

Article

A Novel Health Prognosis Method for a Power System Based on a High-Order Hidden Semi-Markov Model

Qinming Liu ^{1,*} , Daigao Li ¹, Wenyi Liu ¹, Tangbin Xia ² and Jiaxiang Li ¹

¹ Department of Industrial Engineering, Business School, University of Shanghai for Science and Technology, 516 Jungong Road, Shanghai 200093, China; 203781388@st.usst.edu.cn (D.L.); 193860959@st.usst.edu.cn (W.L.); 1913520624@st.usst.edu.cn (J.L.)

² SJTU-Fraunhofer Center, State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; xtbxtb@sjtu.edu.cn

* Correspondence: qmliu@usst.edu.cn

Abstract: Power system health prognosis is a key process of condition-based maintenance. For the problem of large error in the residual lifetime prognosis of a power system, a novel residual lifetime prognosis model based on a high-order hidden semi-Markov model (HOHSMM) is proposed. First, HOHSMM is developed based on the hidden semi-Markov model (HSMM). An order reduction method and a composite node mechanism of HOHSMM based on permutation are proposed. The health state transition matrix and observation matrix are improved accordingly. The high-order model is transformed into the corresponding first-order model, and more node dependency information is stored in the parameter group to be estimated. Secondly, in order to estimate the parameters and optimize the structure of the proposed model, an intelligent optimization algorithm group is used instead of the expectation–maximization (EM) algorithm. Thus, the simplification of the topology of the high-order model by the intelligent optimization algorithm can be realized. Then, the state duration variables in the high-order model are defined and deduced. The prognosis method based on polynomial fitting is used to predict the residual lifetime of the power system when the prior distribution is unknown. Finally, the intelligent optimization algorithm is used to solve the proposed model, and experiments are performed based on a set of power system data sets to evaluate the performance of the proposed model. Compared with HSMM, the proposed model has better performance on the power system health prognosis problem and can get a relatively good solution in a short computation time.

Keywords: high-order hidden semi-Markov model; composite node; model reduction; state duration; polynomial fitting; residual life prognosis



Citation: Liu, Q.; Li, D.; Liu, W.; Xia, T.; Li, J. A Novel Health Prognosis Method for a Power System Based on a High-Order Hidden Semi-Markov Model. *Energies* **2021**, *14*, 8208. <https://doi.org/10.3390/en14248208>

Academic Editor: Andrea Mariscotti

Received: 28 October 2021

Accepted: 3 December 2021

Published: 7 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of technology, the operation of power systems has an increasing demand for higher reliability, lower environmental risks, and higher human safety. Power system failures usually come at the cost of high maintenance costs and uncertain downtime. However, it is difficult to obtain accurate health status and predict failures of a power system in time. Thus, power system health prognosis is an important topic in reliability and maintenance engineering, which determines how to properly integrate positioning degradation data into power system fault detection and failure prevention [1]. Health prognosis involves evaluating the current state, classifying the current state of several failure modes, and predicting the residual lifetime of the power system. Residual lifetime refers to the remaining life from the current health status to the functional failure of the system.

As a probabilistic statistical method, the hidden Markov model (HMM) has a good randomness representation ability and potential structural relationship description ability. It is widely used in the field of complex system modeling. Carey first migrated HMM

from the field of speech recognition to the field of power system fault diagnosis, creating a new research direction of fault diagnosis [2]. Researchers proposed an improved hidden Semi-Markov model (HSMM) to get rid of the constraints of HMM characteristics, which further divided the macro-states into micro-states, and made a priori assumptions about the time distribution of each macro-state to successfully realize the application of HMM in the field of residual lifetime prediction. Liu et al. proposed an HSMM of information fusion, completed the diagnosis and life prediction of equipment, and achieved good results [3].

System health prognosis is a complex process, and the real-time operation of the power system is particularly difficult. Most of the methods in the field of reliability involve intensive calculation and processing of a large number of historical data [4,5]. The methods are usually divided into three categories: model-driven models, physical models, and data-driven models.

The design of the physical model considers the operating conditions to model the operating process of the system. The models integrate the physical features and numeric features of the system under monitoring and develop the mappings between parameters and prognostics features, such as rotating machinery [6,7] and lithium-ion batteries [8]. However, it is difficult to guarantee its accuracy. There is no perfect physical model to adapt to the system.

The data-driven model is suitable for systems with little prior knowledge or a complex structure because of the characteristics of using existing data to predict the health of the system. Moreover, they rely solely on the past observed trajectories [9], including neural networks [10–12], support vector machines [13,14], and Gaussian process regression [15]. Finding the effective part of the original data is an arduous task. The main disadvantages of this model are slow convergence and the ease in which it falls into local minimums; these shortcomings limit the application of this model.

Model-driven models focus on the prediction level fusion of information. For the first time, Dong et al. used HSMM to conduct research on various sensor diagnoses [16]. Then, several variants of Markovian-based models were applied to the diagnosis and prognostics of equipment. Yan et al. used HSMM-based equipment health estimation and proposed several new methods to effectively predict equipment RUL [17]. Li et al. proposed an improved HMM that is improved through performance degradation and successfully applied it to the reliability evaluation of wind turbine bearings [18]. Huang et al. proposed an improved HMM based on a predictive neural network and applied it to a motor drive system to evaluate the proposed model [19]. These studies performed well in different domains, whereas the modification was mostly done in the modeling part, and little consideration was given to signal preprocessing and parameters estimation; furthermore, the recognizing and training processes of model-driven models are usually time-consuming, so they are applied to offline health prognosis.

In order to better refer to the historical statistical information to improve the model recognition rate, a high-order hidden Markov model (HOHMM) was proposed by researchers. For the problem of parameter explosion and a more complex derivation of the high-order model, researchers have established an order reduction algorithm, ORED, to simplify HOHMM. Basically, the fast incremental algorithm was proposed to train HOHMM and the three problems were discussed, respectively [20]. Compared with the successive order reduction of the ORED algorithm, Hadar proposed a more complex algorithm based on the idea of equivalent transformation, which transformed any high-order model into the corresponding first-order model [21]. The above researchers provide a general research method for the research of the high-order model. However, in the process of reducing the order of HOHMM to HMM, it is necessary to re-derive each variable of the model and re-evaluate the parameters based on the Baum–Welch algorithm. With the increase of the order of the model, the workload will increase explosively. Dong et al. established the HO-HMMAR model, deduced variables, and better solved the optimal portfolio problem [22]. Heng et al. established the daily average temperature evolution model and assumed that the asymmetric component was a high-order hidden Markov

process. The results showed that the model can effectively capture the characteristics of temperature data under various conditions. Polynomial fitting was a simple tool for nonlinear fitting, which has efficient data processing ability [23]. Ritesh et al. used the coefficients of polynomial fitting to generate the feature vector of iris recognition and verified the effectiveness of the polynomial method through the benchmark IITD and casia-v4 interval iris data [24]. With the popularization of neural networks and various machine learning methods, the flashpoint of polynomial fitting became dimmer and dimmer in the field of fault diagnosis. However, the polynomial fitting is one of the few methods that can directly write the analytical formula of a nonlinear relationship. It cannot be done by the neural network.

The problem of parameter estimation has always been the most difficult link of the probability model. The EM algorithm with a heavy workload depends on the initial value and can easily fall into local optimization; thus, more and more researchers are turning their attention to various intelligent optimization algorithms. The algorithm of Yu [13] and other genetic algorithms were combined within the grey model. The numerical results showed that the new method combined with a genetic algorithm (GA) can greatly improve the prediction accuracy. Krause [14] et al. used the function approximator based on a GA to estimate the parameters of the induction motor, which achieved better results than other methods. Zhang [15] and others used the artificial immune algorithm to optimize HMM and obtained the initial observation matrix with the highest identification degree. Zhang [16] and others used an adaptive genetic particle swarm optimization algorithm to optimize HSMM and obtained more accurate results than traditional HSMM. How an intelligent simulation algorithm can be used to simplify the research process of HMM is a point that should be paid attention to in future research.

In order to obtain more accurate and practical predictions under the premise of unknown distribution, the contribution of this paper is to propose an improved HOHSMM for the health prognosis of the power system. First, the framework of HOHSMM based on HSMM is established. By considering the ORED algorithm and Hadar's equivalent transformation, a model reduction method based on permutation is proposed, which uses the definition of the high-order hidden Markov group model to transform it into the corresponding first-order model, and the solution of three problems of the low-order model can be used in the high-order complex model. Moreover, the transition probability matrix and observation probability matrix are deformed so that the interdependence information of nodes in the high-order complex model is naturally integrated into the model parameters, and the effect of simplifying the model is achieved. Then, the auxiliary resident variables are defined and deduced, and the parameter group carrying more dependency information is estimated by using the intelligent optimization algorithm group. The maximizing occurrence probability of observation is described as objective, the decomposition dependency of the high-order model is represented by the parameter group, and the complexity is transferred from the model itself to the parameter group. The polynomial fitting method is used to fit each resident variable sequence, and the residual lifetime of the power system is predicted when the distribution is unknown. Finally, the proposed model is verified and evaluated with one power system data set. The results show that the method proposed in this paper is feasible and effective.

