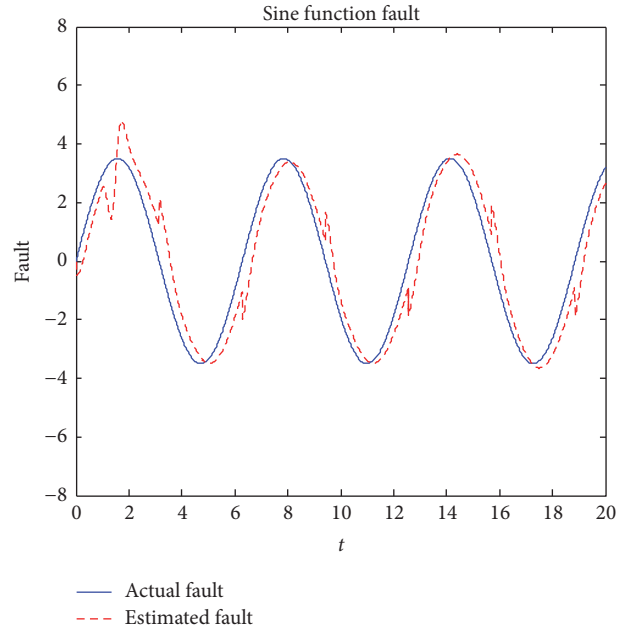


Figure 3: $f_1(t)$ and its estimation.

x_1, x_2. Therefore, the proposed method needs to overcome the oscillation to track states and faults. The simulation results illustrate that the tracking of states and faults can be achieved by the proposed method.

6. Conclusion

In this paper, based on the ESO of the active disturbance rejection controller and linearization of dynamic compensation, a robust fault diagnosis scheme for a class of nonlinear uncertain systems with triangular form is presented.

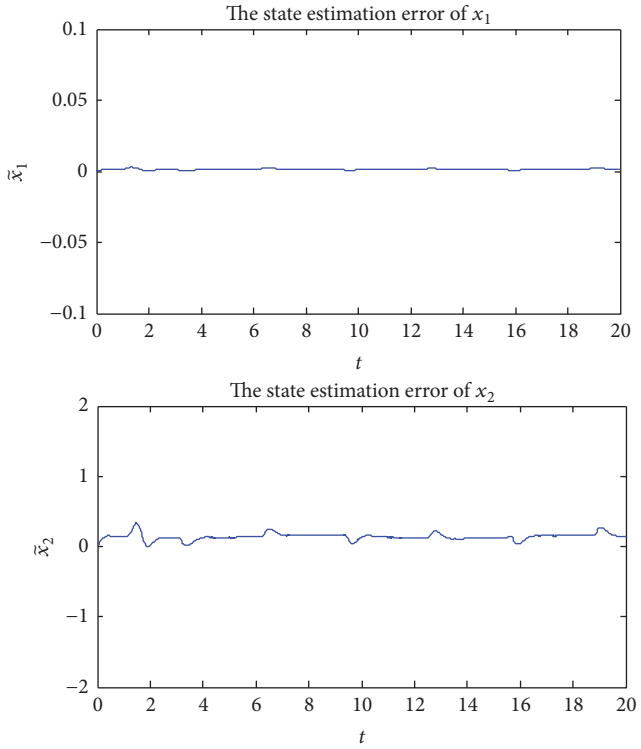


Figure 4: State estimation errors for $f_2(t)$.

