Research Article

Optimal Cost-Effective Maintenance Policy for a Helicopter Gearbox Early Fault Detection under Varying Load

Xin Li, Jing Cai, Hongfu Zuo, and Huaiyuan Li

RMS Centre, Civil Aviation College, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Jing Cai; caijing@nuaa.edu.cn

Most of the existing fault detection methods rarely consider the cost-optimal maintenance policy. A novel multivariate Bayesian control approach is proposed, which enables the implementation of early fault detection for a helicopter gearbox with cost minimization maintenance policy under varying load. A continuous time hidden semi-Markov model (HSMM) is employed to describe the stochastic relationship between the unobservable states and observable observations of the gear system. Explicit expressions for the remaining useful life prediction are derived using HSMM. Considering the maintenance cost in fault detection, the multivariate Bayesian control scheme based on HSMM is developed; the objective is to minimize the long-run expected average cost per unit time. An effective computational algorithm in the semi-Markov decision process (SMDP) framework is designed to obtain the optimal control limit. A comparison with the multivariate Bayesian control chart based on hidden Markov model (HMM) and the traditional age-based replacement policy is given, which illustrates the effectiveness of the proposed approach.

1. Introduction

In a helicopter system, the mechanical drive system is the most efficient and compact device to transmit torque and change angular velocity. As the key mechanical transmission parts, gears are extensively used in aerospace, shipbuilding industry, and wind power industry, such as helicopters, cruisers, and wind turbines. Gear failure will cause machine halts in the whole mechanical system, resulting in great economic losses and even human casualties.

Condition monitoring and fault detection technique can significantly improve the reliability of the gear transmission system and reduce the occurrence of failure. In condition-based maintenance (CBM) and prognostics and health management (PHM), the vibration monitoring data obtained from the accelerometers carry a large amount of information, which were used in the fault detection of the gears for decades and were very helpful to prevent unnecessary machine halts [1, 2]. The accurate early fault detection can effectively prevent the occurrence of secondary damage. Appropriate preventive maintenance can significantly improve equipment availability and save cost for gearbox users. Therefore, the early gear

fault detection and maintenance scheme optimization have become research hotspots in recent years.

In recent research, advanced nonparametric methods were widely applied to gearbox early fault detection, such as the continuous wavelet transform (CWT) [3], the multiscale chirplet path pursuit (MSCPP) [4], and multiresolution Fourier transform [5]. Some researchers applied parametric time series model to gearbox fault detection, most of whom assumed that gearbox ran under constant load conditions. For example, Wang and Wong [6] indicated that the autoregressive (AR) model with a high diagnostic confidence level can detect the gear crack much earlier than the conventional method. Rofe [7] used the autoregressive moving-average (ARMA) model under the condition of load fluctuation for gear fault diagnosis and selected the variance and kurtosis of the residuals of ARMA model as the indicators of failure. It should be noted that the fault diagnosis methods in the literature mostly assume the gears run under constant speed or constant load conditions, meaning that all the signal responses are due to component degradation or failures. This is not always consistent with the reality that gears often run under varying load, such as the wind turbine gearboxes and

the helicopter planetary gearboxes [8, 9]. Under the assumption of constant load, it is difficult to separate the interference signals caused by load changes, which greatly affects the characteristics of the signals. In most cases, the change of the load variation in the feature level is similar to that caused by fault. This increases the difficulty of feature extraction and thereby also increases the difficulty of effective monitoring, diagnosis, and prediction. The gear motion residual (GMR) signal is an effective algorithm and highly insensitive to the varying load [10]. The healthy portion of the GMR signals is used to fit the VAR model and the residuals of the whole data sets are calculated to process the early fault detection scheme in this paper.

The operational states (healthy or unhealthy) cannot be observed directly when the gear system is running. Several approaches have been used to model the degradation process, such as proportional hazards model (PH) [11], stochastic filtering model (SFM) [12], and hidden Markov model (HMM) [13]. In recent years, HMM has been widely applied to speech recognition [14], system fault diagnosis [13, 15], and life prediction [16]. However, in HMM, the sojourn time in each operational state of the system is assumed to follow exponential distribution, that is, the "memoryless" property. As an extension of the HMM, the sojourn time in each operational state is not restricted by exponential distribution in a hidden semi-Markov model (HSMM), and its degradation process is much closer to the real degradation situations. Liu et al. [17, 18] proposed a diagnosis and prediction method of equipment based on the HSMM and the sojourn time follows normal distribution. Jiang et al. [19] carried out a small extension to HMM, considering a HSMM with the sojourn time obeys Erlang $(2, \lambda)$ distribution. A general Erlang HSMM is considered in this paper to model the deterioration process of the gear system. The expectation-maximization (EM) algorithm is employed to estimate the unknown state and observation parameters of the model. After the model parameters are estimated, the posterior probability that the system is in each state is used to derive the expressions for the reliability distribution function of remaining useful life (RUL), the probability density function (PDF), and the mean residual life (MRL) function.

When using the HSMM to model the deterioration of the gear system with the consideration of maintenance policy, it is necessary to consider the best stopping time. The Bayesian methods are used extensively to determine the stopping time. Girshick and Rubin [20] introduced the Bayesian method for the production process control for the first time in 1952, showing that repair was needed when the posterior probability was beyond the control limit. Considering the minimum maintenance cost, Kim et al. [13] put forward the optimal Bayesian forecasting model using the actual oil spectral analysis data. Most of the stochastic models assumed that the system degradation process was modeled by a continuous time HMM and each sojourn time of hidden states was exponentially distributed, which is not always realistic. Relaxing the assumption of Markovian process deterioration, Panagiotidou and Tagaras [21] proposed a model that integrated process control and maintenance for two operational states. However, this method is only applicable to univariate

monitoring. Such a drawback of the existing optimization model motivates us to set up an optimal maintenance model under real circumstances. In order to find the optimal stopping time, based on the HSMM, a Bayesian control scheme is established with the objective of minimizing the long-run expected average cost per unit time to determine the optimal stopping time and when to perform the preventive maintenance actions.

To the best of our knowledge, this is the first paper applying HSMM with general Erlang distribution of sojourn time in two hidden states and optimal Bayesian control scheme to find the cost-effective strategy for early fault detection of the helicopter gearbox. With the condition monitoring information, the proposed approach can not only update the remaining useful life at each sampling epoch, but also process early fault detection of the helicopter gear system and determine the optimal stopping time with minimum cost simultaneously.

The remainder of the paper is organized as follows. In Section 2, the healthy part of GMR signal of the helicopter gearbox life test vibration data is used to fit a VAR model, and the residuals of the whole data sets are obtained. In Section 3, the HSMM is applied, in which each operational state obeys the general Erlang distribution, to describe the degradation process of the gear system, and the EM algorithm is employed to estimate the unknown state and observation parameters of the model. The conditional reliability distribution functions of the RUL and the MRL function are derived. In Section 4, on the basis of hidden semi-Markov degradation modeling, an optimal Bayesian control chart is developed, and the optimal control limit is solved in the semi-Markov decision process framework, followed by a comparison with other methods. The conclusions are presented in Section 5.

2. Residuals Computation for a Helicopter Gearbox

2.1. Experimental Scheme. The overall process of the optimal Bayesian control scheme for the helicopter gearbox early fault detection is illustrated schematically in Figure 1. Firstly, we need to obtain the residuals of the multivariate condition monitoring data. The test data were obtained from the Mechanical Diagnostic Test Bed (MDTB) in Pennsylvania State University, as shown in Figure 2.

Gearbox used in this test includes a pinion gear with 21 teeth and a drive gear with 70 teeth. The gearbox was driven by induction motors (22.38 kW, 1750 rpm) and the torques were provided by alternating current absorption motors (55.95 kW, 1750 rpm). The power of the gearbox is 3.73∼ 14.92 kW and the speed ratio range is 1.2 : 1∼6. #2 and #3 accelerometers were installed on the gearbox body, respectively, in the axial and horizontal directions, as shown in Figure 3.

Samples were taken every 8 minutes (0.1333 h) and stored in a new file with a sampling frequency of 20 kHz. Each sampling time is 10 s. The 16-bit resolution AD converter was used to ensure the accelerometers have sufficient precision. Driving motor speed V01 and gearbox output torque V05 were
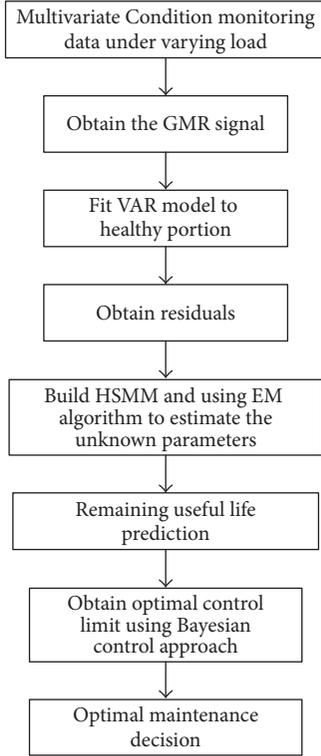
FIGURE 1: The flow chart of Bayesian control scheme.
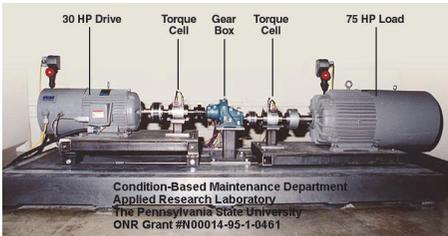
FIGURE 2: Mechanical diagnostic test bed.

also collected with sampling frequency of 1 kHz and sampling time of 10 s. The test is divided into two stages: at a certain rotating speed, gearbox firstly was run at 100% rated output torque for 96 hours (555 in-lbs); it was then run under varying load until failure. In the second stage, the output torque was periodically increased from 50% to 300% rated torque (50%, 100%, 150%, 200%, 250%, and 300%) and then dropped from 300% to 50% rated torque (300%, 250%, 200%, 150%, 100%, and 50%) gradually. The output torque of the gearbox is shown in Figure 4. Figure 5 shows the mean value of the input shaft speed. The input shaft speed fluctuation error in Figure 5 is within 0.06%; therefore the speed can be considered as a constant.

The gearbox was run normally in the first stage for 96 hours; thus only the vibration data collected in the second stage (varying load) were analyzed. Sampling points collected by #2 and #3 sensors were, respectively, saved in files A02 and A03, and each of them contained 145 data files under varying load. The original vibration data is shown in Figure 6. The results of shutdown fault inspection were shown in Figure 7 that the pinion gear was normal while five fully broken and two partially broken teeth were found in the drive gear.