This article aims to develop a more superior power system health prognostics method. The paper is organized as follows: Section 2 introduces first-order and high-order HMM, and Section 3 develops an improved high-order hidden semi-Markov model. Then, the residual lifetime prognosis method of this paper is proposed in Section 4. Section 5 analyzes and discusses the case study of this article, and finally the article is concluded in Section 6.

2. Hidden Markov Model

2.1. First-Order HMM

The first-order HMM can be described as $\lambda = (\pi, A, B)$. It is composed of observable nodes and hidden state nodes. The health state of the node at time t is represented by

$State_t$, then the transition between the hidden state of the hidden state node conforms to the Markov property.

$$\text{Prob}(State_t|State_1 \cdots State_{t-1}) = \text{Prob}(State_t|State_{t-1})$$

Note that the total number of health states is N and the total number of observations is M . π is the initial state distribution vector in the model, and it is the probability value of each state of power system when $t = 1$, $\pi = [\bar{\pi}]_{1 \times N}$. A is the transition probability matrix, and $A = [a_{ij}]_{N \times N}$. a_{ij} is the probability value from state i to state j , and $a_{ij} = \text{Prob}(State_{t+1} = j|State_t = i)$, $\sum_{i=1}^N a_{ij} = 1$, $\sum_{j=1}^N a_{ij} = 1$. B is the observation probability matrix, and $B = [b_i(k)]_{N \times M}$; it is the probability value of the observation generated by the node under state i and $\text{Prob}(o_t = k|State_t = i)$. o_t represents the observation value at time t , and meets $\sum_{i=1}^N b_i(k) = 1$ and $\sum_{k=1}^M b_i(k) = 1$.

In order to solve the evaluation problem of HMM, that is, $\text{Prob}(O|\lambda)$, forward-backward variables based on HMM are usually developed, and the evaluation problem is decomposed into recursive expressions of forward-backward variables. In order to solve the learning problem, the Baum-Welch algorithm is usually used to solve the parameter set $-\lambda$, which may optimize the model and produce the current observations. Its idea is to establish the Lagrange multiplier equation between the current parameter group λ and the maximum parameter group λ , and calculate the partial derivative under the constraints of each parameter to optimize the current parameters. In order to solve the prognosis problem, dynamic programming is applied to HMM to produce the Viterbi algorithm to find the most likely health state path. Due to its own characteristics, the conventional HMM has the defects that it must obey the exponential distribution and cannot describe the deterioration of the power system.

2.2. The High-Order HMM

The high-order hidden Markov model (HOHMM) is a generalization of the first-order Markov model, and it retains more historical statistical information. It is assumed that the current health state of the research object is related to the previous health states. Taking an n -order hidden Markov model as an example, its health state at time t ($t > n + 1$) is related to the health state at the previous n times, and it is expressed as

$$\text{Prob}(State_t|State_1 \cdots State_{t-1}) = \text{Prob}(State_t|State_{t-n} \cdots State_{t-1})$$

HOHMM is different from the conventional HMM, and it is described as $\lambda = (\pi, A, B, \check{A}, \check{B})$, where \check{A} is the state transition matrix only applicable to the previous n times and \check{B} is the observation probability matrix only applicable to the previous n times. HOHMM also has three problems from HMM. However, due to relatively complex node dependencies, the dependencies become more complex with the increase of orders, and the model parameters will increase exponentially. Thus, the research on HOHMM generally focuses on model order reduction. HOHMM improves the model recognition rate on the basis of retaining more historical statistical information, but it still fails to overcome the shortcomings of HMM.

3. Improved High-Order Hidden Semi-Markov Model

In this paper, by considering the shortcomings of HMM and the advantages of high-order modeling, an improved high-order hidden semi-Markov model (HOHSMM) is proposed based on HSMM.

Taking the second-order HSMM as an example, the model can be described as $\lambda = (\pi, A, B, \check{A}, \check{B})$. It is usually assumed that each sub-state conforms to the same time distribution, and the topology of a second-order HSMM is described in Figure 1.

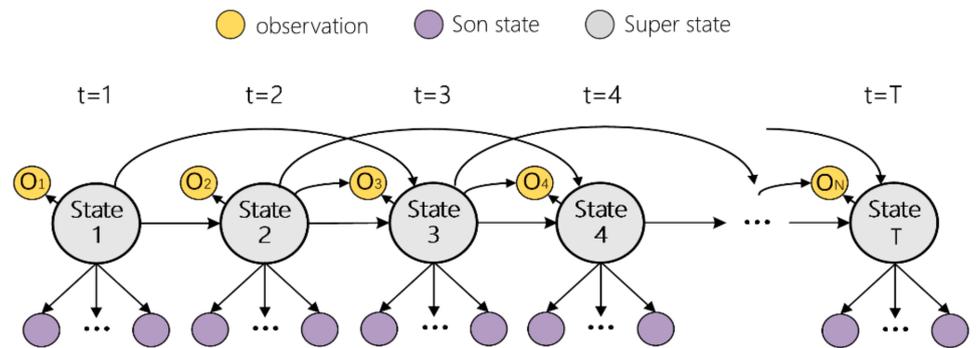


Figure 1. The illustration of two-order HSMM.

3.1. The Model Order Reduction Based on Permutation Mapping

Similarly, taking the second-order HOHSMM as an example, due to the structural changes of the second-order model, the model parameters and related algorithms can be changed accordingly. The respective algorithms for solving the three problems of the low-order model are not applicable to the high-order model. Thus, this paper proposes a model reduction method based on combinatorial mapping, and it is essential to merge the hidden state nodes corresponding to two adjacent time points in the second-order model into one node; then, the merged nodes can be modeled by the Markov process, as shown in Figure 2.

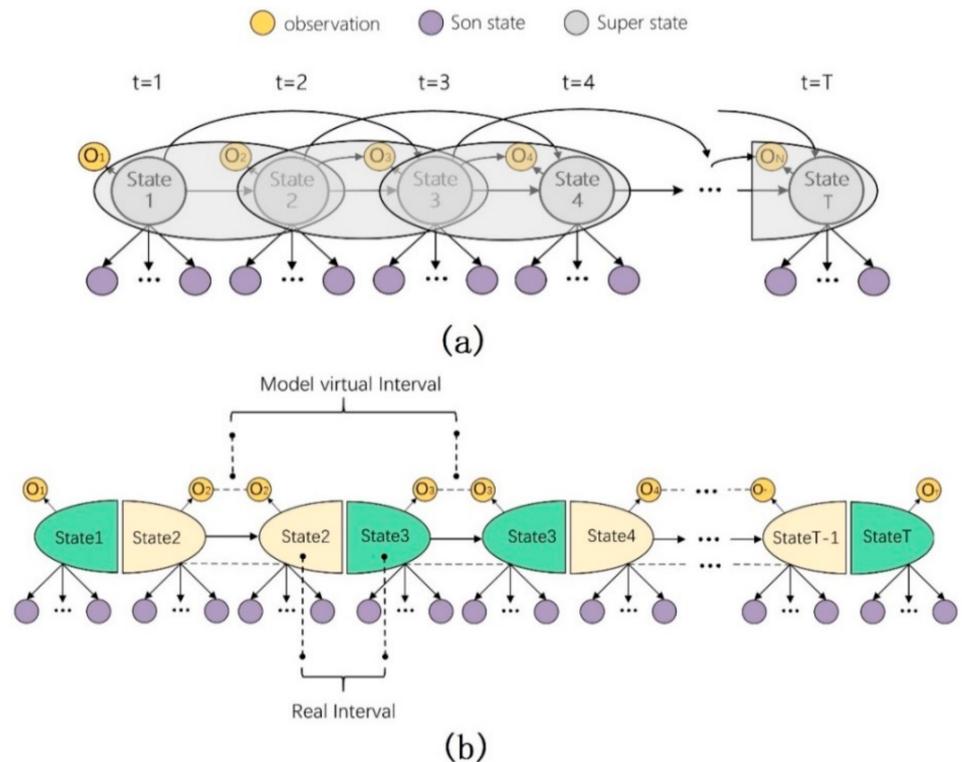


Figure 2. The illustration of the model reduction based on permutation. (a): Division of compound nodes; (b): Markov chain model with reduced order after compound nodes.

