Article

An Automatic Partition Time-Varying Markov Model for Reliability Evaluation

Linlin Kou 1,2, Baiqing Chu 2, Yan Chen 2 and Yong Qin 1,*

1 State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China; koulinlin5829@bjsubway.com
2 Beijing Mass Transit Railway Operation Corp., Ltd., Beijing 100044, China; 18600295169@163.com (B.C.); 12120823@bjtu.edu.cn (Y.C.)
* Correspondence: yqin@bjtu.edu.cn

Abstract: As the service time of mechanical devices is getting longer and longer, the safe and reliability evaluation during operation is highlighted. Moreover, real-time reliability evaluation with consideration of multi-state performance degradation becomes increasingly important nowadays, since the consequences of sudden failures are more unacceptable than ever before. The Markov process is a commonly used model in multi-state reliability evaluation. However, little research of the Markov model can deal with multi-source monitoring data and time-varying properties of device performance degradation, as well as the scientific state number determination. In this article, a real-time reliability evaluation model based on automatic partition and the time-varying Markov chain is proposed to solve the problems of the scientific state number selection and time-varying properties description with the state transition matrix of the Markov process, together with taking advantage of multi-source information. The effectiveness of the proposed algorithm is demonstrated on the bearing with life-long vibration and temperature data. It shows that the proposed automatic partition time-varying Markov model can decide the state number automatically according to the trend of life-long data, and evaluate real-time reliability based on equipment operating hours and operating status. The result of predicted remaining useful life obtained by the proposed model is more accurate, and it also shows great superiority in conformity with reality.

Keywords: real-time reliability; time-varying; automatic partition; remaining useful life; Markov process

1. Introduction

With the continuous improvement of design and manufacturing technology, as well as the materials used, service time of mechanical devices is getting longer and longer, especially in the field of rail transit. For instance, the design service life of bearing on a bogie axle box is more than 20 years. These products usually have only a small number of failures or no failures in life tests, including accelerated life test methods. However, the traditional reliability evaluation methods such as fault tree analysis (FTA) [1], event tree analysis (ETA) [2], and failure mode effects and criticality analysis (FMECA) [3], etc. mainly focus on the “average” characteristics of the same kind of product with fault record data. There is an insurmountable gap of these methods in describing the performance degradation of specific components. Moreover, devices with the same type have different degradation trajectories under different operating environments [4]; even devices of the same type with the same operating environments may also have different degradation trajectories, caused by randomness in the manufacturing process or other unexpected factors. Since real-time reliability evaluation becomes increasingly important nowadays, performance monitoring data is generally used for reliability evaluation of the specific product, instead of traditional failure records. Thanks to the rapid development of sensor technology, more and more data for the service performance evaluation and reliability
analysis has been provided, which greatly improves the possibility of accurate reliability evaluation and prediction.