2.2. Gear Motion Residual (GMR) Signal. TSA algorithm, which extracts the meshing frequency components from the gear vibration data and then synchronously adds and averages them with the rotation of the gear shaft, is widely used in fault diagnosis of gears. The number of sampling points in a single rotational period can be calculated using the following formula:

$$K = \left\lceil \frac{f_s}{\left(f_m / N_g \right)} \right\rceil, \tag{1}$$

where f_s is sampling frequency, f_m is the fundamental meshing frequency of the gear, N_g is the number of teeth of the gear, and $\lceil \cdot \rceil$ is the ceil function.

Suppose that $V(k)$, $k = 1, 2, \ldots, n$ data points were contained in each sampling data file, the rotation period number M of the gear is given by

$$M = \left\lfloor \frac{n}{K} \right\rfloor, \tag{2}$$

where $\lfloor \cdot \rfloor$ represents the floor function.

In a complete revolution of the gear of interest, the TSA signal is given by

$$V_n(k) = \frac{1}{M} \sum_{i=0}^{M-1} V(k + iK), \quad k = 1, 2, \ldots, K. \tag{3}$$

The TSA signal of the helicopter gearbox is shown in Figure 8.

It is efficacious to use the TSA method to reduce the influence of the background noise and the nongear vibration source. GMR signal, which can be obtained by removing the gear meshing frequency and its harmonic from the TSA signals, is highly insensitive to the varying load conditions [10]. The GMR signal can be expressed as follows:

$$R_n = V_n - E_n, \tag{4}$$

where E_n is the signal composed of the eliminated components.

The GMR signal of the gear of interest is shown in Figure 9. It can be seen from the comparison between Figure 9 and the TSA signal that the GMR signal effectively removes the noise, representing more obvious signal change features. Further, we consider a VAR model for the healthy part of the GMR signal to obtain the residual.

2.3. GMR Residuals Computation Using VAR Model. In this section, the healthy portion of the GMR signal is used to fit a VAR model. Two-dimensional data can be expressed as $\{Z_1^i, Z_2^i, \ldots, Z_{t_i}^i\}$, $i = 1, 2$. Assume that the healthy portion of the data obeys a stationary VAR process:

$$Z_n = \mu + \sum_{r=1}^{p} \Phi_r \left(Z_{n-r} - \mu \right) + \varepsilon_n, \quad n \in Z, \tag{5}$$

(a) Side view

(b) Front view
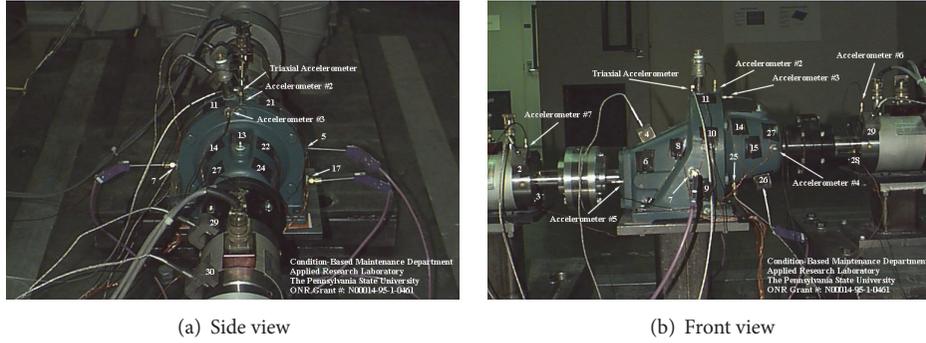
FIGURE 3: Locations of #2 and #3 accelerometers.
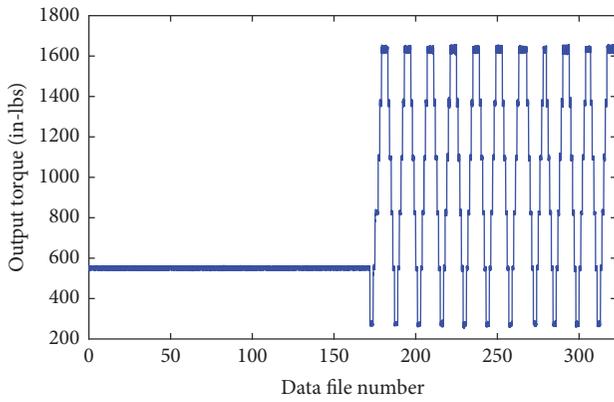
FIGURE 4: Output torque of the gearbox.
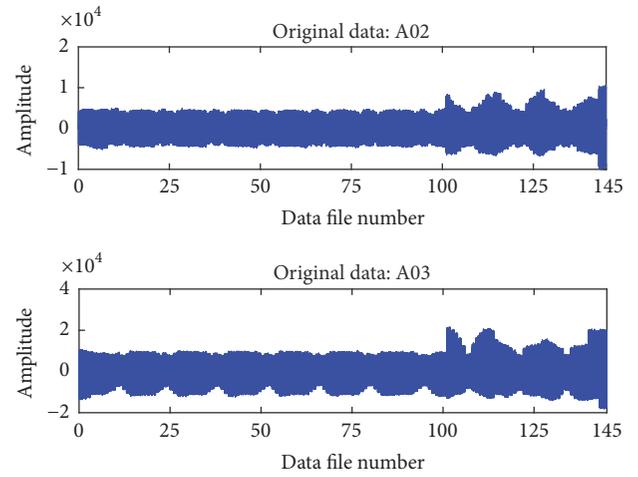
FIGURE 6: Original data of gearbox under varying load.
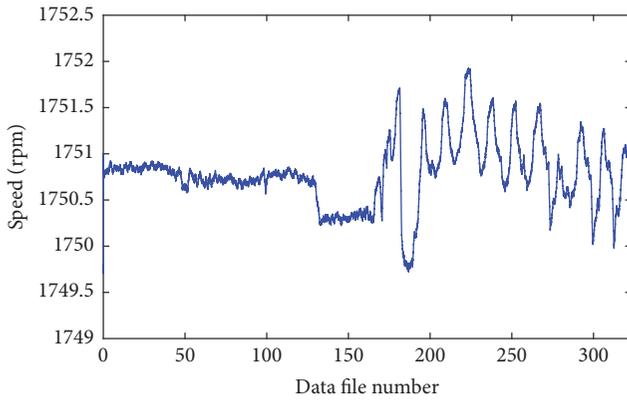
FIGURE 5: Input speed of the driving motor.

where $\mu \in \mathbb{R}^2$ is the mean value; $p \in \mathbb{N}$ is the model order; $\Phi_r \in \mathbb{R}^{2 \times 2}$ is the autocorrelation matrices; ε_n are independent identically distributed (*i.i.d*) and obey $N_2(0, \Sigma)$; $\Sigma \in \mathbb{R}^{2 \times 2}$ is the covariance. Let $\delta = \mu - \sum_{r=1}^{p} \Phi_r \mu$, so (5) can be expressed with the following form:

$$W = \Phi A + B, \qquad (6)$$

where

$$W = \left(Z_{t_2}^2, Z_{t_2-1}^2, \ldots, Z_{p+2}^2, Z_{p+1}^2, Z_{t_1}^1, Z_{t_1-1}^1, \ldots, Z_{p+2}^1, \right.$$
$$\left. Z_{p+1}^1 \right)',$$

$$A = \left(\delta, \Phi_1, \Phi_2, \ldots, \Phi_p \right)',$$

$$B = \left(\varepsilon_{t_2}^2, \varepsilon_{t_2-1}^2, \ldots, \varepsilon_{p+2}^2, \varepsilon_{p+1}^2, \varepsilon_{t_1}^1, \varepsilon_{t_1-1}^1, \ldots, \varepsilon_{p+2}^1, \varepsilon_{p+1}^1 \right)', \qquad (7)$$

$$\Phi = \begin{pmatrix} 1 & 1 & 1 & 1 \\ Z_{t_2-1}^2 & Z_p^2 & Z_{t_1-1}^1 & Z_p^1 \\ Z_{t_2-2}^2 & Z_{p-1}^2 & Z_{t_1-2}^1 & Z_{p-1}^1 \\ \vdots & \vdots & \vdots & \vdots \\ Z_{t_2-p}^2 & Z_1^2 & Z_{t_1-p}^1 & Z_1^1 \end{pmatrix}'.$$

Reinsel [22] indicated that A and Σ can be obtained using least squares estimates, and the AIC criterion is selected in this paper to determine the model order p. Thus the estimated
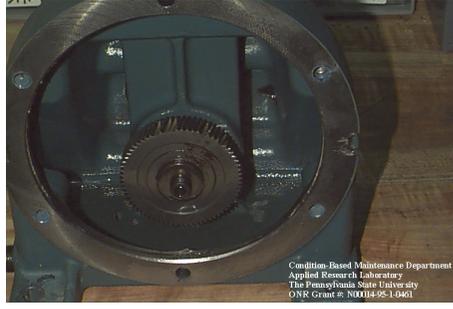
(a) Broken drive gear

(b) Five fully broken and two partially broken teeth

FIGURE 7: Test results of the drive gear.
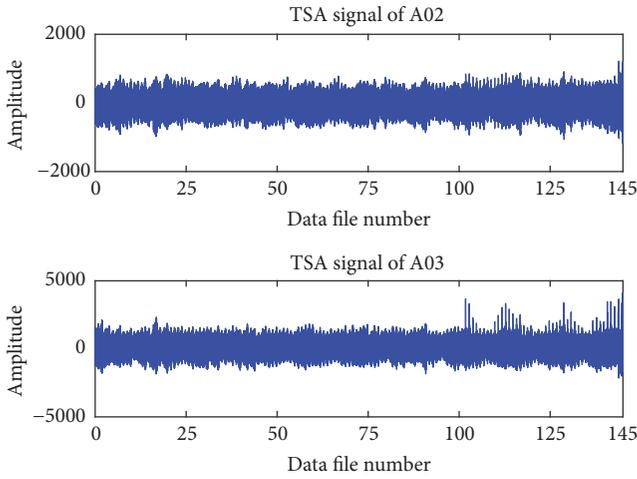
FIGURE 8: TSA signal of the original data.
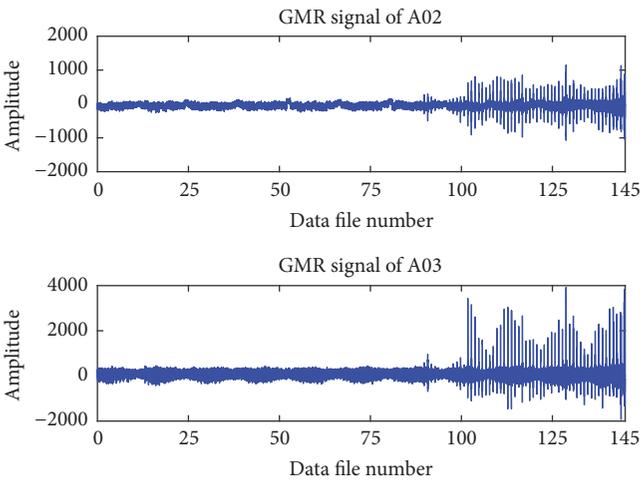
FIGURE 9: GMR signal of the original data.