In this paper, a health state in HSMM generates a segment of observations, as opposed to a single observation in the HMM. Thus, the states in a segment semi-Markov model are called super states. Each super state consists of several single states, which are called son states, which can be seen in Figure 2. Figure 2a is a division of the model on the second-order HOHSMM, and the adjacent state nodes can be combined into a new, larger

node. The relationship between the two health states in the new node has no impact on the Markov property of the whole model. The Markov property of the new model after order reduction can be described as Equation (1).

$$\text{Prob}(State_t, State_{t-1} | State_{t-1} \cdots State_1) = \text{Prob}(State_t, State_{t-1} | State_{t-1}, State_{t-2}) (t \geq 3) \tag{1}$$

The description of time in the new model also changes t to $(t, t + 1)$. If the first element of the time combination $(t, t + 1)$ is the unique index of the time of the new model, the time can also be expressed as \check{t} . \check{t} is different from t in the mathematical sense, but in the physical sense, $\check{t} = t$. When the implicit state node of the new model combination time \check{t} is expressed as $(State_{\check{t}}, State_{t+1})$, it generates two groups of observations, $o_{\check{t}}$ and $o_{\check{t}+1}$, where, $o_{\check{t}+1}$ completely depends on $(State_{\check{t}}, State_{t+1})$ and $o_{\check{t}}$ completely depends on $(State_{\check{t}-1}, State_t)$. Moreover, at time $\check{t} - 1$, the observations connected by dotted lines are the same observation in Figure 2b, and the sub-state set connected by dotted lines is also the same sub-state set. Thus, each group of observations can be obtained from the unique combined hidden state node, while the unique determined observations corresponding to the combined hidden state can be only retained (except the initial time) in its topology.

In the reduced order model, taking the combination of hidden states as the modeling object, different states are essentially an arrangement problem. Based on the above partial definitions of HMM parameters, if a second-order HOHSM has N different super states, and considering the power system has performance degradation and the state is irreversible, then there are $C_N^2 + N$ states appearing in the original model. In this paper, a simple mapping between the arrangement scheme and natural numbers is introduced, as shown in Equation (2), where i, j are states, respectively.

$$Index = \text{Mapping}(i, j) = 4i + j - \sum_{z=0}^i (i, j \in (0, 1 \cdots N - 1), j \geq i) \tag{2}$$

In the parameters of the second-order HOHSM original model, the transition probability matrix becomes the transition probability cube of $N \times N \times N$ and the observation probability matrix becomes the observation probability cube of $N \times N \times M$. The initial probability distribution remains unchanged and the initial transition probability matrix and the initial observation probability matrix are the same as the first-order model. The new model parameters after order reduction are composed of a transition probability matrix $(C_N^2 + N) \times (C_N^2 + N)$, observation probability matrix $(C_N^2 + N) \times M$, initial probability distribution, initial transition probability matrix, and initial observation probability matrix. Letting the reduced-order transition probability matrix be \hat{A} with the element $\hat{a}_{ij}(i, j \in (0, \cdots N - 1))$, the reduced-order observation probability matrix is \hat{B} , and the element is $\hat{b}_{ij}(i, j \in (0, \cdots N - 1))$. For $\lambda = (\pi, A, B, \check{A}, \check{B})$, the element of π is $\pi_i(i \in (0, \cdots N - 1))$, the element of \check{A} is $\check{a}_{ij}(i, j \in (0, \cdots N - 1))$, and the element of \check{B} is $\check{b}_i(j)(i, j \in (0, \cdots N - 1))$.

Based on the order reduction method in Section 3.1, the transition probability matrix $\hat{A} ((C_N^2 + N) \times (C_N^2 + N))$ after dimension reduction can be obtained. The transition probability matrix \hat{A} is sparse, and the actual amount of effective data in the matrix that is not 0 is

$$\sum_{j=0}^{N-1} (-j^2 + jN + N - j) \tag{3}$$

where j is the second state in the state arrangement (i, j) . Taking the second-order HOHSM with $n = 4$ as an example, the effective data of the reduced probability transfer matrix of order reduction are 20. In Table 1, the positions represented by 1 and 1* are the effective data, the composite state represented by the column header at the position of 1* is the main state, and the composite state represented by the column header at the position of 1 is defined as the transition state, while the location represented by 0 is invalid data.

Table 1. Sparse representation of two-order HOHSMM reduced-order transition probability matrix with 4 states.

From/To	Perm	(0,0)	(0,1)	(0,2)	(0,3)	(1,1)	(1,2)	(1,3)	(2,2)	(2,3)	(3,3)
Perm	index/index	0	1	2	3	4	5	6	7	8	9
(0,0)	0	1*	1	1	1	0	0	0	0	0	0
(0,1)	1	0	0	0	0	1*	1	1	0	0	0
(0,2)	2	0	0	0	0	0	0	0	1*	1	0
(0,3)	3	0	0	0	0	0	0	0	0	0	1
(1,1)	4	0	0	0	0	1*	1	1	0	0	0
(1,2)	5	0	0	0	0	0	0	0	1*	1	0
(1,3)	6	0	0	0	0	0	0	0	0	0	1
(2,2)	7	0	0	0	0	0	0	0	1*	1	0
(2,3)	8	0	0	0	0	0	0	0	0	0	1
(3,3)	9	0	0	0	0	0	0	0	0	0	1

*: Main state; Shadow: Valid data.

3.2. The Model Reasoning

In view of the idea of a forward-backward algorithm, a linger time (LT) mechanism is introduced, and an auxiliary variable $\xi_t(i)$ is established, as is described in Equation (4).

$$\xi_t(i, d) = \text{Prob}(O_{[1:t]}, LT((i, i)) = d | \lambda) \quad (t \geq d) \tag{4}$$

The probability that the observation sequence $O_{[1:t]}$ is generated at the cut-off time t and has stayed in the current state for d time under a given model parameter group λ can be obtained by Equation (4).

In this section, the sparse representation of the transition probability matrix of the general second-order HOHSMM is given, and the significance of the main state and the transition state can be described, respectively. The transition among the different main health states of the conventional second-order model is a gradual process.

Main state \rightarrow (Transition state sequence) \rightarrow Next main state

This transformation process needs at most $j - i + 2$ time points to be fully described. Taking the transition from the health state (0,0) to (3,3) as an example, the transition process needs five time points to be described, as shown in Figure 3.

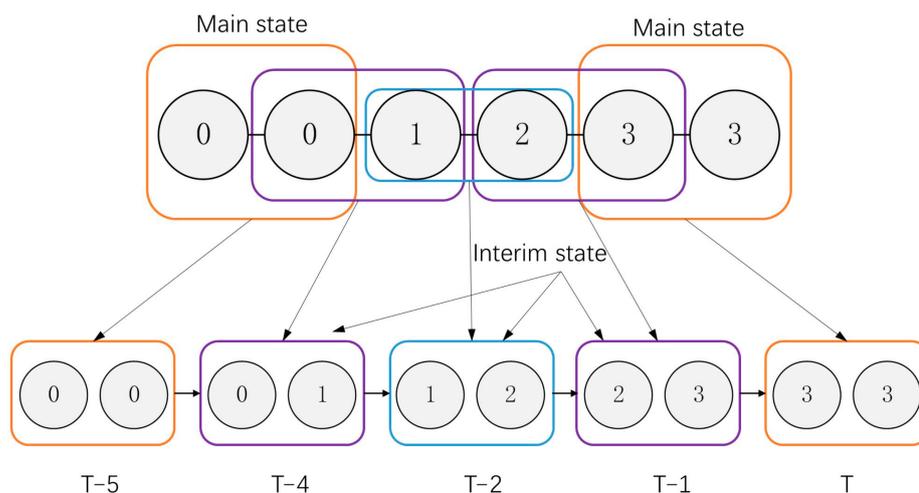


Figure 3. Schematic diagram of main state transition process.

In order to facilitate model reasoning, a transition variable $\vartheta_t(j, i, Road)$ is defined, representing the process intermediate value from the main state (j, j) to the time main state (i, i) of t time. The main states, by being transferred out, can be defined as acceptance states,

and the main states being transferred in can be defined as return states. \vec{t} is defined as the time point when the acceptance states appear.

(1) For $i = 0$ and $d = t$, $\zeta_t(i)$ can be described as

$$\zeta_t(0, d) = \text{Prob}(O_{[1:t]}, LT(0) = d | \lambda) \tag{5}$$

The recursive initial value is

$$\zeta_t(0, 1) = \pi_0 \check{b}_0(o_1) \check{b}_0(o_2) \check{a}_{00} \tag{6}$$

The inner transfer recursion is

$$\zeta_t(0, d) = \zeta_{t-1}(0, d-1) \hat{a}_{(0,0)(0,0)}(t-1) \hat{b}_{(0,0)}(o_t) \tag{7}$$

(2) For $i = 1$, $\zeta_t(i)$ can be described as

$$\zeta_t(1, d) = \text{Prob}(O_{[1:t]}, LT(1) = d | \lambda) \tag{8}$$

The acceptance state can only be (0,0) and the transition state can only be (0,1). Then, the acceptance value is $\zeta_{t-d}(0, t-d)$, $\vartheta_t(0, 1, Road) = \hat{a}_{(0,0)(0,1)}(t-d+1) \hat{b}_{(0,1)}(o_{t-d+1})$. Thus, the recursive median is described as $\zeta_t(1, 0) = \zeta_{t-d}(0, t-d) \hat{a}_{(0,0)(0,1)}(t-d+1) \hat{b}_{(0,1)}(o_{t-d+1}) \hat{a}_{(0,1)(1,1)}(t-d+2) \hat{b}_{(1,1)}(o_{t-d+2})$, and the recursive equation of internal transfer is

$$\zeta_t(1, d) = \zeta_{t-\Delta t}(1, d-1) \hat{a}_{(1,1)(1,1)}(t-\Delta t) \hat{b}_{(1,1)}(o_t) \tag{9}$$