The concept of reliability analysis based on performance degradation data was first proposed by Gertsbackh and Kordonskiy in 1969 [5]. Since then, many experts and scholars have started exploratory research in this area. At present, it can be mainly divided into two kinds of performance degradation reliability analysis methods based on regression model and stochastic process. A simple regression model is used in the escalators Avante model based on time series analysis [6]. The effect of variable amplitude service loading conditions and wear phenomenon are investigated based on nonlinear regression reliability analysis for rail steel [7]. Reference [8] combined the fault data with the state detection data of performance degradation, and used the DTS (Degradation Threshold Shock) method to analyze the reliability with competitive failure phenomenon. The commonly used stochastic processes are mainly the Wiener processes [9,10], Gamma process [11], Poisson process [12], Markov process [13], etc. Some reliability analysis methods generally assume that the device only has two states (normal and failure) [14–16]. The potential changes during device degradation had been ignored, which made the reliability results too crude to guide practical application. Compared with two-state (normal, fault) reliability analysis models, the multi-state model is more suitable for actual equipment operation. Reference [17] provided a reliability evaluation framework based on multi-performance conversion, which is conducted on the conversion capabilities of the components. A multi-state physics model for dynamic reliability evaluation with the crack propagation state transitions and the Forman model is proposed for a multi-cracked structure [18]. As a powerful tool for multi-state performance degradation modeling, the Markov process of different types has obtained many achievements, such as discrete time Markov process (DTMP) [19,20], continuous time Markov process (CTMP) [21,22], discrete-time hidden aging Markov process (DTHAMP) [23,24], nonhomogeneous continuous-time hidden Markov process (NHCTHMP) [25,26], homogeneous discrete-time hidden semi-Markov process (HDTHSMP) [27–29], etc. However, the number of device states was set directly, which may not objectively reflect the overall trend of the device’s performance and may lead to a decrease in model accuracy. Very few pieces of literature have studied the actual number of the device’s states. Reference [30] achieved the optimal number of states by multiple trials under different states number with BIC and AIC as evaluation indicators. However, the time-consuming problem stands out with the increase of available data. Besides that, the state transition matrix of most Markov models is fixed, which means no matter how the device performance degraded, the device reliability does not change when it is detected in a certain state of such a Markov model. In fact, the longer the system or device stays in a certain state, the greater the probability of transitioning out of the state, so these mentioned Markov models cannot reflect the performance degradation within the same state interval over time, which is not in line with reality [31]. That is a great obstacle in dealing with time-varying characteristics of device performance. Moreover, most degradation reliability evaluation methods, especially for mechanical devices, are based on single monitoring data, such as vibration data. Multi-source monitoring is the trend of device performance condition monitoring, and more information is provided for accurate detection and prediction [32]. How to make full use of multi-source data is an urgent problem to be solved.

Hence, in this article, a real-time reliability evaluation model based on the automatic partition method and time-varying Markov chain is proposed, in which the automatic partition is utilized to select the state number of Markov processes scientifically, while the time-varying properties of device performance are described by the time-varying transition matrix of the Markov process. The rest of this article is structured as follows. The basic concept of Tsallis entropy working as data features is presented in Section 2, first. Then the framework of the proposed real-time reliability evaluation algorithm based on automatic partition time-varying Markov process is illustrated. An experiment is carried out in Section 3 to demonstrate the superiority of the proposed algorithm. In particular, it is
conducted on bearing with a condition monitoring vibration signal in two directions and a temperature signal. The article is summarized in Section 4.

2. Materials and Methods

2.1. Tsallis Entropy

Tsallis entropy is non-extensive entropy; it is an expression of entropy with exponent $q$ proposed by Brazilian physicist Constantino Tsallis in 1988 [33]. Its mechanism is based on the non-extensive statistics derived from the Boltzmann–Gibbs theory. The Tsallis entropy performs very well in the description of the system with long-range interactions, long-term memory effects, or systems evolving in a multifractal-like space-time. Due to the above characteristics, the Tsallis entropy has a wide range of applications in many fields, such as physics, medicine, and economics. The Tsallis entropy is concave, stable, and is limited in the rate of entropy generation per unit time [34]. Although Shannon entropy, Renyi entropy, etc., have very rich applications in the field of bearing analysis, these entropies are not good enough to describe the bearing performance degradation with vibration signal. Therefore, the Tsallis entropy is extracted as the data featured from the vibration signal in this article. The definition is as below:

$$H_q^T = \frac{1}{q-1}(1 - \sum_{i=1}^{n} p_i^q)$$  \hspace{1cm} (1)

where $q$ describes the non-extension degree of the target object. In this research, it represents the long-range correlation of time series data. $n$ is the total data number in one sample. $p_i$ is the $i$-th value of the sample.

The purpose of Tsallis entropy is to find a clear $q$ value (generally not equal to 1) to describe as much as possible the phenomenon that is neither regular nor completely chaotic or random. After many experiments, we found that $q = 2$ can better describe the changing characteristics of bearing.