VAR model parameters $\widehat{\gamma} = (\widehat{p}, \widehat{\delta}, \widehat{\Phi}_1, \widehat{\Phi}_2, \ldots, \widehat{\Phi}_p)$ can be utilized to calculate the residual Y_n of the TSA signals:

$$Y_n = Z_n - E_{\widehat{\gamma}}\left(Z_n \mid \overrightarrow{Z}_{n-1}\right), \qquad (8)$$

where $\overrightarrow{Z}_{n-1} = (Z_1, Z_2, \ldots, Z_{n-1})$.
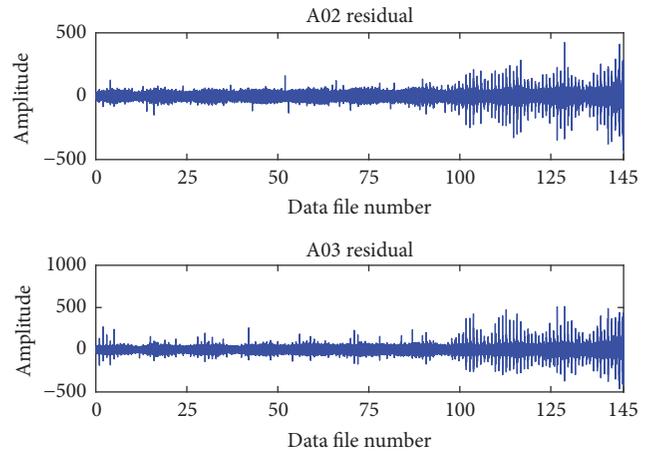
FIGURE 10: Residuals for data files 1–145.

For $n > \widehat{p}$,

$$Y_n = Z_n - \left(\widehat{\delta} + \sum_{r=1}^{\widehat{p}} \widehat{\Phi}_r Z_{n-r}\right). \qquad (9)$$

For $n \leq \widehat{p}$, the residual Y_n can be computed recursively by Kalman filter, for which the details can be found in the Appendix.

Using the above method, we obtain the GMR residual signal of the gear, which is shown in Figure 10. To compress the huge amount of data, we select the standard deviation, which is commonly used characteristic parameter in the vibration signal processing, to process the residual of the gear, as shown in Figure 11.

Figure 11 illustrates that the residual standard deviation values of A02 and A03 are relatively stable from files 1 to 100; therefore, we assume the system is in the healthy state in this portion, while from files 101 to 145 the system operates in the unhealthy state as the standard deviation values of this part show an obvious increase compared with those of the healthy portion. Normality and independence tests are taken for these two parts of data, the results of which are shown in Table 1.

From Table 1, both the standard deviation values for A02 and A03 passed the normality and independence tests,
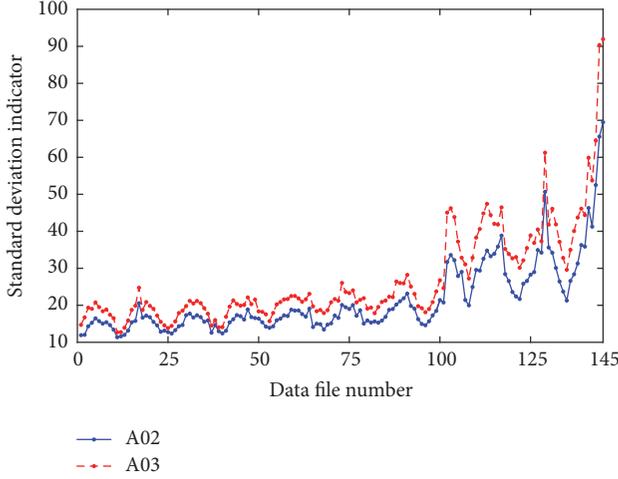
FIGURE 11: Standard deviation indicator of residuals.

TABLE 1: p value of the normality and independence tests.

RMS of residuals	Healthy portion (files 1–100)	Unhealthy portion (files 101–145)
Normality	0.1641	0.4032
Independence	0.3321	0.1861

indicating that standard deviation values of the residual are independent and obey the multivariate normal distribution, which is consistent with the assumption that the observation of the HSMM described in Section 3 obeys multivariate normal distribution.

3. Residual Life Prediction Based on Hidden Semi-Markov Model

3.1. Hidden Semi-Markov Model. It has been demonstrated from in [13, 19, 23] that a 3-state Markov model is adequate to model the deterioration process of the system. We assume that the degradation process of the system is described by a 3-state HSMM, the state space of which is $S = \{0, 1, 2\}$. States 0 and 1 are unobservable and represent the good and warning system states, respectively. Only state 2 is assumed to be observable and represents the failure state. Suppose the system starts in the healthy state, that is, $P(X_0 = 0) = 1$. While the sampling interval is fixed, the observation of the system $Y = (Y_k : k \in \mathbb{N})$ is independent given the state of the gear system. $y_\Delta, y_{2\Delta}, \ldots, y_{k\Delta} \in \mathbb{R}^d$ is used to denote the d-dimension observation vector at each sampling epoch; thus $y_{k\Delta}$ given the state x obeys d-dimension multivariate normal distribution with $N_d(\mu_x, \Sigma_x)$, $x = 0, 1$, and the probability density function is given by

$$
f_{y_{k\Delta}|X_{n\Delta}}(y_{k\Delta} \mid x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_x)}}
$$
$$
\cdot \exp\left(-\frac{1}{2}(y_{k\Delta} - \mu_x)' \Sigma_x^{-1}(y_{k\Delta} - \mu_x)\right),
$$
(10)
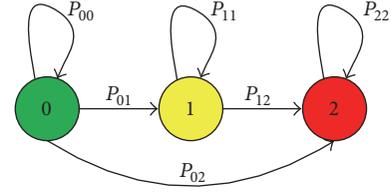
FIGURE 12: Schematic for system state transition of the 3-state HSMM.

where $\mu_0, \mu_1 \in \mathbb{R}^2$, $\Sigma_0, \Sigma_1 \in \mathbb{R}^{2\times2}$ are unknown observation parameters.

The state transition of the gear system is illustrated in Figure 12. The transition probability matrix is $P = [P_{ij}]$ ($i, j \in S$), where P_{ij} represents the probability that the system transits to state j after it leaves from state i.

Generally, the transition probability matrix of the system is assumed as follows:

$$
P = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0 & p_{01} & p_{02} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}
$$
(11)

Eq. (11) indicates that the system starts in the healthy state, where $p_{01} + p_{02} = 1$.

Suppose that in state i for $i = 0, 1$, the sojourn time obeys the general Erlang distribution, and the probability density function is given by

$$
f_i(t) = \frac{\lambda^{k_i} t^{k_i - 1} e^{-\lambda t}}{(k_i - 1)!} \quad \text{for } t \geq 0,
$$
(12)

where $k_i \in \mathbb{N}^+$ is the unknown shape parameter and $\lambda > 0$ is unknown rate parameter. Erlang distribution can also be modeled as a series of a given number of exponential phases running one after another until the end of the sojourn time. In order to record the phases in the Markov model, we need to enlarge the state space. Let $\Theta = \{K_1, K_2, K_3\}$ be the new state space, where $K_1 = \{1, \ldots, k_1\}$ represents the set of healthy states, $K_2 = \{k_1 + 1, \ldots, k_1 + k_2\}$ represents the set of unhealthy states, and $K_3 = \{k_1 + k_2 + 1\}$ represents the failure state.

The system's failure time is denoted by $\xi = \inf\{t \geq 0 : X_t = 2\}$. Let O denote the data histories and $\Lambda = (p_{01}, p_{02}, k_1, k_2, \lambda)$ and $\Psi = (\mu_0, \mu_1, \Sigma_0, \Sigma_1)$ denote the sets of state and observation parameters to be estimated, respectively. Since the sample path $(X_t : t \geq 0)$ of the state process in HSMM is unobservable, it is difficult to determine the analytical expression of the maximum likelihood function. The expectation-maximization (EM) algorithm can solve it by iteratively maximizing the pseudo likelihood function. Let $\Lambda_0 = (\hat{p}_{01}, \hat{p}_{02}, \hat{k}_1, \hat{k}_2, \hat{\lambda})$ and $\Psi_0 = (\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma}_0, \hat{\Sigma}_1)$ be the initial values; the EM algorithm is shown in Figure 13.

Updated parameters Λ^*, Ψ^* in each step are then used as new initial values to the E-step, leading the iteration in E-step and M-step until the Euclidean norm satisfies $|(\Lambda^*, \Psi^*) - (\hat{\Lambda}, \hat{\Psi})| < \varepsilon$, for small given $\varepsilon > 0$. Detailed expressions of likelihood function and pseudo likelihood function can
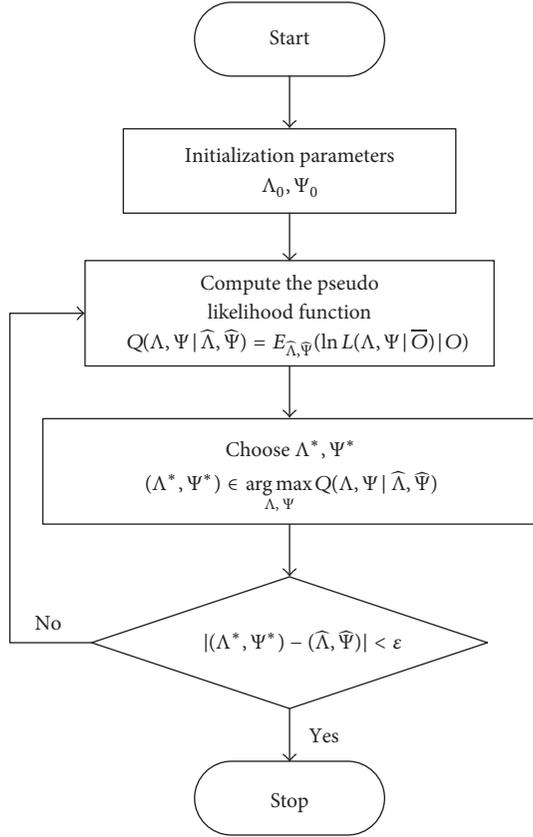
FIGURE 13: Flow chart of the EM algorithm.