(3) For $I = 2$, the acceptance states can be (0,0) and (1,1) and $\zeta_t(i)$ can be described as

$$\zeta_t(2, d) = \text{Prob}(O_{[1:t]}, LT(2) = d | \lambda) \tag{10}$$

The inner transfer recursion is

$$\zeta_t(2, d) = \zeta_{t-1}(2, d-1) \hat{a}_{(2,2)(2,2)}(t-1) \hat{b}_{(2,2)}(o_t) \tag{11}$$

Sit1: When the acceptance state is (0,0), it needs four time points to describe from (0,0) to (2,2) and the acceptance value is

$$\zeta_{\vec{t}}(0, \vec{t}) \vec{t} \in \{t-d-3, t-d-2\} \tag{12}$$

The corresponding $\vartheta_t(0, 2, Road)$ can be described as

$$\begin{cases} \hat{a}_{(0,0)(0,1)}(\vec{t}) \hat{b}_{(0,1)}(o_{\vec{t}+1}) \hat{a}_{(0,1)(1,2)}(\vec{t}+1) \hat{b}_{(1,2)}(o_{\vec{t}+2}) \hat{a}_{(1,2)(2,2)}(\vec{t}+2) \hat{b}_{(2,2)}(o_{\vec{t}+3}) \vec{t} = t-d-3 \\ \hat{a}_{(0,0)(0,2)}(\vec{t}) \hat{b}_{(0,2)}(o_{t-d-1}) \hat{a}_{(0,2)(2,2)}(\vec{t}+1) \hat{b}_{(2,2)}(o_{\vec{t}+2}) \vec{t} = t-d-2 \end{cases} \tag{13}$$

The corresponding return value is given by Equation (14).

$$\begin{cases} \zeta_{\vec{t}}(0, \vec{t}) \hat{a}_{(0,0)(0,1)}(\vec{t}) \hat{b}_{(0,1)}(o_{\vec{t}+1}) \hat{a}_{(0,1)(1,2)}(\vec{t}+1) \hat{b}_{(1,2)}(o_{\vec{t}+2}) \hat{a}_{(1,2)(2,2)}(\vec{t}+2) \hat{b}_{(2,2)}(o_{\vec{t}+3}) \vec{t} = t-d-3 \\ \zeta_{\vec{t}}(0, \vec{t}) \hat{a}_{(0,0)(0,2)}(\vec{t}) \hat{b}_{(0,2)}(o_{t-d-1}) \hat{a}_{(0,2)(2,2)}(\vec{t}+1) \hat{b}_{(2,2)}(o_{\vec{t}+2}) \vec{t} = t-d-2 \end{cases} \tag{14}$$

Sit2: When the acceptance state is (1,1), it needs three time points to describe from (1,1) to (2,2) and the corresponding $\vec{t} = t-d-2$ acceptance value is

$$\sum_{du=1}^{\vec{t}-1} \zeta_{\vec{t}}(1, d_u)$$

$\vartheta_t(1, 2, Road)$ is

$$\hat{a}_{(1,1)(1,2)}(\vec{t}) \hat{b}_{(0,1)}(o_{\vec{t}+1}) \hat{a}_{(1,2)(2,2)}(\vec{t} + 1) \hat{b}_{(2,2)}(o_{\vec{t}+2}) \tag{15}$$

The return value is

$$\sum_{du=1}^{\vec{t}-1} \xi_{\vec{t}}(1, d_u) \hat{a}_{(1,1)(1,2)}(\vec{t}) \hat{b}_{(0,1)}(o_{\vec{t}+1}) \hat{a}_{(1,2)(2,2)}(\vec{t} + 1) \hat{b}_{(2,2)}(o_{\vec{t}+2}) \tag{16}$$

(4) For $I = 3$, the acceptance states can be (0,0), (1,1), and (2,2) and $\xi_t(i)$ can be described as

$$\xi_t(3, d) = \text{Prob}(O_{[1,t]}, LT(3) = d | \lambda) \tag{17}$$

Inner transfer recursion is

$$\xi_t(3, d) = \xi_{t-1}(3, d - 1) \hat{a}_{(3,3)(3,3)}(t - 1) \hat{b}_{(3,3)}(o_t) \tag{18}$$

Sit1: When the acceptance state is (0,0), it needs five time points to describe from (0,0) to (3,3), and the corresponding acceptance value is

$$\xi_{\vec{t}}(0, \vec{t}) \vec{t} \in \{t - d - 4, t - d - 3, t - d - 2\} \tag{19}$$

$\vartheta_t(0, 3, Road)$ can be described as

$$\left\{ \begin{array}{l} \hat{a}_{(0,0)(0,1)}(\vec{t}) \hat{b}_{(0,1)}(o_{\vec{t}+1}) \hat{a}_{(0,1)(1,2)}(\vec{t} + 1) \hat{b}_{(1,2)}(o_{\vec{t}+2}) \hat{a}_{(1,2)(2,3)}(\vec{t} + 2) \hat{b}_{(2,3)}(o_{\vec{t}+3}) \hat{a}_{(2,3)(3,3)}(\vec{t} + 3) \hat{b}_{(3,3)}(o_{\vec{t}+4}) \vec{t} = t - d - 4 \\ \sum_{i=1}^2 \hat{a}_{(0,0)(0,i)}(\vec{t}) \hat{b}_{(0,i)}(o_{\vec{t}+1}) \hat{a}_{(0,i)(i,3)}(\vec{t} + 1) \hat{b}_{(i,3)}(o_{\vec{t}+2}) \vec{t} = t - d - 3 \\ \hat{a}_{(0,0)(0,3)}(\vec{t}) \hat{b}_{(0,3)}(o_{\vec{t}+1}) \hat{a}_{(0,3)(3,3)}(\vec{t} + 1) \hat{b}_{(3,3)}(o_{\vec{t}+2}) \vec{t} = t - d - 2 \end{array} \right. \tag{20}$$

The return value is the product of the corresponding acceptance value and the intermediate value.

Sit2: When the acceptance state is (1,1), it needs four time points to describe from (1,1) to (3,3), and the corresponding acceptance value is

$$\xi_{\vec{t}}(1, \vec{t}) \vec{t} \in \{t - d - 3, t - d - 2\} \tag{21}$$

The corresponding intermediate value $\vartheta_t(0, 3, Road)$ is

$$\left\{ \begin{array}{l} \hat{a}_{(1,1)(1,2)}(\vec{t}) \hat{b}_{(1,2)}(o_{\vec{t}+1}) \hat{a}_{(1,2)(2,3)}(\vec{t} + 1) \hat{b}_{(2,3)}(o_{\vec{t}+2}) \hat{a}_{(2,3)(3,3)}(\vec{t} + 2) \hat{b}_{(3,3)}(o_{\vec{t}+3}) \vec{t} = t - d - 3 \\ \hat{a}_{(1,1)(1,3)}(\vec{t}) \hat{b}_{(1,3)}(o_{\vec{t}+1}) \hat{a}_{(1,3)(3,3)}(\vec{t} + 1) \hat{b}_{(3,3)}(o_{\vec{t}+2}) \vec{t} = t - d - 2 \end{array} \right. \tag{22}$$

The recursive median is

$$\left\{ \begin{array}{l} \xi_{\vec{t}}(1, \vec{t}) \hat{a}_{(1,1)(1,2)}(\vec{t}) \hat{b}_{(1,2)}(o_{\vec{t}+1}) \hat{a}_{(1,2)(2,3)}(\vec{t} + 1) \hat{b}_{(2,3)}(o_{\vec{t}+2}) \hat{a}_{(2,3)(3,3)}(\vec{t} + 2) \hat{b}_{(3,3)}(o_{\vec{t}+3}) \vec{t} = t - d - 3 \\ \xi_{\vec{t}}(1, \vec{t}) \hat{a}_{(1,1)(1,3)}(\vec{t}) \hat{b}_{(1,3)}(o_{\vec{t}+1}) \hat{a}_{(1,3)(3,3)}(\vec{t} + 1) \hat{b}_{(3,3)}(o_{\vec{t}+2}) \vec{t} = t - d - 2 \end{array} \right. \tag{23}$$

Sit3: When the acceptance state is (2,2), it needs three time points to describe from (2,2) to (3,3), and the corresponding acceptance value is

$$\xi_{\vec{t}}(2, \vec{t}) \vec{t} \in \{t - d - 2\} \tag{24}$$

The corresponding intermediate value $\vartheta_t(2, 3, Road)$ is

$$\hat{a}_{(2,2)(2,3)}(\vec{t})\hat{b}_{(1,3)}(o_{\vec{t}+1})\hat{a}_{(2,3)(3,3)}(\vec{t}+1)\hat{b}_{(3,3)}(o_{\vec{t}+2}) \tag{25}$$

The recursive median is

$$\zeta_{\vec{t}}(2, \vec{t})\hat{a}_{(2,2)(2,3)}(\vec{t})\hat{b}_{(1,3)}(o_{\vec{t}+1})\hat{a}_{(2,3)(3,3)}(\vec{t}+1)\hat{b}_{(3,3)}(o_{\vec{t}+2}) \tag{26}$$

The general recursive equation of $\xi_t(i, d)$ can be expressed by Equation (27):

$$\zeta_t(i, d) = \sum_{j=0}^i \left[\left(\prod_{k=t-d}^t \hat{a}_{(i,i)(i,i)}\hat{b}_{(i,i)}(o_k) \right) \sum_{d_i=1}^{ct} \sum_{Road \in R} \xi_{ct}(j, d_i)\theta_t(j, i, Road) \right] \tag{27}$$

where $ct = t - d - lr + 1$, lr is the length of *Road*.