2.2. The Automatic Partition Time-Varying Markov Model for Reliability Evaluation

The reliability analysis method based on the automatic partition time-varying Markov process mainly includes five steps: degradation feature extraction, state number decision, training data preparation, time-varying Markov process modeling, and reliability assessment. The automatic partition method is used to automatically divide the component degradation sample data into multiple stages, to avoid the artificial and time-consuming selection of the number of Markov process states. In the data processing procedure, in order to deal with the different degradation processes of bearings, the slope-based dynamic time warping method is used to effectively regularize the collected samples, so as to get more accurate bearing Markov model. On the basic Markov process, a time-varying transition probability matrix is proposed to describe the changing trend of the device’s own characteristics with time, which provides a necessary means for more accurate assessment of device state reliability and life prediction. The functions for reliability and remaining useful life are provided at the last step.

The basic steps are provided as follows and shown in Figure 1.

1. Extract feature on the full-life-running vibration data.
2. The automatic partition method is used to decide the state number of Markov model, corresponding to the degradation state of device.
3. Prepare training data. The slope-based dynamic time warping method is used for sample time warping to complete the sample data processing in this step.
4. Establish the time-varying Markov model. The model parameters are re-estimated using the conditional probability formula for calculating the initial forward probability, backward probability, and state sequence.
5. Reliability and remaining useful life calculation and prediction. Determine the current state and obtain the sojourn time in the current state, $i$. Then, calculate it in the
remaining states \( i + 1, i + 2, \ldots, N \). Calculate the reliability and remaining useful life of the equipment with current state, \( i \), and sojourn time, \( st_i \).

![Figure 1. Basic steps of the proposed method.](image)

### 2.2.1. The Automatic Partition Method for State Number Determination of the Markov Model

The automatic partition method based on fuzzy time series segmentation is used to automatically obtain the state number of the Markov model.

It is assumed that \( x_t \) is an \( M \) dimensional multivariate sample, and \( S \) is the sample sequence with length, \( L \), where \( S = \{ d_l = [l_1, x_{l1}] \mid 1 \leq l \leq L \}, x_l = \{ x_{lm} \mid m = 1, 2, \ldots, M \} \).

According to the trend of data change, that is based on the equipment performance degradation, sample sequence \( S \) can be partition into several segments.

In this article, sample sequence \( S \) is assumed to be divided into \( P \) segments, noted as \( Sub^p_S = \{ Sub^p_S | p = 1, 2, \ldots, P \} \), and the boundaries are \( s_1 < s_2 < \cdots < s_P \), where \( s_1, s_2, \ldots, s_P \in S \) and \( s_1 = d_1, s_P = d_L \). That is, \( Sub^p_S = (s_{p-1} + 1, s_p), p = 1, 2, \ldots, P \).

Define the sample sequence segmentation objective function as \( C(Sub^p_S) \).

\[
C(Sub^p_S) = \sum_{p=1}^{P} C(Sub^p_S) = \sum_{p=1}^{P} C(s_{p-1} + 1, s_p) \tag{2}
\]

The objective function is generally represented as the distances between the monitoring data in each segment and their corresponding fitting function.

\[
C(Sub^p_S) = \sum_{p=1}^{P} \sum_{l=s_{p-1}+1}^{s_p} \left\| x_l - \nu^p_{l} \right\|^2 = \sum_{p=1}^{P} \sum_{l=1}^{L} \beta_p(t_l) D^2(x_l, \nu^p_{l}) \tag{3}
\]

where \( \nu^p_{l} \) is the cluster center of the \( p \)-th segment, \( D^2(x_l, \nu^p_{l}) \) is the distance between \( x_l \) and \( \nu^p_{l} \), \( \beta_p(t_l) \) is the membership function of \( x_l \) subordinate to the \( p \)-th segment.

The objective function mentioned above is minimized to obtain the segment boundaries with optimal segment positions \( s_{p-1} + 1, s_p, p = 1, 2, \ldots, P \). It is generally conducted by various heuristic algorithms.