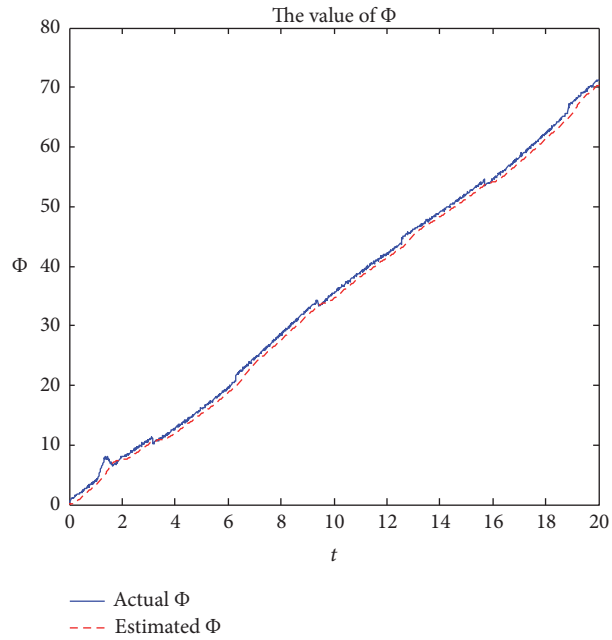


Figure 5: Combination $\Phi(x_1, x_2, t)$ estimate and true value for $f_2(t)$.

Compared with other methods, a good advantage is that the proposed method is more robust. The combination of uncertainty, fault, and nonlinear function is defined as a new system state. The ESO has a good tracking performance for system states and the combination. Then nonlinear system is linearized with state estimations by linearization of dynamic

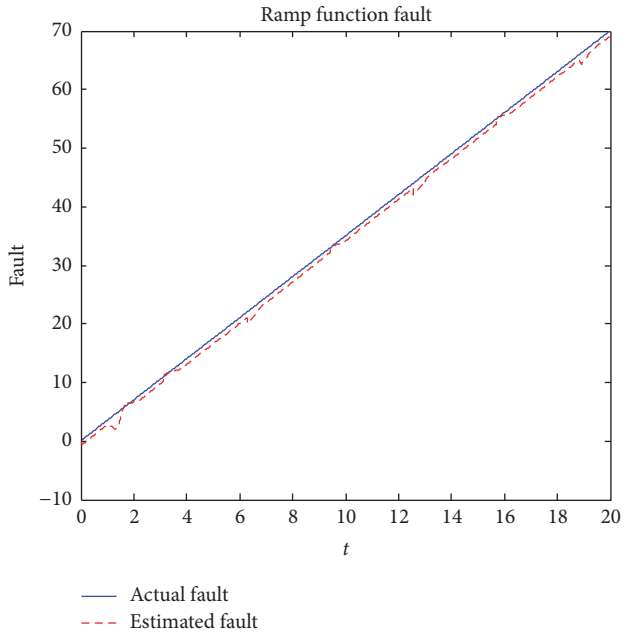


FIGURE 6: $f_2(t)$ and its estimation.

compensation to realize fault diagnosis. The robustness and fault detection are enhanced via the appropriately designed threshold in the diagnostic decision-making stage. A simulation example is used to illustrate the effectiveness of the proposed method.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by National Natural Science Foundation (NNSF) of China under Grant no. 61305117, 973 subproject, and National Key Laboratory Foundation of China under Grant 9140c59030411ht05.

References

[1] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[2] J. Chen and R. J. Patton, "Fault diagnosis of non-linear dynamic systems," in *Robust Model-Based Fault Diagnosis for Dynamic Systems*, vol. 3 of *The International Series on Asian Studies in Computer and Information Science*, pp. 251–295, Springer, Boston, MA, USA, 1999.

[3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis part I: quantitative model-based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.

[4] X. Zhang, M. M. Polycarpou, and T. Parisini, "Fault diagnosis of a class of nonlinear uncertain systems with Lipschitz nonlinearities using adaptive estimation," *Automatica. A Journal of IFAC, the International Federation of Automatic Control*, vol. 46, no. 2, pp. 290–299, 2010.

[5] J. S. Tsai, M.-H. Lin, C.-H. Zheng, S.-M. Guo, and L.-S. Shieh, "Actuator fault detection and performance recovery with Kalman filter-based adaptive observer," *International Journal of General Systems*, vol. 36, no. 4, pp. 375–398, 2007.

[6] J. Guo, X. Huang, and Y. Cui, "Design and analysis of robust fault detection filter using LMI tools," *Computers and Mathematics with Applications*, vol. 57, no. 11-12, pp. 1743–1747, 2009.