Based on the above classification and derivation of recursion, the auxiliary variables ζ_t in the higher state are decomposed into the expression of lower auxiliary variables. In the recursive process, all the transition process state paths meeting the main state transition can be generated, and $\xi_t(i, d)$ can be calculated by traversing all possible transition state paths. The transition state paths included by each main state are given in Table 2, where x is not included in the path length.

Table 2. Path generator for each main state transition process.

From	To	Road	From	To	Road
(0,0)	(1,1)	[0,0,1,1]	(0,0)	(3,3)	[x,x,0,0,3,3]
(0,0)	(2,2)	[x,0,0,2,2]	(1,1)	(2,2)	[1,1,2,2]
(0,0)	(2,2)	[0,0-2,2]	(1,1)	(3,3)	[1,1-3,3]
(0,0)	(3,3)	[0,0-3,3]	(1,1)	(3,3)	[x,1,1,3,3]
(0,0)	(3,3)	[x,0,0,1,3,3]	(2,2)	(3,3)	[2,2,3,3]
(0,0)	(3,3)	[x,0,0,2,3,3]	-	-	-

By the given model parameters, the probability of generating observations O is

$$\text{Prob}(O|\lambda) = \sum_{i=0}^{N-1} \sum_{d=1}^D \zeta_T(i, d)$$

Auxiliary variable $\tau_t(index)$ is the probability of being in Mapping⁻¹(*index*) at time t under the premise of the given model parameters and observations.

$$\tau_t(index) = \text{Prob}(\text{State}_t = \text{Mapping}^{-1}(index) | \lambda, O)$$

Where $\tau_0(index)$ can be obtained by calculating $\pi, \check{A}, \check{B}$, and the $\tau_t(index)$ recursive equation can be shown as follows.

$$\tau_{t+1}(index) = \frac{\text{prob}(o_{\vec{t}+1}, \text{State}_t = \text{Mapping}^{-1}(index) | \lambda)}{\sum_{indexi=0}^9 \text{prob}(o_{\vec{t}+1}, \text{State}_t = \text{Mapping}^{-1}(indexi) | \lambda)} = \frac{\hat{b}_{(index)}(o_{\vec{t}+1}) \sum_{i=0}^9 \tau_t(i)\hat{a}_{(i)(index)}}{\sum_{indexi=0}^9 \hat{b}_{(index)}(o_{\vec{t}+1}) \sum_{i=0}^9 \tau_t(i)\hat{a}_{(i)(indexi)}} \tag{28}$$

For the corresponding low-order model, the probability in a single state $i (i \in (1, 2, 3, 4))$ at time t can be described by the high-order model.

$$\text{Prob}(S_t = i) = \sum_{indexi \in I} \tau_{t-1}(indexi)$$

I is the index set corresponding to the second sub state i under different composite states, and the index correspondence is given in Table 3.

Table 3. Index correspondence.

Son State	Index Gather (I)	Perm Gather
0	{0}	{(0,0)}
1	{1,4}	{(0,1),(1,1)}
2	{2,5,7}	{(0,2),(1,2),(2,2)}
3	{3,6,8,9}	{(0,3),(1,3),(2,3),(3,3)}

For the parameter estimation problem, this paper selects the intelligent simulation algorithm group to replace the Baum–Welch algorithm and carries out a two-stage estimation. First, a one-stage likelihood function $L(\lambda, \check{A}, \check{B})$ is established, and it is described as

$$L(\lambda, \check{A}, \check{B}) = \text{Prob}(O_{[1:2]} | \lambda, \check{A}, \check{B})$$

For the situation of the current $\lambda, \check{A}, \check{B}$, if the probability of the first two observations is generated, then the optimal $\lambda, \check{A}, \check{B}$ in one stage is

$$\lambda, \check{A}, \check{B} = \underset{\lambda, \check{A}, \check{B}}{\text{argmax}} \text{Prob}(O_{[1:2]} | \lambda, \check{A}, \check{B})$$

The two-stage likelihood function is

$$L(A, B) = \text{Prob}(O_{[3:]} | \lambda, \check{A}, \check{B}, A, B)$$

In the situation of the current $\lambda, \check{A}, \check{B}, A, B$, the probability from the third to the final observation is generated, and the corresponding optimal parameters A, B in the second stage are

$$A, B = \underset{A, B}{\text{argmax}} \text{Prob}(O_{[3:]} | \lambda, \check{A}, \check{B}, A, B)$$

Different intelligent simulation algorithms are used to optimize and compare the two-stage likelihood function, and finally to select the best result.

4. Residual Life Prognosis under Uncertain Distribution

For the recursive reasoning on the duration of each state in Section 3.2, the system duration of each health state and the joint probability value of the observation can be obtained. In fact, there is a certain correlation between the observations generated by the system and the system duration in a single state, and it can be described by the observation and state transition matrix through a specific mode. In order to obtain the edge probability of the system duration in each health state, it is advisable to assume that system observation and dwell are independent of each other. For the above obtained $\check{\zeta}_t(i, d)$, by calculating the generation probability of the corresponding observation, the conditional probability equation can be obtained.

$$\text{Prob}(LT(i) = d | \lambda) = \frac{\check{\zeta}_t(i, d)}{\text{Prob}(O | \lambda)}$$

The time represented by each time point is used to calculate the probability that a certain state produces the duration, and a group of duration probability sequences are obtained. Generally, when the discrete data points are known, an a priori distribution assumption is conducive to study the continuity characteristics of the data. However, if the data distribution is incorrectly assumed, there will be great errors in the fitting and the original properties of the data will be lost. Thus, this paper proposes a method to fit

the data under an unknown distribution based on polynomial regression. The polynomial regression can be described as

$$f^n(x) = w_0 + w_1x + w_2x^2 + w_3x^3 \dots + w_mx^m$$

Letting the m -order polynomial function fitted by duration probability sequences generated by health i be described as $Lds_m^i(t)$, the residual lifetime of system at time t can be obtained.

$$RULL_1(t) = \sum_{i=0}^{N-1} \left[\int_t^D \frac{Lds_n^i(x)R_i}{\int_0^D Lds_n^i(y)dy} (x-t)dx + \sum_{j=i+1}^{N-1} \int_t^D xLds_m^j(x)R_jdx \right] \text{Prob}(s_t = i) \quad (29)$$

where, D is the maximum duration of each state and $\text{Prob}(s_t = i)$ is the probability in health state i at time t . R_i, R_j are the integral scaling coefficients after continuous discrete data.

5. Case Study

This paper verifies and evaluates the proposed model and method through the example of the system health diagnosis and life prediction of one power station. A LW42A-40.5 high voltage circuit breaker system was selected as the experimental platform; it was equipped with a CT14 spring operating mechanism, three-phase linkage during opening and closing, and SF6 was used as the arc extinguishing medium. An LW42A-40.5 high voltage circuit breaker system is mainly used in 35 kV power transmission and distribution systems for protection and control. The vibration signals of the system in the laboratory were collected by a hydraulic accelerometer installed parallel to the rotating shaft of the power system. In the application example, the power system was filled with 20, 40, 60, and 80 mg of micro dust, respectively, and a fixed length time window every 10 min to collect a vibration signal (P6) of about 1 min was used, as shown in Figure 4. Then, the vibration signal was divided into five layers by using 10 dB wavelets and the array of high-frequency and low-frequency wavelet coefficients were obtained. The dimensionally reduced wavelet coefficients were used as the input feature sequence vector of DGHMM. During the whole experiment, the states of the power system could be divided into four types: Baseline, Cont1, Cont2, and Cont3. Cont3 is the complete failure state of system. The whole experimental analysis platform was Python 3 and the platform running environment was Windows 10.

The power system state monitoring information was used to predict the state of health. In order to obtain a reliable health prognosis, the features to be monitored should be sensitive to vibration trends. We used wavelet transform to remove noise from the original signal and perform feature extraction in this paper [25]. It can generate a suitable framework to study the multi-scale transient representation of the signal. It is also good at time-frequency analysis and processing non-stationary signals. More than one characteristic should usually be used for a healthy prognosis. Here, a wavelet amplitude model based on the overall monitoring signal was used to demonstrate how to use features for health prognosis.