be found in Khaleghei and Makis [23]. The state parameters p_{01}, p_{02} can be updated by calculating the unique stationary point from

$$\frac{\partial Q\left(\Lambda, \Psi \mid \widehat{\Lambda}, \widehat{\Psi}\right)}{\partial p_{01}} = \frac{\partial Q\left(\Lambda, \Psi \mid \widehat{\Lambda}, \widehat{\Psi}\right)}{\partial p_{02}} = \mathbf{0}. \quad (13)$$

The explicit expressions of updated p_{01}^* in each step are given by

$$p_{01}^* = \frac{\sum_{i=1}^N \widehat{\alpha}_1{}^i + \sum_{j=1}^M \widehat{\beta}_1{}^j}{\sum_{i=1}^N \left(\widehat{\alpha}_1 + \widehat{\omega}\right)^i + \sum_{j=1}^M \widehat{\beta}_1{}^j}. \quad (14)$$

The updated k_1^*, k_2^*, λ^* in each step can be obtained by

$$\max_{k_1, k_2, \lambda} \left(\sum_{i=1}^N \left(\widehat{\alpha}_1 + \widehat{\omega}\right)^i + \sum_{j=1}^M \left(\widehat{\beta}_1 + \widehat{\gamma}_1\right)^j\right) k_1 \ln \lambda$$

$$+ \sum_{i=1}^N \widehat{\alpha}_1{}^i k_2 \ln \lambda + \sum_{i=1}^N \widehat{\alpha}_3{}^i \left(k_2 - 1\right) + \sum_{j=1}^M \frac{\langle a_4, \widehat{g}\rangle^j}{\widehat{f}^j}$$

$$+ \left(\sum_{i=1}^N \left(\widehat{\alpha}_2 + \widehat{\omega} \ln t\right)^i + \sum_{j=1}^M \left(\widehat{\beta}_3 + \widehat{\gamma}_3\right)^j\right) \left(k_1 - 1\right)$$

$$- \left(\sum_{i=1}^N \widehat{\alpha}_1{}^i + \sum_{j=1}^M \widehat{\beta}_1{}^j\right) \ln \left(k_2 - 1\right)!$$

$$- \left(\sum_{i=1}^N \left(\widehat{\alpha}_1 + \widehat{\omega}\right)^i + \sum_{j=1}^M \left(\widehat{\beta}_1 + \widehat{\gamma}_1\right)^j\right) \ln \left(k_1 - 1\right)!$$

$$- \left(\sum_{i=1}^N t \left(\widehat{\alpha}_1 + \widehat{\omega}\right)^i + \sum_{j=1}^M \left(\widehat{\beta}_2 + \widehat{\gamma}_2\right)^j\right) \lambda. \quad (15)$$

For observation parameters, the unique stationary point can be obtained by

$$\frac{\partial Q\left(\Lambda, \Psi \mid \widehat{\Lambda}, \widehat{\Psi}\right)}{\partial \mu_0} = \mathbf{0},$$

$$\frac{\partial Q\left(\Lambda, \Psi \mid \widehat{\Lambda}, \widehat{\Psi}\right)}{\partial \mu_1} = \mathbf{0},$$

$$\frac{\partial Q\left(\Lambda, \Psi \mid \widehat{\Lambda}, \widehat{\Psi}\right)}{\partial \Sigma_0^{-1}} = \mathbf{0}, \quad (16)$$

$$\frac{\partial Q\left(\Lambda, \Psi \mid \widehat{\Lambda}, \widehat{\Psi}\right)}{\partial \Sigma_1^{-1}} = \mathbf{0}.$$

Thus the explicit expressions of the updated $\Psi^* = (\mu_0^*, \mu_1^*, \Sigma_0^*, \Sigma_1^*)$ in each step are given by

$$\mu_0^* = \frac{\sum_{i=1}^N \langle n_1, \widehat{C}\rangle + \sum_{j=1}^M \langle n_1, \widehat{D}\rangle}{\sum_{i=1}^N \langle s_1, \widehat{C}\rangle + \sum_{j=1}^M \langle s_1, \widehat{D}\rangle},$$

$$\Sigma_0^* = \frac{\sum_{i=1}^N \langle n_3, \widehat{C}\rangle + \sum_{j=1}^M \langle n_3, \widehat{D}\rangle}{\sum_{i=1}^N \langle s_1, \widehat{C}\rangle + \sum_{j=1}^M \langle s_1, \widehat{D}\rangle},$$

$$\mu_1^* = \frac{\sum_{i=1}^N \langle n_2, \widehat{C}\rangle + \sum_{j=1}^M \langle n_2, \widehat{D}\rangle}{\sum_{i=1}^N \langle s_2, \widehat{C}\rangle + \sum_{j=1}^M \langle s_2, \widehat{D}\rangle}, \quad (17)$$

$$\Sigma_1^* = \frac{\sum_{i=1}^N \langle n_4, \widehat{C}\rangle + \sum_{j=1}^M \langle n_4, \widehat{D}\rangle}{\sum_{i=1}^N \langle s_2, \widehat{C}\rangle + \sum_{j=1}^M \langle s_2, \widehat{D}\rangle},$$

where the form $\langle a, b\rangle = a'b$, $n_1 = (\mathbf{0}, y_1, \sum_{n=1}^2 y_n, \ldots, \sum_{n=1}^T y_n)'$, $s_1 = (0, 1, \ldots, T)$, $n_2 = (\sum_{n=1}^T y_n, \sum_{n=2}^T y_n, \ldots, y_T, \mathbf{0})'$, $s_2 = (T, \ldots, 1, 0)'$, $n_3 = (\mathbf{0}, (y_1 - \widehat{\mu}_1)'(y_1 - \widehat{\mu}_1), \ldots, \sum_{n=1}^T (y_1 - \widehat{\mu}_1)'(y_1 - \widehat{\mu}_1))'$, and $n_4 = (\sum_{n=1}^T (y_n - \widehat{\mu}_2)'(y_n - \widehat{\mu}_2), \sum_{n=2}^T (y_n - \widehat{\mu}_2)'(y_n - \widehat{\mu}_2), \ldots, (y_T - \widehat{\mu}_2)'(y_T - \widehat{\mu}_2), \mathbf{0})'$.

The initial values to $\widehat{\Lambda}$ and $\widehat{\Psi}$ were assigned and EM algorithm was employed to solve the parameters to be estimated, the convergence criterion for which is $|(\Lambda^*, \Psi^*) - (\widehat{\Lambda}, \widehat{\Psi})| < 1e - 6$. We assigned different initial values and the results were very similar, which are shown in Tables 2 and 3.

3.2. Remaining Useful Life (RUL) Prediction.

Let $\Pi_k(i)$ be the posterior probability that given the observation data $y_\Delta, y_{2\Delta}, \ldots, y_{k\Delta} \in \mathbb{R}^2$ the system is in state i ($1 \le i \le k_1 + k_2$) at sampling time $k\Delta$:

$$\Pi_k(i) = P\left(\Theta_{k\Delta} = i \mid \xi > k\Delta, y_\Delta, y_{2\Delta}, \ldots, y_{k\Delta}, \vec{\Pi}_{k-1}\right), \quad (18)$$

TABLE 2: Model parameters estimation with EM algorithm.

Parameters	Initial values	First iteration	Second iteration	Last iteration
$\widehat{p}_{01}$	0.8	0.6897	0.9187	1.0000
$\widehat{p}_{02}$	0.2	0.3103	0.0813	$7.6502e - 05$
$\widehat{k}_1$	3	2	2	2
$\widehat{k}_2$	3	1	2	2
$\widehat{\lambda}$	0.5	0.4063	0.2918	0.2069
$\widehat{\mu}_0$	$\begin{pmatrix} 10 \\ 15 \end{pmatrix}$	$\begin{pmatrix} 13.6567 \\ 19.3175 \end{pmatrix}$	$\begin{pmatrix} 13.9976 \\ 19.3198 \end{pmatrix}$	$\begin{pmatrix} 15.9207 \\ 19.4560 \end{pmatrix}$
$\widehat{\mu}_1$	$\begin{pmatrix} 20 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 25.0970 \\ 33.5655 \end{pmatrix}$	$\begin{pmatrix} 27.0138 \\ 36.3763 \end{pmatrix}$	$\begin{pmatrix} 29.9528 \\ 38.8550 \end{pmatrix}$
$\widehat{\Sigma}_0$	$\begin{pmatrix} 5 & 8 \\ 8 & 10 \end{pmatrix}$	$\begin{pmatrix} 5.2117 & 5.9879 \\ 5.9879 & 9.7890 \end{pmatrix}$	$\begin{pmatrix} 5.2980 & 6.3219 \\ 6.3219 & 9.7924 \end{pmatrix}$	$\begin{pmatrix} 5.3835 & 6.6110 \\ 6.6110 & 9.2166 \end{pmatrix}$
$\widehat{\Sigma}_1$	$\begin{pmatrix} 50 & 100 \\ 100 & 100 \end{pmatrix}$	$\begin{pmatrix} 89.6879 & 138.6759 \\ 138.6759 & 200.8768 \end{pmatrix}$	$\begin{pmatrix} 102.8798 & 139.0890 \\ 139.0890 & 203.3276 \end{pmatrix}$	$\begin{pmatrix} 123.1435 & 159.1407 \\ 159.1407 & 212.8811 \end{pmatrix}$
$Q \times 10^3$	—	-2.48	-1.89	-1.45

TABLE 3: Model parameters estimation with EM algorithm using different initial values.