\( \beta_p(t_l) \) is defined as a \((0, 1)\) binary function in references [35,36], which is not appropriate in practice. In this article, the multivariate mixture Gaussian function is introduced into a cluster prototype as the sequence fitting function to describe the fuzzy characteristic of...
boundaries between segments. The optimal segment positions are obtained by minimizing the weighted sum of squared distances between the monitoring data in each segment and the center of the cluster prototype.

\[
C(\text{Sub}_L^p) = \sum_{p=1}^{P} \sum_{l=1}^{L} \left( \mu_{p,l} \right)^m D^2(d_l, \eta_p)
\]

(4)

where \( \mu_{p,l} \) is the membership of \( d_l = [t_l, x_l]^T \) belonging to the \( p \)-th segment, \( p = 1, 2, \ldots, P \); \( m = [1, \infty) \) is the weighted index of fuzzy clustering, generally taking \( m = 2 \) \([37]\); \( D^2(d_l, \eta_p) \) represents the distance between the monitoring data in each segment and the center of the cluster prototype; \( \eta_p \) is the clustering prototype function of \( p \)-th segment.

It is assumed that the expectation and the covariance matrix of the used multivariate Gaussian distribution for sample sequence are \( \eta_p \) and \( F_p \), respectively. The probability density function \( f(d_l|\eta_p) \) denoting the adscription of monitoring data to the \( p \)-th segment is

\[
f(d_l|\eta_p) = \sum_{p=1}^{P} f(d_l|\eta_p) f(\eta_p)
\]

(5)

where \( f(d_l|\eta_p) = \frac{1}{\sqrt{(2\pi)^{n} \frac{1}{2}} \sqrt{\det(F_p)}} \exp(-\frac{1}{2} (d_l - \eta_p)^T F_p^{-1} (d_l - \eta_p)) \); \( f(\eta_p) \) is the unconditional clustering probability, \( \sum_{p=1}^{P} f(\eta_p) = 1 \); \( \eta_p \) is the clustering prototype function of the \( p \)-th segment, \( \eta_p = \{ f(\eta_p), \eta_p, F_p | p = 1, 2, \ldots, P \} \).

According to probability theory, the distance function and the membership in the \( p \)-th segment are inversely proportional based on the Gath–Geva clustering algorithm \([38]\). The time and feature variables of sample sequence are independent with each other. Then,

\[
f(d_l|\eta_p) = \frac{1}{D^2(d_l, \eta_p)} \alpha_p \frac{1}{\sqrt{2\pi \sigma_{p,t}^2}} \exp(-\frac{1}{2} (t_l - \eta_{p,t})^2) \times \frac{1}{\sqrt{\det(A_p)}} \exp(-\frac{1}{2} (x_l - \eta_{p,v}^T)^T A_p^{-1} (x_l - \eta_{p,v}^T))
\]

(6)

where \( \alpha_k \) is the initial probability of clustering, \( \frac{1}{\sqrt{2\pi \sigma_{p,t}^2}} \exp(-\frac{1}{2} (t_l - \eta_{p,t})^2) \) is the distance between the time variable and the clustering center of time variable, \( \frac{1}{\sqrt{\det(A_p)}} \exp(-\frac{1}{2} (x_l - \eta_{p,v}^T)^T A_p^{-1} (x_l - \eta_{p,v}^T)) \) is the distance between the feature variable in the \( l \)-th sample and the feature cluster center, \( \eta_{p,v} \) is the feature cluster center of \( p \)-th segment, and \( r \) is the rank of the feature variable distance norm, \( A_p \).

There are many ways to calculate the distance norm \( A_p \). Mahalanobis distance is a good choice and can be used to adjust the correlation between variables. Let \( A_p = F_p \), where \( F_p \) is the fuzzy covariance matrix of the multivariate Gaussian distribution,

\[
F_p = \frac{1}{L} \sum_{l=1}^{L} (\mu_{p,l})^m (x_l - \eta_{p,v}^T)(x_l - \eta_{p,v}^T)^T
\]

(7)

Principal Component Analysis (PCA) is introduced to eliminate the strong correlations between variables while reducing dimensionality.
It is assumed that \( \lambda_{p,q}, q = 1, 2, \ldots, Q \) are the \( Q \) non-zero eigenvalues of \( F_p \) in descending order. Their corresponding eigenvectors are \( u_{p,q}, q = 1, 2, \ldots, Q \). It can be obtained that:

\[
F_p = U_p \Lambda_p U_p^T
\]  

(8)

\[
\Lambda_p = \begin{bmatrix}
\lambda_{p,1} & 0 & \cdots & 0 \\
0 & \lambda_{p,2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_{p,Q}
\end{bmatrix}
\]

(9)

\[
U_p = [u_{p,1}, u_{p,2}, \ldots, u_{p,Q}]
\]

(10)

The feature variables of the sample sequence are reduced to \( Q \)-dimension by the PCA method, and \( y_{p,l} = W_p^{-1}(x_l) = W_p^T(x_l) \) is obtained, where \( W_p = U_p \Lambda_p^{1/2} \).