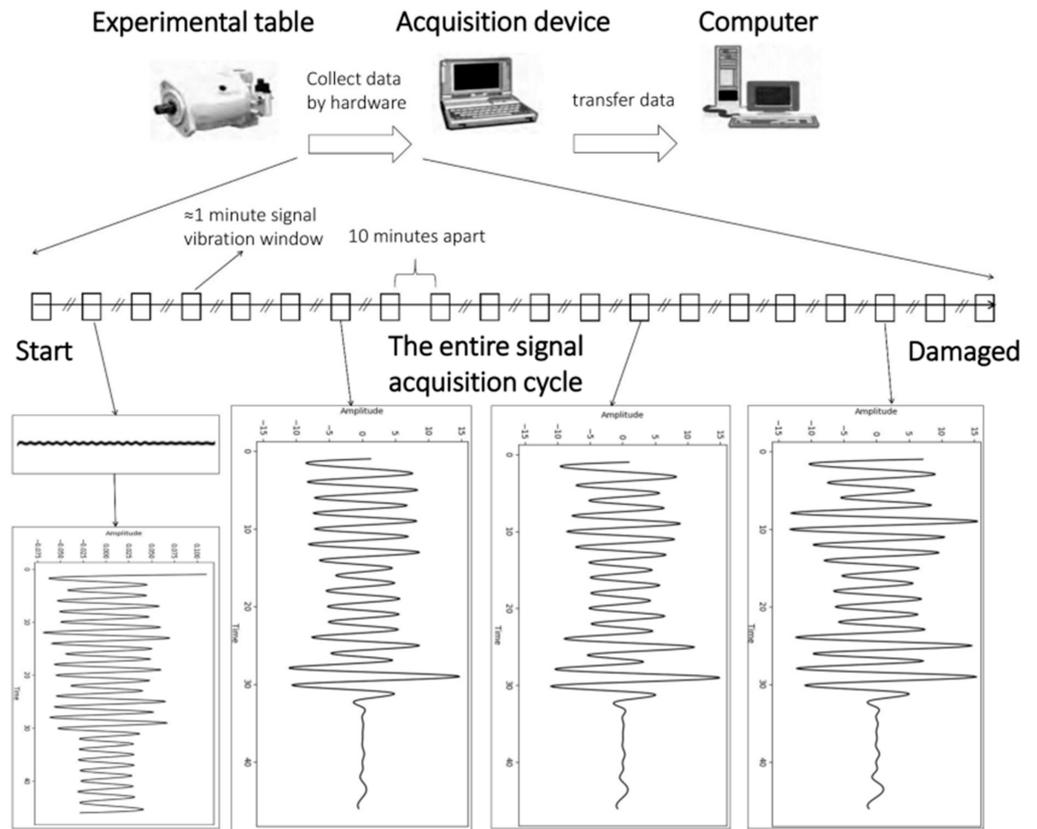


Figure 4. Experimental bench and signal acquisition.

5.1. Partial Data Preview

Using the vibration signal monitoring data from the hydraulic accelerometer, the health status diagnosis and residual life prognosis of the power system were carried out. Part of the vibration data (P6) after wavelet transform is shown in Table 4.

Table 4. Partial wavelet transform data of P6.

Time Spot	Sen 1	Sen 7	Sen 15	Sen 23	Sen 32
1	2.62	19.53	5.66	0.06	0.96
3	2.65	20.24	6.10	0.06	1.04
5	2.39	17.45	4.92	0.06	0.87
7	2.07	16.15	3.98	0.06	0.75
9	2.25	18.53	4.51	0.06	0.81
11	2.24	21.44	4.38	0.06	0.87
13	2.41	5.37	6.84	0.08	1.66
15	2.58	6.16	7.96	0.09	1.84
17	2.53	6.07	7.63	0.09	1.84
19	2.44	6.24	7.71	0.10	1.84
21	20.66	9.68	8.42	0.11	2.17
23	5.06	8.75	7.20	0.10	1.89
25	7.53	8.82	8.15	0.10	2.00
27	5.74	8.37	7.19	0.10	1.89
29	3.46	8.25	7.16	0.10	1.92
31	41.78	7.98	7.12	0.15	1.86
33	31.08	7.90	6.98	0.14	1.84
35	60.15	7.99	7.03	0.15	1.85
37	77.27	7.80	7.12	0.17	1.84

5.2. Parameter Estimation of the Intelligent Optimization Algorithm Group

Considering that the maximum expectation algorithm has a large dependence on the initial value, we used a genetic algorithm (GA), particle swarm algorithm (PSO), and artificial fish swarm algorithm (AFSA) to estimate the parameters of the proposed model under the same set of full observations, respectively.

The maximum number of iterations was 300, and the number of population (the number of particles in a particle swarm) was 30. The GA adopted a large parameter adaptive value encoding and decoding strategy. The PSO adopted the fixed inertia weight strategy, with an inertia weight of 0.5 and a learning factor of 2. The AFSA perception field was 1 and the crowding factor was 0.6. The objective optimization model was to maximize the occurrence probability of observation. The iterative process and results are shown in Figure 5. Under the premise of the same population number and iteration times, GA had good results with a continuous evolution and maximum likelihood. The likelihood values of the parameter groups found by the three optimization algorithms are shown in Table 5, and the reduced probability transition probability matrix is given in Table 6.

Table 5. The maximum likelihood of each algorithm.

Algorithm	Maximum Likelihood
GA	9.8665×10^{-8}
PSO	2.44412×10^{-13}
AFSA	7.22939×10^{-13}

Table 6. The optimal probability transfer matrix is represented sparsely.

From/To	Perm	(0,0)	(0,1)	(0,2)	(0,3)	(1,1)	(1,2)	(1,3)	(2,2)	(2,3)	(3,3)
Perm	index/ index	0	1	2	3	4	5	6	7	8	9
(0,0)	0	7.78×10^{-1}	4.45×10^{-2}	4.50×10^{-1}	2.67×10^{-2}	0.00	0.00	0.00	0.00	0.00	0.00
(0,1)	1	0.00	0.00	0.00	0.00	8.53×10^{-1}	1.40×10^{-1}	7.00×10^{-3}	0.00	0.00	0.00
(0,2)	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.44×10^{-1}	2.56×10^{-1}	0.00
(0,3)	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
(1,1)	4	0.00	0.00	0.00	0.00	8.66×10^{-1}	1.32×10^{-1}	2.45×10^{-3}	0.00	0.00	0.00
(1,2)	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.97×10^{-1}	3.44×10^{-3}	0.00
(1,3)	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
(2,2)	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.87×10^{-1}	1.13×10^{-1}	0.00
(2,3)	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
(3,3)	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

5.3. Residual Lifetime Prognosis

Based on the model reasoning, the duration of each state was analyzed, and the full probability equation was used for equivalent replacement in the calculation process. The different duration values of each state at different time points were separated according to the state to obtain the duration probability sequence of each state, and the polynomial regression was used to fit the sequence to obtain their respective analytical formulas. In the fitting process, the order m with less of a resonance effect on the prediction of the residual life of the subsequent state was preferentially selected (priority selection principle). The polynomial fitting of four different states at the final time point is given as an example in

Figure 6. The optimal values of fitting in the four states were 20, 15, 19, and 17 respectively, and the detailed polynomial regression coefficients are given in Table 7.

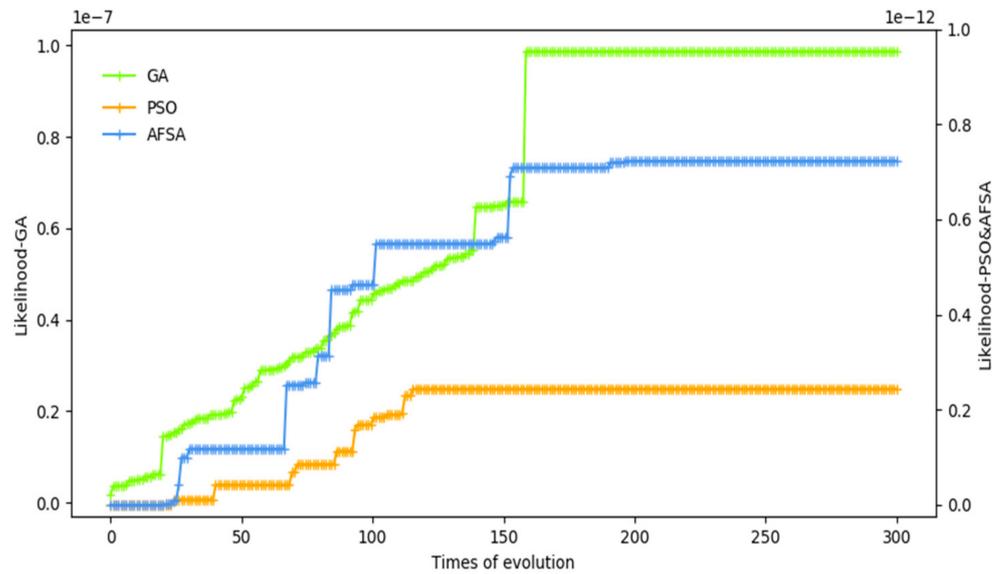


Figure 5. Swarm intelligence algorithm iteration process.