[7] H. R. Karimi, M. Zapateiro, and N. Luo, "A linear matrix inequality approach to robust fault detection filter design of linear systems with mixed time-varying delays and nonlinear perturbations," *Journal of the Franklin Institute. Engineering and Applied Mathematics*, vol. 347, no. 6, pp. 957–973, 2010.

[8] S. X. Ding, "Fault Isolation Schemes," in *Model-Based Fault Diagnosis Techniques*, Advances in Industrial Control, pp. 405–440, Springer, 2013.

[9] T. Mulumba, A. Afshari, K. Yan, W. Shen, and L. K. Norford, "Robust model-based fault diagnosis for air handling units," *Energy and Buildings*, vol. 86, pp. 698–707, 2015.

[10] L. Li and D. Zhou, "Fast and robust fault diagnosis for a class of nonlinear systems: Detectability analysis," *Computers and Chemical Engineering*, vol. 28, no. 12, pp. 2635–2646, 2004.

[11] F. Bagheri, H. Khaloozaded, and K. Abbaszadeh, "Stator fault detection in induction machines by parameter estimation, using adaptive Kalman filter," in *Proceedings of the 2007 Mediterranean Conference on Control and Automation, MED*, July 2007.

[12] P. Seiler, J. Bokor, B. Vanek, and G. J. Balas, "Robust model matching for geometric fault detection filters," in *Proceedings of the 2011 American Control Conference, ACC 2011*, pp. 226–231, San Francisco, CA, USA, July 2011.

[13] W. Wang and C. Wen, "Adaptive actuator failure compensation control of uncertain nonlinear systems with guaranteed transient performance," *Automatica. A Journal of IFAC, the International Federation of Automatic Control*, vol. 46, no. 12, pp. 2082–2091, 2010.

[14] W. Wang and C. Wen, "Adaptive compensation for infinite number of actuator failures or faults," *Automatica. A Journal of IFAC, the International Federation of Automatic Control*, vol. 47, no. 10, pp. 2197–2210, 2011.

[15] R. Sun, S. Liu, and Y.-F. Zhang, "Design of fault diagnosis observer for a class of nonlinear systems," *Kongzhi Lilun Yu Yingyong/Control Theory and Applications*, vol. 30, no. 11, pp. 1462–1466, 2013.

[16] R. Patton J, J. Chen, and B. Nielsen S, "Model-based methods for fault diagnosis: some guide-lines," *Transactions of the Institute of Measurement Control*, vol. 17, no. 2, pp. 73–83, 1995.

[17] X. Wu and B. Zhou, "Fault tolerance control for proton exchange membrane fuel cell systems," *Journal of Power Sources*, vol. 324, pp. 804–829, 2016.

[18] J. Huang, Z. Jiang, and T. Xu L, "Method on component fault diagnosis for nonlinear systems," *Control and Decision*, vol. 29, no. 5, pp. 925–928, 2014.

[19] X. Qing-shi and G. Ying-qing, "Fault diagnosis of aircraft engines based on adaptive sliding mode observers," *Aeronautical Computing Technique*, 2016.

[20] J. Guo W, L. Gou F, and P. Shi Y, "Research on Fault Detection Method Based on Adaptive Observer," *Computer Simulation*, vol. 31, no. 2, pp. 136-130, 2014.

[21] G. Stefatos and A. Ben Hamza, "Dynamic independent component analysis approach for fault detection and diagnosis," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8606–8617, 2010.

[22] D. Ye, P. Lin, and Z. Gao, "Research on fault diagnosis using extended state observer," *Chinese Journal of Scientific Instrument*, 2008.

[23] K. Chang and WU. Guo-Qing, "Fault diagnosis method for nonlinear systems via ESO," *Journal of Hebei University of Science*, 2012.

[24] K. Chang and G. Q. Wu, "Application of ESO in fault detection and identification for the nonlinear systems," *Information Technology*, 2013.

[25] Y. Yan B, *Research on several methods of fault diagnosis for nonlinear systems and their applications*, Shanghai Jiao Tong University, 2010.

[26] J. Chu, J. Yu, and J. C. Wang, "Design of nonlinear robust controller based on state transformation," *Acta Automatica Sinica*, vol. 18, no. 5, pp. 590–594, 1992.