According to probabilistic principal component analysis [39],

\[
W_p = U_p Q (\Lambda_p Q - \sigma^2_{p,x} I)^{1/2} R_p S, \text{ where } R_p \text{ is a } Q \times Q \text{ orthogonal transformation matrix,}
\]

\[
\sigma^2_{p,x} = \frac{1}{M-Q} \sum_{i=Q+1}^{M} \lambda_{p,q}.
\]

Until now, the automatic partitioning algorithm of fuzzy segmentation has been transformed into an operational research optimization problem. The model is shown below,

\[
C(Sub_p^S) = \sum_{p=1}^{P} \sum_{l=1}^{L} \left( \mu_{p,l} \right)^{m} D^2(d_l, \eta_p)
\]

(11)

\[
U = [\mu_{p,l}]_{P \times L}, \quad \mu_{p,l} \in [0, 1], \forall p, l
\]

\[
\sum_{p=1}^{P} \mu_{p,l} = 1, \quad \forall l
\]

\[
0 < \sum_{l=1}^{L} \mu_{p,l} < L, \quad \forall p
\]

(12)

The alternate optimization algorithm and the Picard iterative algorithm can be used to solve this problem.

The basic steps are as follows:

1. Initialization

   The initial segments number of \( S \) and the dimension of the eigenvector space are given at first. Then, select the appropriate stopping condition, \( \epsilon, \epsilon' > 0 \), and initialize \( W_p, v_p^x, \sigma^2_{p,x}, \mu_{p,l}; \)

2. Loop calculation.

   (i) Calculate the cluster prototype \( \eta_p \).

      Cluster initial probability:

      \[
      \alpha_p = \frac{1}{T} \sum_{l=1}^{L} \mu_{p,l}
      \]

      (13)

      Cluster center:

      \[
      v_p^x = \frac{\sum_{l=1}^{L} \left( \mu_{p,l} \right)^{m} (x_l - W_p \langle y_{p,l} \rangle)}{\sum_{l=1}^{L} \left( \mu_{p,l} \right)^{m}}
      \]

      (14)

      where \( \langle y_{p,l} \rangle = M_p^{-1} W_p^T (x_l - v_p^x) \), \( M_p = \sigma^2_{p,x} I + W_p^T W_p \).

      Weight update:

      \[
      \tilde{W}_p = F_p W_p (\sigma^2_{p,x} I + M_p^{-1} W_p^T F_p W_p)^{-1}
      \]

      (15)
where $U_{i}$ is the similarity of each state space as measured by the fuzzy decision algorithm proposed by Kaymak et al. [40], in which the overall similarity matrix is $H_{a,b} = \{ h_{a,b} | 1 \leq a \leq P, 1 \leq b \leq P \}$,

$$h_{a,b} = \sqrt{ \left( \frac{1}{p(p-1)} \sum_{a=1}^{p} \sum_{b=1}^{p} S_{a,b}^{PCDA} \right)^{2} + \frac{1}{p(p-1)} \sum_{a=1}^{p} \sum_{b=1}^{p} D(v_{a}^{p}, v_{b}^{p}) }$$  \hspace{1cm} (21)

If the clustering similarity factor, $h_{a,b+1}$, of two adjacent segments, $Sub_{S_{a}}^{b}, Sub_{S_{a}}^{b+1}$, is exceeded the set threshold, $Sub_{S_{a}}^{b}, Sub_{S_{a}}^{b+1}$ will be merged. It ends when the algorithm termination condition, $\varepsilon$, is reached.