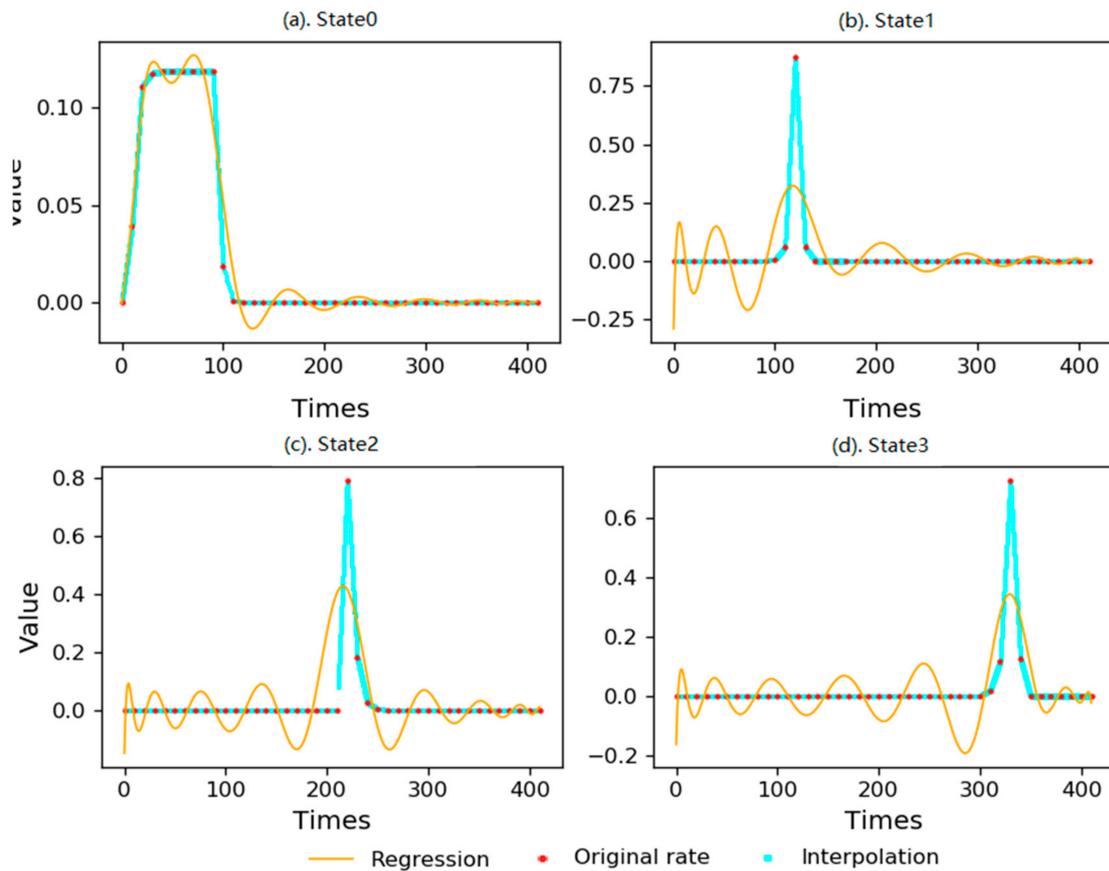


Figure 6. The polynomial fitting of each state at the final time.

Table 7. The polynomial fitting coefficients of each state at the final time.

Coeff. Index	State Type			
	State0(20)	State1(15)	State2(19)	State3(17)
0	-2.38819×10^{-45}	1.89034×10^{-32}	-3.01553×10^{-41}	-4.21246×10^{-37}
1	7.47909×10^{-42}	-6.14991×10^{-29}	1.02846×10^{-37}	1.23709×10^{-33}
2	-9.90244×10^{-39}	9.03133×10^{-26}	-1.54556×10^{-34}	-1.56243×10^{-30}
3	6.85507×10^{-36}	-7.90982×10^{-23}	1.31921×10^{-31}	1.0738×10^{-27}
4	-2.21019×10^{-33}	4.59824×10^{-20}	-6.68384×10^{-29}	-3.92341×10^{-25}
5	-1.73946×10^{-31}	-1.86838×10^{-17}	1.65986×10^{-26}	2.81649×10^{-23}
6	4.20376×10^{-28}	5.44157×10^{-15}	2.2419×10^{-24}	4.69387×10^{-20}
7	-1.12487×10^{-25}	-1.14607×10^{-12}	-3.79886×10^{-21}	-2.7376×10^{-17}
8	-3.1407×10^{-23}	1.73894×10^{-10}	1.72559×10^{-18}	8.37058×10^{-15}
9	3.36199×10^{-20}	-1.8709×10^{-8}	-4.83436×10^{-16}	-1.67201×10^{-12}
10	-1.2786×10^{-17}	1.3854×10^{-6}	9.40639×10^{-14}	2.29806×10^{-10}
11	2.98005×10^{-15}	-6.72669×10^{-5}	-1.31442×10^{-11}	-2.18798×10^{-8}
12	-4.69272×10^{-13}	1.982402×10^{-3}	1.32469×10^{-9}	1.41766×10^{-6}
13	5.09843×10^{-11}	$-3.1177213 \times 10^{-2}$	-9.50595×10^{-8}	-6.00431×10^{-5}
14	-3.77167×10^{-9}	$2.04641444 \times 10^{-1}$	4.72263×10^{-6}	1.5492×10^{-3}
15	1.82137×10^{-7}	$-2.9072277 \times 10^{-1}$	1.55018×10^{-4}	2.1526379×10^{-2}
16	-5.28119×10^{-6}	---	3.118172×10^{-3}	$1.26386828 \times 10^{-1}$
17	7.74612×10^{-5}	---	3.3825859×10^{-2}	$1.61790099 \times 10^{-1}$
18	4.00078×10^{-4}	---	$1.54167196 \times 10^{-1}$	---
19	4.512709×10^{-3}	---	$1.47377792 \times 10^{-1}$	---
20	6.50429×10^{-4}	---	---	---

In principle, negative values are not allowed for probability integration, but polynomial regression shows obvious fluctuation characteristics, positive and negative values have certain offset effect, and resonance also exists at the same time point in different states. It is not difficult to predict that the predicted residual life value may have certain fluctuation characteristics. In order to better fit the polynomial, the original data points were linearly interpolated combined with the characteristics of the original data. In Figure 7, the red point is the original data point, that is, the adaptive duration generation probability value, the light blue is the interpolation point, and the yellow line is the polynomial fitting line.

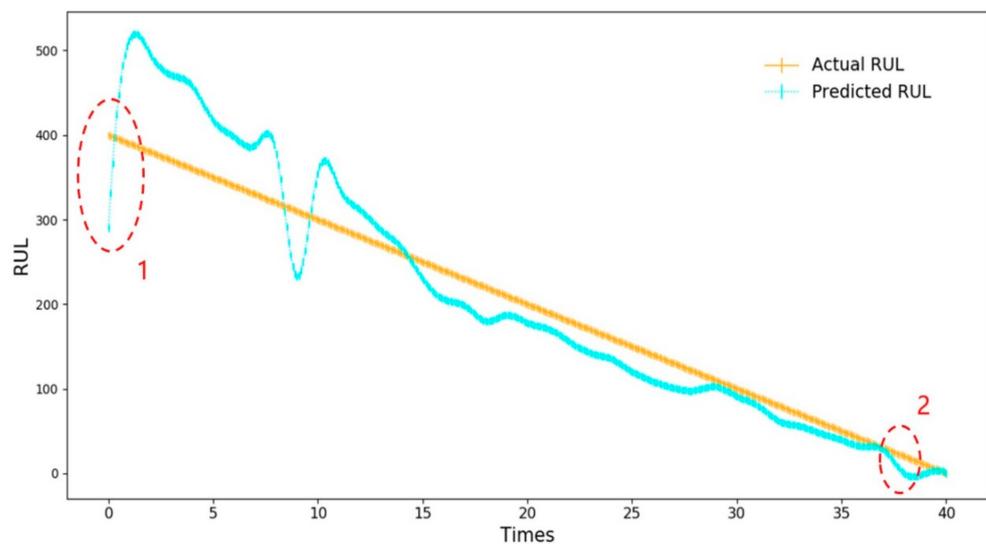


Figure 7. Residual lifetime prognosis of P6.

Finally, the residual life prognosis of P6 is shown in Figure 7, in which the discrete prediction points have been interpolated and smoothed. Mark 1 shows that there is a large

deviation in the predicted value at the final time point. Corresponding to the phenomenon of the initial fluctuation range of the polynomial regression formula in each subgraph of Figure 7 being large but the fluctuation gradually decreasing with the increase of the time point, this is the result of the priority selection principle. The phenomenon of early damage appears at Mark 2, corresponding to the phenomenon of residence early reduction in Figure 6d.