Parameters	Initial values	First iteration	Second iteration	Last iteration
$\widehat{p}_{01}$	0.5	0.3176	0.8920	1.0000
$\widehat{p}_{02}$	0.5	0.6824	0.1080	$2.2018e - 08$
$\widehat{k}_1$	5	3	3	2
$\widehat{k}_2$	5	2	2	2
$\widehat{\lambda}$	0.3	0.2663	0.2199	0.2070
$\widehat{\mu}_0$	$\begin{pmatrix} 15 \\ 15 \end{pmatrix}$	$\begin{pmatrix} 13.8769 \\ 16.7687 \end{pmatrix}$	$\begin{pmatrix} 14.8768 \\ 18.9879 \end{pmatrix}$	$\begin{pmatrix} 15.9214 \\ 19.4561 \end{pmatrix}$
$\widehat{\mu}_1$	$\begin{pmatrix} 25 \\ 35 \end{pmatrix}$	$\begin{pmatrix} 28.1879 \\ 36.9768 \end{pmatrix}$	$\begin{pmatrix} 29.0180 \\ 37.9671 \end{pmatrix}$	$\begin{pmatrix} 29.9731 \\ 38.8583 \end{pmatrix}$
$\widehat{\Sigma}_0$	$\begin{pmatrix} 5 & 10 \\ 10 & 5 \end{pmatrix}$	$\begin{pmatrix} 3.9112 & 7.0879 \\ 7.0879 & 6.0788 \end{pmatrix}$	$\begin{pmatrix} 5.1939 & 6.9078 \\ 6.9078 & 8.6892 \end{pmatrix}$	$\begin{pmatrix} 5.4165 & 6.6329 \\ 6.6329 & 9.1875 \end{pmatrix}$
$\widehat{\Sigma}_1$	$\begin{pmatrix} 100 & 150 \\ 150 & 200 \end{pmatrix}$	$\begin{pmatrix} 113.6877 & 152.3243 \\ 152.3243 & 209.7609 \end{pmatrix}$	$\begin{pmatrix} 119.0451 & 157.9880 \\ 157.9880 & 210.6791 \end{pmatrix}$	$\begin{pmatrix} 122.8959 & 159.2142 \\ 159.2142 & 212.7090 \end{pmatrix}$
$Q \times 10^3$	—	-1.98	-1.53	-1.32

where $\sum_{i=1}^{k_1+k_2} \Pi_k(i) = 1$, $\vec{\Pi}_k = (\Pi_k(1), \Pi_k(2), \dots, \Pi_k(k_1 + k_2))$ denotes the posterior probability vector. Suppose the system always starts in the first phase of the healthy state, that is, $\Pi_0(1) = 1$. According to the Bayes' theorem, the posterior probability $\Pi_k(i)$ can be updated iteratively at each sampling epoch by the following formula:

$$\Pi_k(i) = \frac{g\left(y_{k\Delta} \mid \Theta_{k\Delta} = i, \xi > k\Delta, y_{(k-1)\Delta}, \vec{\Pi}_{k-1}\right) \times P\left(\Theta_{k\Delta} = i \mid \xi > k\Delta, y_{(k-1)\Delta}, \vec{\Pi}_{k-1}\right)}{\sum_{j=1}^{k_1+k_2} g\left(y_{k\Delta} \mid \Theta_{k\Delta} = j, \xi > k\Delta, y_{(k-1)\Delta}, \vec{\Pi}_{k-1}\right) \times P\left(\Theta_{k\Delta} = j \mid \xi > k\Delta, y_{(k-1)\Delta}, \vec{\Pi}_{k-1}\right)}, \tag{19}$$

where $P(\Theta_{k\Delta} = i, \xi > k\Delta, y_{(k-1)\Delta}, \vec{\Pi}_{k-1}) = \sum_{1 \le m \le i} P_{mi}(\Delta) \Pi_{k-1}(m) / \sum_{m=1}^{k_1+k_2} \sum_{j=m}^{k_1+k_2} P_{mj}(\Delta) \Pi_{k-1}(m)$.

It can be obtained from (10) that $y_{k\Delta} \mid \Theta_{k\Delta} \sim N_d(\mu_x, \Sigma_x)$ for $x = 1, 2$; thus

$$g\left(y_{k\Delta} \mid \Theta_{k\Delta} = i, \xi > k\Delta, y_{(k-1)\Delta}, \vec{\Pi}_{k-1}\right)$$
$$= \begin{cases} f\left(y_{k\Delta} \mid \mu_1, \Sigma_1\right) & \text{if } i \in K_1 \\ f\left(y_{k\Delta} \mid \mu_2, \Sigma_2\right) & \text{if } i \in K_2, \end{cases} \tag{20}$$

where $f(y_{k\Delta} \mid \mu_x, \Sigma_x) = (1/\sqrt{(2\pi)^d \det(\Sigma_x)}) \exp(-(1/2)(y_{k\Delta} - \mu_x)' \Sigma_x^{-1}(y_{k\Delta} - \mu_x))$, $x = 1, 2$.

For the gearbox degeneration system, the estimated shape parameters of Erlang distribution in the hidden states are $k_1 = k_2 = 2$ (see Table 2); therefore the transition probability matrix $P_{ij}(\Delta)$ can be obtained by solving Kolmogorov backward differential equations:

$$P_{ij}(t) = \begin{pmatrix} e^{-\lambda t} & \lambda t e^{-\lambda t} & p_{01}\dfrac{(\lambda t)^2}{2!}e^{-\lambda t} & p_{01}\dfrac{(\lambda t)^3}{3!}e^{-\lambda t} & 1 - e^{-\lambda t}\left(1 + \lambda t + p_{01}\dfrac{(\lambda t)^2}{2!} + p_{01}\dfrac{(\lambda t)^3}{3!}\right) \\ 0 & e^{-\lambda t} & p_{01}\lambda t e^{-\lambda t} & p_{01}\dfrac{(\lambda t)^2}{2!}e^{-\lambda t} & 1 - e^{-\lambda t}\left(1 + p_{01}\lambda t + p_{01}\dfrac{(\lambda t)^2}{2!}\right) \\ 0 & 0 & e^{-\lambda t} & \lambda t e^{-\lambda t} & 1 - e^{-\lambda t}(1 + \lambda t) \\ 0 & 0 & 0 & e^{-\lambda t} & 1 - e^{-\lambda t} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{21}$$

Suppose at sampling time $k\Delta$, the gear system has not failed, that is, $\xi > k\Delta$; then for $t \ge 0$ and $k_1 = k_2 = 2$, the conditional reliability function of RUL of the gear system is

$$R\left(t \mid \vec{\Pi}_k\right) = P\left(\xi - k\Delta > t \mid \xi > k\Delta, \vec{\Pi}_k\right)$$
$$= P\left(\Theta_{k\Delta+t} \notin K_3 \mid \xi > k\Delta, \vec{\Pi}_k\right)$$
$$= \Pi_k(1) e^{-\lambda t}\left(1 + \lambda t + p_{01}\frac{(\lambda t)^2}{2!} + p_{01}\frac{(\lambda t)^3}{3!}\right) \tag{22}$$
$$+ \Pi_k(2) e^{-\lambda t}\left(1 + p_{01}\lambda t + p_{01}\frac{(\lambda t)^2}{2!}\right)$$
$$+ \Pi_k(3) e^{-\lambda t}(1 + \lambda t) + \Pi_k(4) e^{-\lambda t}.$$

Thus the PDF is:

$$f\left(t \mid \vec{\Pi}_k\right) = \frac{d\left(1 - R\left(t \mid \vec{\Pi}_k\right)\right)}{dt}. \tag{23}$$

Given the model parameters $\Lambda = (p_{01}, p_{02}, k_1, k_2, \lambda)$ and $\Psi = (\mu_0, \mu_1, \Sigma_0, \Sigma_1)$, (22) and (23) can be used to update the conditional reliability function and the PDF using the condition monitoring data obtained from each sampling epoch. Taking file 88 to file 93 for analysis, the corresponding conditional reliability functions are shown in Figure 14 and the probability density functions in Figure 15, where "$*$" represents the actual residual values corresponding to each sampling file. As can be seen from Figure 15, from file 89, the RUL distribution of the gear system is highly concentrated; the main cause is that with the steady accumulation of condition monitoring data, the uncertainty of the remaining distribution lifetime gradually decreases while the accuracy of the RUL prediction continually improves.

The MRL of the gear system is given by:

$$\mu_{k\Delta} = E\left(\xi - k\Delta > t \mid \xi > k\Delta, \vec{\Pi}_k\right)$$
$$= \frac{1}{\lambda}\left(2\Pi_k(1)(1 + p_{01}) + \Pi_k(2)(1 + 2p_{01})\right. \tag{24}$$
$$\left. + 2\Pi_k(3) + \Pi_k(4)\right).$$

The data from file 1 to file 95 are selected for MRL prediction. After each sampling completed, the life prediction is performed when the new observations are available, and then the predicted value is compared with the values that are obtained using HMM. Meanwhile, as illustrated in Table 4, the relative error (RE) between actual remaining useful lives and the prediction results using different models are calculated. As can be seen in Table 4, with the increase of collected data, the RE in the prediction results using HSMM are smaller than those using HMM, which is closer to the actual values.

4. Optimal Bayesian Control Scheme

4.1. Optimal Bayesian Control Chart Based on HSMM. In this section, we design a multivariate Bayesian control chart to detect the early fault of the helicopter gearbox. The optimal control limit is used to determine the stopping time, of which the objective is to minimize the long-run expected average cost per unit time. In partially observable Markov decision process (POMDP), it is well known that the posterior probability that the system is in the unhealthy state provides enough information for maintenance decision making [13]. Let Π_k^1 be the posterior probability that the helicopter gear system is in the unhealthy state:

$$\Pi_k^1 = \sum_{i \in K_2} \Pi_k(i) = \Pi_k(3) + \Pi_k(4), \tag{25}$$

where the initial value $\Pi_0^1 = 0$.

TABLE 4: Comparison of RUL prediction and RE using different models.

File number	ARL	RUL		RE	
		HSMM	HMM	HSMM	HMM
1	19.34	19.801	19.800	0.024	0.024
2	19.206	19.013	19.942	0.010	0.038
3	19.071	18.203	19.801	0.046	0.038
4	18.937	18.019	19.689	0.049	0.040
5	18.803	16.981	19.332	0.097	0.028
6	18.669	15.002	19.340	0.196	0.036
7	18.535	13.200	19.185	0.288	0.035
⋮	⋮	⋮	⋮	⋮	⋮
88	7.662	8.013	18.343	0.045	1.394
89	7.527	8.003	18.271	0.063	1.427
90	7.393	7.830	18.001	0.059	1.435
91	7.259	7.482	7.701	0.031	0.061
92	7.125	7.450	7.645	0.045	0.073
93	6.990	7.308	7.632	0.046	0.091
94	6.856	7.225	7.545	0.054	0.100
95	6.722	7.031	7.416	0.046	0.103
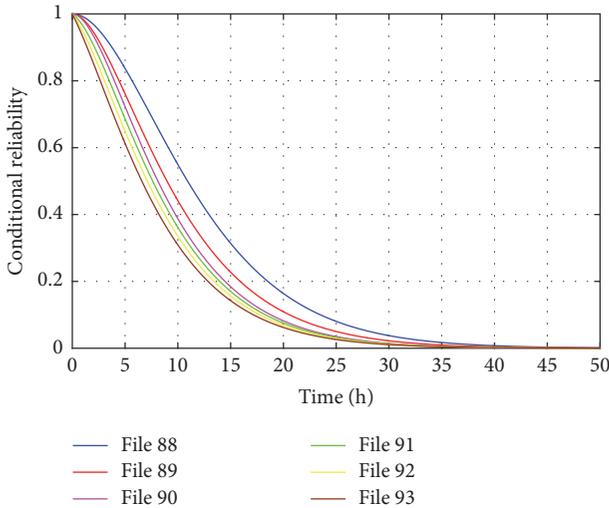
FIGURE 14: Conditional reliability function in files 88~93.