\subsection*{2.2.2. The Time-Varying Markov Model}

The automatic segmentation method provides a scientific basis for the states number selection of the Markov process. Moreover, we argue that the transition probability between states is not only related to the state that the device stays, but also related to the sojourn time that the equipment stays in the present state. In practical situations, the longer a device stays in a state, the greater the probability of its transition from that state to another states. That is to say, the state sojourn time affects the state transition probability to a certain extent. Therefore, the state transition matrix (Sojourn time-based state transition probabilities) related to the state sojourn time is introduced to establish a time-varying Markov process for real-time reliability evaluation. The proposed Markov process is noted

\begin{align*}
\sigma_{p,x}^{2} &= \frac{1}{Q} tr(F_{p} - F_{p}W_{p}M_{p}^{-1}\tilde{W}_{p}^{T}) \\
D_{p} &= \sigma_{p,x}^{2} I + \tilde{W}_{p}\tilde{W}_{p}^{T} \\
\end{align*}

\begin{align*}
\psi_{p} &= \frac{1}{c_{p}} \sum_{i=1}^{c_{p}} (\mu_{p,i})^{m} t_{i} \\
\sigma_{p,t}^{2} &= \frac{1}{c_{p}} \sum_{i=1}^{c_{p}} (\mu_{p,i})^{m} (t_{i} - \psi_{p})^{2} \\
\end{align*}

\begin{align*}
A_{p} &= \sigma_{p,x}^{2} I + \tilde{W}_{p}\tilde{W}_{p}^{T} \\
\end{align*}

\begin{align*}
\psi_{p} &= \frac{1}{c_{p}} \sum_{i=1}^{c_{p}} (\mu_{p,i})^{m} t_{i} \\
\sigma_{p,t}^{2} &= \frac{1}{c_{p}} \sum_{i=1}^{c_{p}} (\mu_{p,i})^{m} (t_{i} - \psi_{p})^{2} \\
\end{align*}
as \( M = (\pi, A, ST) \), where, \( \pi \) is the initial state probability distribution \( \pi = \{ \pi_i \}, 1 \leq i \leq N \) and \( \pi_i = P(s_1 = i), 1 \leq i \leq N \); \( N \) is the state number, which is determined by the automatic partition method mentioned above; \( A \) is the time-varying state transition matrix, \( A = \{ a_{ij}(st^t) \}, 1 \leq i, j \leq N, 1 \leq st^t \leq ST^t \), and \( a_{ij}(st^t) \) represents the state transition probability of the device transitioning from state \( i \) to state \( j \) when the device stays in state \( i \) at time \( t \), \( \sum_{j=1}^{N} a_{ij}(st^t) = 1, 1 \leq st^t \leq ST^t, 1 \leq i, j \leq N \). \( ST^t \) is the maximum sojourn time in state \( i \), and \( ST \) is the state sojourn time distribution, which is assumed to obey a Gaussian distribution.

Firstly, according to the sample set, the sojourn time distribution of each state is determined, and then the model parameters are estimated by the improved Baum–Welch algorithm.

1) When the model parameters \( M = (\pi, A, ST) \) are known, the probability that the device stays in state \( i \) for a sojourn time, \( T^* \), at time \( t \) is \( a_t(i, T_t^{it*}) = P(q_t = i, st^t = T_t^{it*} | M) \), When \( t = 1 \), the initial forward probability is

\[
a_1(1, st^1_1) = 1
\]

\[
a_1(j, 1) = a_1(1, st^1_1)a_{ij}(st^1_t)
\]

\[
a_1(N, st^N_1) = \sum_{i=2}^{N} a_1(i, st^1_t = 1)P(st^1_t = 1)a_{iN}(st^1_t = 1)
\]

2) At time \( t \), when the equipment stays in state \( i \) for a sojourn time \( st^t_j \), the backward probability of producing the state sequence in time \( T - t \) is \( \beta_t(i, st^t_j) = P(s_t = i, st^t_j | M) \). When \( t = T \), the initial backward probability formula is

\[
\beta_T(N, st^N_T) = 1
\]

\[
\beta_T(i, st^t_j) = P(st^t_j | a_{iN}(st^t_j))
\]

\[
\beta_T(1, st^t_1) = \sum_{j=2}^{N} a_{1j}(st^t_1)\beta_T(j, st^t_1)
\]