Compared with the low-order HSMM results [3], the residual life prognosis by the proposed model in this paper is shown in Table 8. From the sampling time point, the overall effect of the residual life prognosis method based on HOHSMM and polynomial fitting is significantly better than that of the conventional HSMM. The case shows that the life prognosis method based on HOHSMM and polynomial regression is effective and feasible.

Table 8. Relative error analysis of the predicted RU.

Actual RUL	Model of This Paper		HSMM	
	Predicted RUL	The Relative Error (%)	Predicted RUL	The Relative Error (%)
300	357.010	19.00	302.558	0.85
260	266.729	2.59	299.643	15.25
220	180.068	18.15	297.954	35.43
170	142.383	16.25	194.981	14.69
150	120.147	19.90	192.081	28.05
120	98.159	18.20	188.666	57.22
110	102.573	6.75	102.471	6.84
90	79.956	11.16	100.291	11.43
50	39.600	20.80	97.675	95.35
Mean relative error/%		14.76		29.4592

6. Conclusions

The diagnostic condition of mechanical systems is that the training data are sufficient and the samples of different categories in the training data have a balanced distribution. In the whole life cycle of a high-voltage circuit breaker system, it is in normal state most of the time, and the number of normal samples in the actually monitored signal will be more than the number of fault samples, resulting in the imbalance of training data and unknown distribution. Thus, for the problem of large error in the residual lifetime prognosis of power systems, a novel residual lifetime prognosis model based on HOHSMM and polynomial fitting was proposed. Based on HSMM, an order reduction method and composite node mechanism of HOHSMM based on permutation were proposed. The order reduction method of the permutation and combination model is simple and intuitive and uses the definition of the high-order hidden Markov group model. The high-order model can be transformed into the corresponding first-order model by changing the observation angle, and the solution of the three problems of the low-order model can be used in the high-order complex model. The intelligent optimization algorithm group can be used to replace the EM algorithm to estimate the parameters and optimize the structure of the proposed model, and the simplification of the topology of the high-order model by the intelligent optimization algorithm can be realized. The complex dependency information in the high-order model is transferred to the deformed parameter group. It effectively simplifies the model and provides a new idea for the study of this kind of model. Finally, a case was studied to verify the proposed model. From the experimental results, the comparison between the proposed model and HSMM showed several advantages of the proposed model, indicating that the remaining life prediction based on polynomial fitting has better performance for the health prognosis problem of the power system.

Author Contributions: Conceptualization, Q.L., D.L. and W.L.; methodology, T.X.; investigation, Q.L. and D.L.; resources, J.L.; writing—original draft preparation, Q.L.; writing—review and editing, D.L.; visualization, W.L.; supervision, T.X.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (Nos. 71632008, 71840003, and 51875359), the Natural Science Foundation of Shanghai (No. 19ZR1435600 and 20ZR1428600), the Humanity and Social Science Planning Foundation of the Ministry of Education of China (No. 20YJAZH068), and the Science and Technology Development Project of the University of Shanghai for Technology and Science (No. 2020KJFZ038).

Acknowledgments: The authors are indebted to the reviewers and the editors for their constructive comments which greatly improved the contents and exposition of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Duan, C.; Makis, V.; Deng, C. Optimal Bayesian early fault detection for CNC equipment using hidden semi-Markov process. *Mech. Syst. Signal Process.* **2019**, *122*, 290–306. [[CrossRef](#)]
2. Carey, B.; Dan, M.; Tarik, A. Condition-based maintenance of machines using Hidden Markov models. *Mech. Syst. Signal Process.* **2000**, *14*, 597–612.
3. Liu, Q.; Dong, M.; Lv, W.; Geng, X.; Li, Y. A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis. *Mech. Syst. Signal Process.* **2015**, *64–65*, 217–232. [[CrossRef](#)]
4. Huang, W.; Dietrich, D. An Alternative Degradation Reliability Modeling Approach Using Maximum Likelihood Estimation. *IEEE Trans. Reliab.* **2005**, *54*, 310–317. [[CrossRef](#)]
5. Liu, Q.; Dong, M.; Peng, Y. A novel method for online health prognosis of equipment based on hidden semi-Markov model using sequential Monte Carlo methods. *Mech. Syst. Signal Process.* **2012**, *32*, 331–348. [[CrossRef](#)]
6. Wu, B.; Li, W.; Qiu, M. Remaining useful life prediction of bearing with vibration signals based on a novel indicator. *Shock. Vib. Shock. Vib.* **2017**, *90*, 1–10. [[CrossRef](#)]
7. Kacprzynski, G.J.; Sarlashkar, A.; Roemer, M.J. Predicting remaining life by fusing the physics of failure modeling with diagnostics. *J. Miner. Met. Mater. Soc.* **2004**, *56*, 29–35. [[CrossRef](#)]
8. Guha, A.; Patra, A.; Vaisakh, K.V. Remaining useful life estimation of lithiumion batteries based on the internal resistance growth model. In Proceedings of the 2017 Indian Control Conference (ICC), Guwahati, India, 4–6 January 2017; pp. 33–38.
9. Yang, T.; Zheng, Z.; Qi, L. A method for degradation prediction based on Hidden semi-Markov models with mixture of Kernels. *Comput. Ind.* **2020**, *122*, 103295. [[CrossRef](#)]
10. Yang, Z.; Baraldi, P.; Zio, E. A multi-branch deep neural network model for failure prognostics based on multimodal data. *J. Manuf. Syst.* **2021**, *59*, 42–50. [[CrossRef](#)]
11. Behera, S.; Misra, R.; Sillitti, A. Multiscale deep bidirectional gated recurrent neural networks based prognostic method for complex non-linear degradation systems. *Inf. Sci.* **2021**, *554*, 120–144. [[CrossRef](#)]
12. Li, J.; Ping, Y.; Li, H.; Li, H.; Liu, Y.; Liu, B.; Wang, Y. Prognostic prediction of carcinoma by a differential-regulatory-network-embedded deep neural network. *Comput. Biol. Chem.* **2020**, *88*, 107317. [[CrossRef](#)] [[PubMed](#)]
13. Lv, X.; Wang, H.; Zhang, X.; Liu, Y.; Jiang, D.; Wei, B. An evolutionary SVM method based on incremental algorithm and simulated indicator diagrams for fault diagnosis in sucker rod pumping systems. *J. Pet. Sci. Eng.* **2021**, *203*, 108806. [[CrossRef](#)]
14. Zeng, N.; Qiu, H.; Wang, Z.; Liu, W.; Zhang, H.; Li, Y. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer’s disease. *Neurocomputing* **2018**, *320*, 195–202. [[CrossRef](#)]
15. Saha, S.; Saha, B.; Saxena, A.; Goebel, K. Distributed prognostic health management with Gaussian process regression. In Proceedings of the 2010 IEEE Aerospace Conference, Big Sky, MT, USA, 6–13 March 2010; pp. 1–8.
16. Dong, M.; He, D. Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *Eur. J. Oper. Res.* **2007**, *178*, 858–878. [[CrossRef](#)]
17. Yan, Y.; Cai, J.; Li, T.; Zhang, W.; Sun, L. Fault prognosis of HVAC air handling unit and its components using hidden-semi Markov model and statistical process control. *Energy Build.* **2021**, *240*, 110875. [[CrossRef](#)]
18. Li, J.; Zhang, X.; Zhou, X.; Lu, L. Reliability assessment of wind turbine bearing based on the degradation-Hidden-Markov model. *Renew. Energy* **2018**, *132*, 1076–1087. [[CrossRef](#)]
19. Huang, D.; Ke, L.; Chu, X.; Zhao, L.; Mi, B. Fault diagnosis for the motor drive system of urban transit based on improved Hidden Markov Model. *Microelectron. Reliab.* **2018**, *82*, 179–189.
20. Du Preeze, J.A. Efficient training of high-order hidden Markov models using first-order representations. *Comput. Speech Lang.* **1998**, *12*, 23–39. [[CrossRef](#)]
21. Hadar, U.; Messer, H. High-order hidden Markov models—estimation and implementation. In Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing (SSP 2009), Cardiff, UK, 31 August–3 September 2009; pp. 249–252.

22. Zhu, D.-M.; Lu, J.; Ching, W.-K.; Siu, T.-K. Discrete-time optimal asset allocation under Higher-Order Hidden Markov Model. *Econ. Model.* **2017**, *66*, 223–232. [[CrossRef](#)]
23. Heng, X.; Mamon, R. A self-updating model driven by a higher-order hidden Markov chain for temperature dynamics. *J. Comput. Sci.* **2016**, *17*, 47–61.
24. Vyas, R.; Kanumuri, T.; Sheoran, G.; Dubey, P. Efficient iris recognition through curvelet transform and polynomial fitting. *Optik* **2019**, *185*, 859–867. [[CrossRef](#)]
25. Dong, M.; He, D. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mech. Syst. Signal Process.* **2007**, *21*, 2248–2266. [[CrossRef](#)]