FIGURE 15: Probability density function in files 88~93.

By renewal theory, the cost minimization problem with fixed sampling interval is equivalent to seeking an optimal value of $\Pi^* \in [0, 1]$, such that

$$g\left(\Pi^*\right) = \inf_{\Pi \in [0,1]} \frac{E_{\Pi^*}\left(\text{CC}\right)}{E_{\Pi^*}\left(\text{CL}\right)}, \qquad (26)$$

where CC and CL denote, respectively, the cycle cost and cycle length.

The Bayesian control scheme is illustrated in Figure 16. For the fixed control limit $\overline{\Pi} \in (0, 1)$, at each decision epoch, the posterior probability vector $\vec{\Pi}_k$ is updated by Bayes' theorem. At the sampling time $k\Delta$, if $\Pi_k^1 < \overline{\Pi}$, the system will continue to run, of which the cost for each sampling is C_S. If $\Pi_k^1 \geq \overline{\Pi}$, system should be stopped and full inspection

is performed to determine whether the system is in state 0 or not, for which the inspection cost rate and inspection time are, respectively, C_I and T_I. The gear system has probability Π_k^1 to be in state 1 and probability $1 - \Pi_k^1$ to be in state 0. If the system is found to be in state 0 after a full inspection, then no maintenance measures are taken and the system continues operating in the healthy condition. If the system is found to be in state 1, preventive maintenance measures will be taken with the maintenance cost rate C_{PM} and the maintenance time T_{PM}. If the posterior probability Π_k^1 is not beyond the fixed control limit $\overline{\Pi}$ but still fails, that is, a random failure occurs, we need to immediately take failure replacement measures with corresponding maintenance cost rate and replacement time C_F and T_F, respectively. When the system takes full inspection, preventive maintenance, or failure replacement, the lost production cost rate C_{LP} will be incurred. When the system operates in the warning state, additional operating
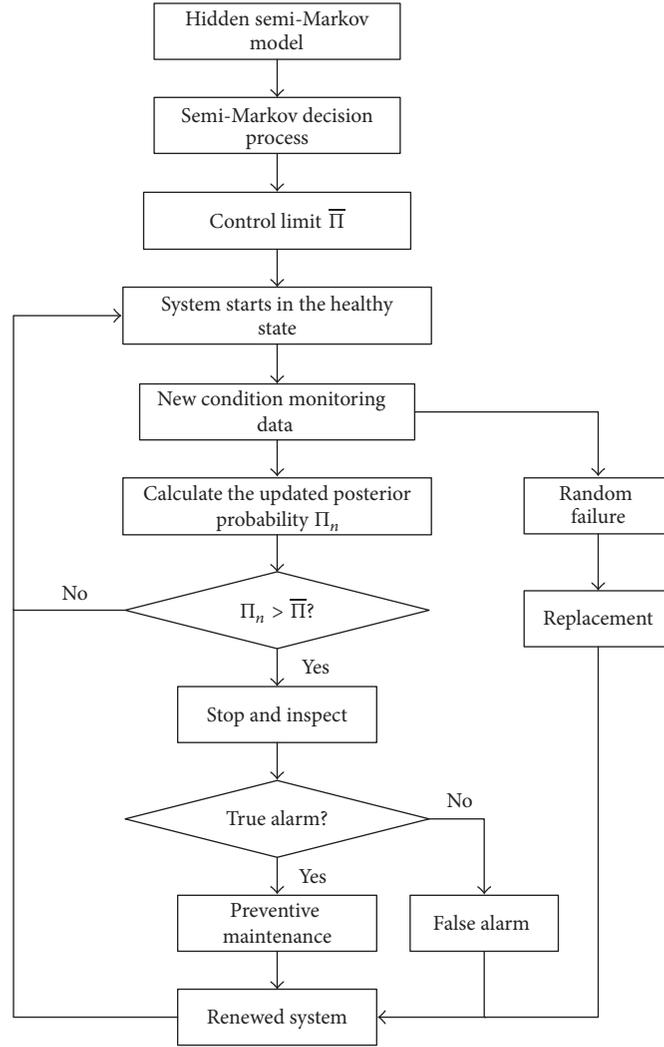
FIGURE 16: Schematic of Bayesian control approach.

cost rate C_{AO} and maintenance cost rate C_{AM} will be generated. We further assume that after the inspection, the preventive maintenance, or failure replacement, the system will begin from a new cycle.

We design an effective algorithm in the semi-Markov decision process (SMDP) framework to obtain the optimal control limit. In order to discretize the interval $[0, 1]$, it is sufficient to choose the number of subintervals $L = 30$ to provide effective accuracy. For $L = 30$, we define the SMDP in state $(i_1, i_2, \ldots, i_{k_1+k_2-1})$; that is, if the posterior probability in state $\{1\}$ lies in the interval $[(i_1 - 1)/L, i_1/L]$, the posterior probability in state $\{2\}$ lies in the interval $[(i_2 - 1)/L, i_2/L], \ldots$, and the posterior probability in state $\{k_1 + k_2 - 1\}$ lies in the interval $[(i_{k_1+k_2-1} - 1)/L, i_{k_1+k_2-1}/L]$. If $\Pi_k^1 \geq \overline{\Pi}$, the preventive maintenance is performed, then the SMDP is defined in state PM. If the system fails, the SMDP is defined in state F. Let $\overline{S} = \{(i_1, i_2, \ldots, i_{k_1+k_2-1}) : \sum_{m=1}^{k_1+k_2-1} i_m \leq L, \ i_m \in \mathbb{N}^+\}$, then the state space for the SMDP is defined as $S = \{(1, 0, \ldots, 0)\} \cup \overline{S} \cup \{\text{PM}\} \cup \{F\}$, where $(1, 0, \ldots, 0)$ presents the gear system is in state one at the initial moment.

For the fixed control limit $\overline{\Pi}$, $g(\overline{\Pi})$ can be obtained by solving the following linear equations:

$$v_x = c_x - g\left(\overline{\Pi}\right)\tau_x + \sum_{x' \in S} P_{x,x'} v_{x'} \quad \text{for } x, x' \in S, \tag{27}$$

$$v_0 = 0, \tag{28}$$

where v_x is the relative value until the next decision epoch given the current state $x \in S$. c_x is the expected cost incurred until the next decision epoch given the current state $x \in S$. τ_x is the expected sojourn time until the next decision epoch given the current state $x \in S$. $P_{x,x'}$ is the probability that at the next decision epoch the system will be in state $x' \in S$ given the current state is $x' \in S$.

Next, we derive the closed form expressions of each quantity in (27). For each $(i_1, i_2, \ldots, i_{k_1+k_2-1})$, $(i'_1, i'_2, \ldots, i'_{k_1+k_2-1}) \in \overline{S}$, and $j_a^- = (i'_a - 1)/L$, $j_a^+ = i'_a/L$ for $a = 1, \ldots, k_1 + k_2 - 1$,

the transition probabilities of SMDP are calculated by the following formula:

$$
\begin{aligned}
P_{(i_1,i_2,\ldots,i_{k_1+k_2-1})(i'_1,i'_2,\ldots,i'_{k_1+k_2-1})} &= P\left(j_1^- \le \Pi_k(1) < j_1^+,\ldots,\right. \\
j_{k_1+k_2-1}^- &\le \Pi_k(k_1+k_2-1) < j_{k_1+k_2-1}^+, \xi > k\Delta \mid \xi \\
&> (k-1)\Delta, \vec{\Pi}_{k-1}\right) = \sum_{m=1}^{k_1+k_2} P\left(j_1^- \le \Pi_k(1)\right. \\
&< j_1^+,\ldots,j_{k_1+k_2-1}^- \le \Pi_k(k_1+k_2-1) < j_{k_1+k_2-1}^+ \mid \xi \\
&> k\Delta, \vec{\Pi}_{k-1}, \Theta_{k\Delta} = m\right) \times P\left(\Theta_{k\Delta} = m \mid \vec{\Pi}_{k-1}\right) \\
&\times R\left(\Delta \mid \vec{\Pi}_{k-1}\right).
\end{aligned}
\tag{29}
$$

From (10) we can further obtain:

$$
\begin{aligned}
\frac{f\left(y_{k\Delta} \mid \mu_1, \Sigma_1\right)}{f\left(y_{k\Delta} \mid \mu_0, \Sigma_0\right)} \\
= \left(|\Sigma_0| \cdot |\Sigma_1|^{-1}\right)^{1/2} \exp\left(-\frac{1}{2}(V_k + C)\right),
\end{aligned}
\tag{30}
$$

where $V_k = (Y_{k\Delta} - B)'A(Y_{k\Delta} - B)$, $A = \Sigma_1^{-1} - \Sigma_0^{-1}$, $B = (\Sigma_1^{-1} - \Sigma_0^{-1})^{-1}(\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0)$, $C = (\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_0^T\Sigma_0^{-1}\mu_0) - B^T(\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0)$.

Thus (18) can be expressed as follows:

$$
\Pi_k(m) = \begin{cases}
\dfrac{c_{\Pi_{k-1}}^m}{\sum_{\forall j \in K_1} c_{\Pi_{k-1}}^j + \left(f\left(y_{k\Delta} \mid \mu_1, \Sigma_1\right)/f\left(y_{k\Delta} \mid \mu_0, \Sigma_0\right)\right)\sum_{\forall j \in K_2} c_{\Pi_{k-1}}^j} & \text{if } m \in K_1 \\[4ex]
\dfrac{c_{\Pi_{n-1}}^m}{\left(f\left(y_{k\Delta} \mid \mu_0, \Sigma_0\right)/f\left(y_{k\Delta} \mid \mu_1, \Sigma_1\right)\right)\sum_{\forall j \in K_1} c_{\Pi_{k-1}}^j + \sum_{\forall j \in K_2} c_{\Pi_{k-1}}^j} & \text{if } m \in K_2,
\end{cases}
\tag{31}
$$

where $c_{\Pi_{k-1}}^m = \sum_{n=1}^m \Pi_{k-1}(n)P_{nm}(\Delta)$.