When \( 0 < t < T \), the backward probability recursion formula is

\[
\beta_t(i, st^t_j) = P(st^t_j) \sum_{j=2}^{N} a_{ij}(st^t_j)\beta_{t+1}(j, 1) + \beta_{t+1}(i, st^t_{j+1}), 1 < i < N
\]

\[
\beta_t(1, st^t_1) = \sum_{j=2}^{N} a_{1j}(st^t_1)\beta_t(j, 1)
\]
Based on the forward probability formula and the backward probability formula, the conditional probability formula \( S = (s_1, s_2, \ldots, s_T), s_1, s_2, \ldots, s_T \in \{1, 1 \leq i \leq N\} \) of the state vector can be obtained as

\[
P_S = P(S|M) = \sum_{i=1}^{N} \sum_{s_t^i} a_t(i, s_t^i) \beta_t(i, s_t^i)
\]

(33)

(3) The re-estimated parameters of the time-varying Markov model can be obtained by the formula

\[
\xi_t(i, j, s_t) = p(s_t = i, s_{t'} = j, s_t^i = s_t|S, M)
\]

\[
= \left\{ \begin{array}{ll}
\frac{1}{P_S} a_t(1, s_t^i) a_t(1, j) \beta_t(i, 1), & i = 1 \\
\frac{1}{P_S} a_t(1, s_t^i) p(s_t) a_t(s_t^i) \beta_{t+1}(j, 1), & 2 \leq i, j < N \\
\frac{1}{P_S} a_t(i, s_t) p(s_t) a_tN(s_t) \beta_{t+1}(j, s_t), & 2 \leq i < N, j = N
\end{array} \right.
\]

(34)

\( \zeta_t(i, j, s_t) \) is the probability with a given model, \( M \), and state sequence of \( S = (s_1, s_2, \ldots, s_T) \) and \( s_1, s_2, \ldots, s_T \in \{1, 1 \leq i \leq N\} \) that show the performance state of equipment will transit to state \( j \) after a sojourn time \( s_t \) in state \( i \).

\( \gamma_t(i, s_t) \) is the probability that when the device stays in state \( i \) for a sojourn time, \( s_t \), at time \( t \), with a given model of \( M \) and state sequence of \( S, \gamma_t(i, s_t) = p(s_{t+1} = j|s_t = i, s_t^i = s_t|S, M) = \frac{1}{P_S} a_t(i, s_t) \beta_t(i, s_t) \).

\( \eta_t(i, s_t) \) is the probability when the device stays in state \( i \) for a sojourn time of \( s_t \) at time of \( t \),

\[
\eta_t(i, s_t) = \frac{1}{P_S} a_t(i, s_t) p(s_t) = s_t|S, M) \left[ \sum_{j=2}^{N-1} a_{ij}(s_t) \beta_{i+1}(j, 1) + a_{iN}(s_t) \beta_{i}(N, s_t^N) \right]
\]

(35)

It can be seen from \( a_{ij}(s_t) = p(s_{t+1} = j|s_t = i, s_t^i = s_t) \) that the revaluation formula of \( a_{ij}(s_t) \) is

\[
\pi_{ij}(s_t) = \frac{\sum_{t=1}^{T} \xi_t(i, j, s_t)}{\sum_{t=1}^{T} \eta_t(i, s_t)}
\]

(36)

where \( \sum_{t=1}^{T} \xi_t(i, j, s_t) \) is noted as the expected transition number from state \( i \) to state \( j \), when it stays in state \( i \) for a sojourn time, \( s_t \). \( \sum_{t=1}^{T} \eta_t(i, s_t) \) is the sum of expected transition number out of state \( i \), when it stays in state \( i \) for a sojourn time, \( s_t \).