In (29), for $m \in K_1$, $P(j_m^- \le \Pi_k(m) < j_m^+ \mid \xi > k\Delta, \vec{\Pi}_{k-1})$ is computed as follows:

$$
\begin{aligned}
P\left(j_m^- \le \Pi_k(m) < j_m^+ \mid \xi > k\Delta, \vec{\Pi}_{k-1}\right) &= \sum_{m=1}^{k_1+k_2} P\left(-2\right. \\
&\cdot \ln\left(\frac{\left(c_{\Pi_{k-1}}^m - j_m^- \sum_{j \in K_1} c_{\Pi_{k-1}}^j\right)\left(|\Sigma_0|^{-1} \cdot |\Sigma_1|\right)^{1/2}}{j_m^- \sum_{j \in K_2} c_{\Pi_{k-1}}^j}\right) \\
&- C \le V_k < -2 \\
&\cdot \ln\left(\frac{\left(c_{\Pi_{k-1}}^m - j_m^+ \sum_{j \in K_1} c_{\Pi_{k-1}}^j\right)\left(|\Sigma_0|^{-1} \cdot |\Sigma_1|\right)^{1/2}}{j_m^+ \sum_{j \in K_2} c_{\Pi_{k-1}}^j}\right) \\
&\left.- C \mid \Theta_{k\Delta} = m\right) \frac{c_{\Pi_{k-1}}^m}{\sum_{j=1}^{k_1+k_2} c_{\Pi_{k-1}}^j} = \sum_{m=1}^{k_1+k_2} P\left(T\left(j_m^-\right)\right. \\
&\left.\le V_k < T\left(j_m^+\right) \mid \Theta_{k\Delta} = m\right) \frac{c_{\Pi_{k-1}}^m}{\sum_{j=1}^{k_1+k_2} c_{\Pi_{k-1}}^j}.
\end{aligned}
\tag{32}
$$

For $m \in K_2$, the probability $P(j_m^- \le \Pi_k(m) < j_m^+ \mid \xi > k\Delta, \vec{\Pi}_{k-1})$ can be derived using the similar method in (32) and is omitted. So the first item in (29) can be written as follows:

$$
\begin{aligned}
P\left(j_1^- \le \Pi_k(1) < j_1^+,\ldots,j_{k_1+k_2-1}^- \le \Pi_k(k_1+k_2-1)\right. \\
\left.< j_{k_1+k_2-1}^+ \mid \xi > k\Delta, \vec{\Pi}_{k-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{m=1}^{k_1+k_2} P\left(\max\left(T\left(j_1^-\right),\ldots,T\left(j_{k_1+k_2-1}^-\right)\right) \le V_k\right. \\
&\left.< \min\left(T\left(j_1^+\right),\ldots,T\left(j_{k_1+k_2-1}^+\right)\right) \mid \Theta_{k\Delta} = m\right) \\
&\cdot \frac{c_{\Pi_{k-1}}^m}{\sum_{j=1}^{k_1+k_2} c_{\Pi_{k-1}}^j}.
\end{aligned}
\tag{33}
$$

Since $Y_{k\Delta} - B \mid \Theta_{k\Delta} \in K_1 \sim N(\mu_0 - B, \Sigma_0)$, $Y_{k\Delta} - B \mid \Theta_{k\Delta} \in K_2 \sim N(\mu_1 - B, \Sigma_1)$, the probability in (33) can be calculated using Theorem 3.1 proposed by Provost and Rudiuk [24]. Next, the other transition probabilities in SMDP are derived.

If $\Pi_{k-1}^1 \ge \Pi^*$, after full inspection, the probability of system in state 1, that is, true alarm occurs, is given by

$$
P_{(i_1,i_2,\ldots,i_{k_1+k_2-1}),\text{PM}} = \Pi_{k-1}^1 = \Pi_k(3) + \Pi_k(4).
\tag{34}
$$

Also, after full inspection, the gear system may be found in healthy state; the probability is given by

$$
\begin{aligned}
P_{(i_1,i_2,\ldots,i_{k_1+k_2-1}),(1,0,\ldots,0)} &= 1 - \Pi_{k-1}^1 \\
&= 1 - \Pi_k(3) - \Pi_k(4).
\end{aligned}
\tag{35}
$$

If $\Pi_{k-1}^1 < \Pi^*$, the probability of failure for the gear system is given by

$$
\begin{aligned}
P_{(i_1,i_2,\ldots,i_{k_1+k_2-1}),F} &= 1 - R\left(\Delta \mid \vec{\Pi}_{k-1}\right) = 1 - \Pi_{k-1}(1) \\
&\cdot e^{-\lambda\Delta}\left(1 + \lambda\Delta + p_{01}\frac{(\lambda\Delta)^2}{2!} + p_{01}\frac{(\lambda\Delta)^3}{3!}\right)
\end{aligned}
$$

$$- \Pi_{k-1}(2) e^{-\lambda\Delta} \left(1 + p_{01}\lambda\Delta + p_{01}\frac{(\lambda\Delta)^2}{2!} \right)$$

$$- \Pi_{k-1}(3) e^{-\lambda\Delta} (1 + \lambda\Delta) - \Pi_{k-1}(4) e^{-\lambda\Delta}. \tag{36}$$

The remainder transition probabilities are given by

$$P_{F,(1,0,\ldots,0)} = 1,$$
$$P_{PM,(1,0,\ldots,0)} = 1. \tag{37}$$

If $\Pi^1_{k-1} \geq \Pi^*$, the corresponding expected cost and expected sojourn time are given by

$$c_{(i_1,i_2,\ldots,i_{k_1+k_2-1})} = (C_I + C_{LP}) \cdot T_I, \tag{38}$$
$$\tau_{(i_1,i_2,\ldots,i_{k_1+k_2-1})} = T_I.$$

If $\Pi^1_{k-1} < \Pi^*$, the expected cost and expected sojourn time can be computed as follows:

$$c_{(i_1,i_2,\ldots,i_{k_1+k_2-1})}$$
$$= C_s \sum_{(i'_1,i'_2,\ldots,i'_{k_1+k_2-1})\in\bar{S}} P_{(i_1,i_2,\ldots,i_{k_1+k_2-1}),(i'_1,i'_2,\ldots,i'_{k_1+k_2-1})}$$
$$+ (C_F + C_{LP}) \cdot T_F \cdot P_{(i_1,i_2,\ldots,i_{k_1+k_2-1}),F}$$
$$+ (C_{AO} + C_{AM}) \int_0^\Delta \sum_{j\in K_2}\sum_{i=1}^{j} \Pi_{k-1}(i) P_{ij}(t)\, dt, \tag{39}$$

$$\tau_{(i_1,i_2,\ldots,i_{k_1+k_2-1})}$$
$$= \Delta \sum_{(i'_1,i'_2,\ldots,i'_{k_1+k_2-1})\in\bar{S}} P_{(i_1,i_2,\ldots,i_{k_1+k_2-1})(i'_1,i'_2,\ldots,i'_{k_1+k_2-1})}$$
$$+ \int_0^\Delta (T_F + t) f\left(t \mid \vec{\Pi}_{k-1}\right) dt.$$

If the gear system is in PM or F state, the corresponding expected costs and expected sojourn times are given by

$$c_{PM} = (C_P + C_{LP}) \cdot T_{PM},$$
$$c_F = (C_F + C_{LP}) \cdot T_F,$$
$$\tau_{PM} = T_{PM}, \tag{40}$$
$$\tau_F = T_F.$$

By setting the maintenance time parameters $T_I = 3\,$h, $T_{PM} = 15\,$h, and $T_F = 30\,$h and maintenance cost parameters $C_S = 20$, $C_I = 100$, $C_{PM} = 300$, $C_F = 1000$, $C_{LP} = 300$, $C_{AO} = 100$, and $C_{AM} = 200$, we coded the computational algorithm in MATLAB R2015b to calculate the different expected average costs under different control limits $\overline{\Pi}$. The optimal control limit was $\Pi^* = 0.71$ and the corresponding average cost was equal to 113.59. The Bayesian control chart of the helicopter gear system is shown in Figure 17. It is observed that based on HSMM the posterior probability Π^1_k firstly exceeds the optimal control limit at file 89, which indicates that the system need to be stopped to take full inspection at this moment.
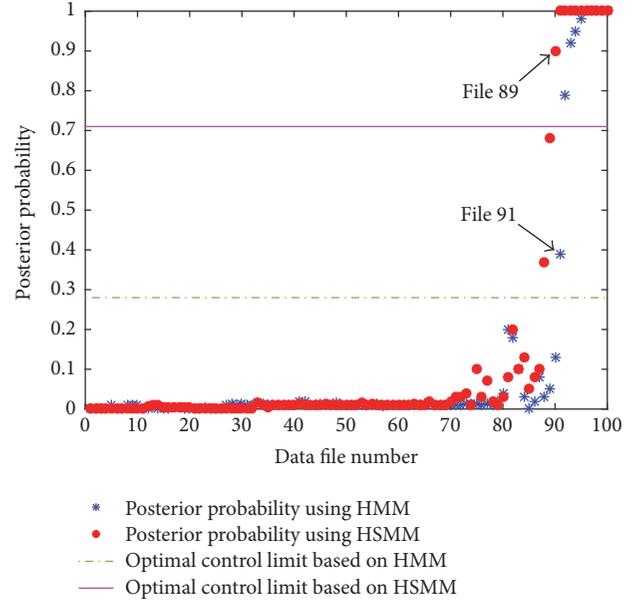
FIGURE 17: Optimal multivariate Bayesian control chart of the gearbox.

4.2. Comparison with Other Maintenance Policies. Using the same maintenance time and cost parameters, the optimal control limit $\Pi^* = 0.28$ was obtained in the Bayesian control chart based on HMM, in which the sojourn time in each hidden states follows exponential distribution. The corresponding average cost was equal to 121.72. It should be noted that the average cost concluded by HSMM was less than that by HMM. As shown in Figure 17, while based on HMM, the posterior probability Π^1_k firstly exceeds the optimal control limit at file 91. The multivariate Bayesian control chart indicates that the proposed Bayesian control scheme can detect the fault of the gear of interest much earlier than the method based on HMM.

Next, we considered the traditional age-based replacement policy, which is well known that it does not take into account condition monitoring information. For the helicopter gearbox, the expected cycle cost (CC) under age-based replacement policy is the sum of failure cost, additional operating and maintenance costs while the system operates in unhealthy state and the preventive maintenance cost:

$$E_\tau(CC) = (C_F + C_{LP}) T_F \left(1 - R\left(\tau \mid \vec{\Pi}_0\right) \right)$$
$$+ (C_{AO} + C_{AM}) \int_0^\tau P_{13}(t) + P_{14}(t)\, dt \tag{41}$$
$$+ (C_P + C_{LP}) T_P R\left(\tau \mid \vec{\Pi}_0\right),$$

where τ is the stopping time.

The expected cycle length (CL) is the sum of the expected time under normal operation, replacement time caused by failure, and preventive maintenance time:

$$E_\tau(CL) = \int_0^\tau R\left(t \mid \vec{\Pi}_0\right) dt + T_F \left(1 - R\left(\tau \mid \vec{\Pi}_0\right) \right)$$
$$+ T_P R\left(\tau \mid \vec{\Pi}_0\right). \tag{42}$$

TABLE 5: The optimal maintenance cost under different maintenance policies.

Policy	Optimal parameters		Expected average cost
	Π^*	τ^*	
Bayesian (HSMM)	0.71	—	113.59
Bayesian (HMM)	0.28	—	121.72
Aged-based replacement	—	17.24	140.23

Then the long-run expected average cost per unit time under age-based replacement policy is equivalent to finding the optimal stopping time τ^*, such that

$$\tau^* = \arg\inf \frac{E_\tau\,(\text{CC})}{E_\tau\,(\text{CL})}. \tag{43}$$

Table 5 shows the optimal values obtained under different policies, which are, respectively, the multivariate Bayesian control chart based on HSMM and HMM and the age-based replacement policy. It is interesting to find that the expected average cost using the Bayesian control chart based on the HSMM is considerably lower than the other policies, while the cost obtained by the age-based replacement policy is the highest due to the irrespective of condition monitoring information. This is reasonable because, for the helicopter gear system, the cost incurred by failure is always higher than the cost incurred by inspection and preventive maintenance measures.

5. Conclusions

In this paper, we have proposed a novel optimal Bayesian control scheme for the early fault detection of the partially observable helicopter gear system. The GMR signal was selected for data preprocessing to eliminate the influence of load variation from the original signal of the helicopter gearbox. The healthy portion of GMR signal was used to fit a VAR model; thus the residuals of whole data history were obtained. Without the restriction that the sojourn time in each hidden state was exponentially distributed, the general Erlang distribution was considered for modeling the gear system's sojourn time in each of the hidden states in a 3-state HSMM. The EM algorithm was employed to estimate the unknown parameters of the model. Using HSMM to describe the deterioration process of the gear system, several important quantities for the gearbox residual life prediction, such as conditional reliability distribution function, PDF and MRL were derived in terms of the posterior probability that the system was in each hidden state. The comparison with the life prediction method based on HMM indicated that the relative errors were smaller using HSMM. An optimal multivariate Bayesian control chart with cost-effective criterion was developed. Moreover, the optimal control limit was solved in the SMDP framework. The comparison with the Bayesian control chart based on HMM as well as the age-based replacement policy indicated that the proposed multivariate Bayesian control scheme provided accurate early fault

detection to the helicopter gear system with the minimum cost.

In this research, we have considered the multivariate Bayesian control chart under fixed sampling interval; further improvement can be obtained by considering two or more sampling intervals or the availability maximization criterion to obtain the optimal control limit, which can be a suitable topic for the future work.

Appendix

Formula (6) can be rewritten as a state space model; the corresponding state and observation equations are given by

$$\begin{aligned} \alpha_n &= D + T\alpha_{n-1} + E_n, \\ Z_n &= H\alpha_n, \end{aligned} \tag{A.1}$$

where

$$\begin{aligned} \alpha_n &= \left(Z_n, Z_{n-1}, \ldots, Z_{n-\widehat{p}+1} \right)', \\ D &= \left(\widehat{\delta}, 0, \ldots, 0 \right)', \\ E_n &= \left(\varepsilon_n, 0, \ldots, 0 \right)', \\ H &= \left(I_2, 0, \ldots, 0 \right), \\ T &= \begin{pmatrix} \widehat{\Phi}_1 & \widehat{\Phi}_2 & \cdots & \widehat{\Phi}_{\widehat{p}} \\ I_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_2 & 0 \end{pmatrix}, \end{aligned} \tag{A.2}$$

and ε_n are i.i.d. $N_2(0, \widehat{C})$.

For each $m \geq 0$, we define

$$\begin{aligned} \alpha_{n+m|n} &= E\left(\alpha_{n+m} \mid \overrightarrow{Z}_n \right), \\ P_{n+m|n} \\ &= E\left(\left(\alpha_{n+m} - \alpha_{n+m|n}\right)\left(\alpha_{n+m} - \alpha_{n+m|n}\right)^T \mid \overrightarrow{Z}_n \right), \\ \eta_{n+m|n} &= Z_{n+m} - E\left(Z_{n+m} \mid \overrightarrow{Z}_n \right), \\ f_{n+m|n} &= E\left(\eta_{n+m|n} \eta_{n+m|n}^T \mid \overrightarrow{Z}_n \right). \end{aligned} \tag{A.3}$$

Then, the following recursive equations of Kalman filter are given by

$$\begin{aligned} \alpha_{n+1|n} &= D + T\alpha_{n-1}, \\ f_{n+1|n} &= HP_{n+1|n}H', \end{aligned}$$

$$P_{n+1|n} = TP_{n|n}T' + \widehat{C},$$

$$\alpha_{n+1|n+1} = \alpha_{n+1|n} + \left(P_{n+1|n}H'\right)\left(f_{n+1|n}^{-1}\right)\eta_{n+1|n},$$

$$\eta_{n+1|n} = Z_{n+1} - H\alpha_{n+1|n},$$

$$P_{n+1|n+1} = P_{n+1|n} - \left(P_{n+1|n}H'\right)\left(f_{n+1|n}^{-1}\right)HP_{n+1|n} \tag{A.4}$$

and the initial values are given by

$$\alpha_{0|0} = \left(I_{2p} - T\right)^{-1}D,$$

$$P_{0|0} = \text{vec}^{-1}\left[\left(I_{(2p)^2} - T \otimes T\right)^{-1}\text{vec}\left(\widehat{C}\right)\right]. \tag{A.5}$$

For each $n \leq \widehat{p}$, using the recursive equations above we obtain

$$Y_n = Z_n - E_{\widehat{\gamma}}\left(Z_n \mid \overrightarrow{Z}_{n-1}\right) = \eta_{n|n-1}. \tag{A.6}$$

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—reviews, methodology and applications," *Mechanical Systems and Signal Processing*, vol. 42, no. 1-2, pp. 314–334, 2014.

[2] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: a review on data driven approaches," *Mathematical Problems in Engineering*, vol. 2015, Article ID 793161, 17 pages, 2015.

[3] Z. K. Zhu, R. Yan, L. Luo, Z. H. Feng, and F. R. Kong, "Detection of signal transients based on wavelet and statistics for machine fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1076–1097, 2009.

[4] J. Luo, D. Yu, and M. Liang, "Application of multi-scale chirplet path pursuit and fractional Fourier transform for gear fault detection in speed up and speed-down processes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4971–4986, 2012.

[5] C. Kar and A. R. Mohanty, "Vibration and current transient monitoring for gearbox fault detection using multiresolution Fourier transform," *Journal of Sound and Vibration*, vol. 311, no. 1-2, pp. 109–132, 2008.

[6] W. Wang and A. K. Wong, "Autoregressive model-based gear fault diagnosis," *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 124, no. 2, pp. 172–179, 2002.

[7] S. Rofe, "Signal processing methods for gearbox fault detection," Tech. Rep., Defence Science and Technology Organization, Canberra, Australia, 1997.

[8] J. Igba, K. Alemzadeh, C. Durugbo, and K. Henningsen, "Performance assessment of wind turbine gearboxes using in-service data: current approaches and future trends," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 144–159, 2015.

[9] K. Lyu, X. Tan, G. Liu, and C. Zhao, "Sensor selection of helicopter transmission systems based on physical model and sensitivity analysis," *Chinese Journal of Aeronautics*, vol. 27, no. 3, pp. 643–654, 2014.

[10] Y. Shao and C. K. Mechefske, "Gearbox vibration monitoring using extended Kalman filters and hypothesis tests," *Journal of Sound and Vibration*, vol. 325, no. 3, pp. 629–648, 2009.

[11] K. Barker and H. Baroud, "Proportional hazards models of infrastructure system recovery," *Reliability Engineering and System Safety*, vol. 124, pp. 201–206, 2014.

[12] Z. Wang, C. Hu, W. Wang, Z. Zhou, and X. Si, "A case study of remaining storage life prediction using stochastic filtering with the influence of condition monitoring," *Reliability Engineering and System Safety*, vol. 132, pp. 186–195, 2014.

[13] M. J. Kim, R. Jiang, V. Makis, and C.-G. Lee, "Optimal Bayesian fault prediction scheme for a partially observable system subject to random failure," *European Journal of Operational Research*, vol. 214, no. 2, pp. 331–339, 2011.

[14] M. Gales and S. Young, "The application of hidden Markov Models in speech recognition," *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195–304, 2007.

[15] F. Cartella, J. Lemeire, L. Dimiccoli, and H. Sahli, "Hidden semi-Markov models for predictive maintenance," *Mathematical Problems in Engineering*, vol. 2015, Article ID 278120, 23 pages, 2015.

[16] D. N. Nguyen, L. Dieulle, and A. Grall, "Remaining useful lifetime prognosis of controlled systems: a case of stochastically deteriorating actuator," *Mathematical Problems in Engineering*, vol. 2015, Article ID 356916, 16 pages, 2015.

[17] Q. Liu, M. Dong, and Y. Peng, "A novel method for online health prognosis of equipment based on hidden semi-Markov model using sequential Monte Carlo methods," *Mechanical Systems and Signal Processing*, vol. 32, pp. 331–348, 2012.

[18] Q. Liu, M. Dong, W. Lv, X. Geng, and Y. Li, "A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis," *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 217–232, 2015.

[19] R. Jiang, M. J. Kim, and V. Makis, "Maximum likelihood estimation for a hidden semi-markov model with multivariate observations," *Quality and Reliability Engineering International*, vol. 28, no. 7, pp. 783–791, 2012.

[20] M. A. Girshick and H. Rubin, "A Bayes approach to a quality control model," *Annals of Mathematical Statistics*, vol. 23, pp. 114–125, 1952.

[21] S. Panagiotidou and G. Tagaras, "Optimal integrated process control and maintenance under general deterioration," *Reliability Engineering and System Safety*, vol. 104, pp. 58–70, 2012.

[22] G. C. Reinsel, *Elements of Multivariate Time Series Analysis*, Springer Series in Statistics, Springer-Verlag, New York, 2nd edition, 1997.

[23] A. Khaleghei and V. Makis, "Model parameter estimation and residual life prediction for a partially observable failing system," *Naval Research Logistics*, vol. 62, no. 3, pp. 190–205, 2015.

[24] S. B. Provost and E. M. Rudiuk, "The exact distribution of indefinite quadratic forms in noncentral normal vectors," *Annals of the Institute of Statistical Mathematics*, vol. 48, no. 2, pp. 381–394, 1996.