Assume that the state sojourn time probability obeys a Gaussian distribution \((\mu_i, \sigma_i^2)\),

\[
\overline{\mu}_i = \frac{\sum_{t=1}^{T} \sum_{s_t=1}^{T_t^s} \eta_t(i, s_t) \cdot s_t}{\sum_{t=1}^{T} \sum_{s_t=1}^{T_t^s} \eta_t(i, s_t)}
\]

(37)

\[
\overline{\sigma}_i^2 = \frac{\sum_{t=1}^{T} \sum_{s_t=1}^{T_t^s} \eta_t(i, s_t) \cdot s_t^2}{\sum_{t=1}^{T} \sum_{s_t=1}^{T_t^s} \eta_t(i, s_t)} - \overline{\mu}_i^2
\]

(38)

(4) Reliability and remaining useful life prediction
The device in the completely failed segment means that it cannot complete the specified task as required. Therefore, it can be considered that the probability $\zeta_{t}(i, N, st)$ of the device transferring to the completely failed segment, $N$, at time, $t$, with a sojourn time, $st$, is its unreliability. The sum probabilities for device from the current state transferring to every other states is its reliability. The reliability of a device at different times can be defined as

$$R(t) = 1 - \zeta_{t}(i, N, st)$$  \hspace{1cm} (39)

The failure rate function is noted as $\lambda(t)$, and its life distribution function is as $F(t)$; density function is $f(t), f(t) = F'(t)$, the reliability function is $R(t), F(t) + R(t) = 1$, and $\lambda(t) = f(t) / R(t)$. For $\forall t$, there is $R(t) > 0$. Therefore, it can be considered that $\lambda(t)$ is an estimate of the interval fault conditional probability.

Let $L$ denote the life cycle of the device and $RUL(t)$ denote the remaining life expectancy of the equipment running normally at time, $t$, from time, $t$, to the failure time; there is $RUL(t) = E(L - t | L > t) = (1 / R(t)) \int_{t}^{\infty} R(x)dx$.

At time $t^*$, if the device stays in state $i$ with a sojourn time of $st$, its remaining useful life is equal to the sum of the effective remaining sojourn time of running state $i$ and the remaining time in other remaining states (assuming that it traverses $i + 1, i + 2, \ldots, N - 1$ states).

Among them, the effective remaining sojourn time in operating state $i$ is

$$rul_{t^*+st}^i = rul^i \cdot (1 - \lambda(t^* + st) \cdot \Delta t) = rul^i \cdot (1 - \frac{\zeta_{t^*+st}(i, j, st)}{\eta_{t^*+st}(i, st)})$$  \hspace{1cm} (40)

where $rul_{t^*+st}^i$ is the effective remaining sojourn time when the device stays in state $i$ with a sojourn time of $st$ at time $t^*$; $rul^i$ is the expected sojourn time in state $i$, $rul^i = \mu_i$.

Therefore, the remaining useful life for device at time $t^*$, when it stays in state $i$ with a sojourn time of $st$ can be obtained,

$$RUL_{t^*+st} = rul_{t^*+st}^i + \sum_{j=i+1}^{N-1} \mu_j$$  \hspace{1cm} (41)

3. Results

Rolling bearing is one of the most widely used general mechanical components in rail vehicles [35]. Only 10~20% of bearings can reach their design life time, according to statistics. They usually cannot work properly due to fatigue, wear, strain, electrical corrosion, fracture, and gluing, etc. Therefore, the bearing is taken as an example to verify the science of the proposed method in this article.

3.1. Experiment Description

The full life data of bearings collected on the PRONOSTIA experimental platform (Figure 2), specially designed and implemented by the AS2M department of the French FEMTO-ST Institute, are used. The PRONOSTIA experimental platform consists of three main parts: the rotating part, the loading part (where radial force is applied on the test bearing), and the measuring part.
where *ts trul + is the effective remaining sojourn time when the device stays in state *i with a sojourn time of *st at time *t ∗; irul is the expected sojourn time in state *i, \[ \mu = . \]

Therefore, the remaining useful life for device at time *t ∗, when it stays in state *i with a sojourn time of *s ∗ can be obtained, **1

3. Results

Rolling bearing is one of the most widely used general mechanical components in rail vehicles [35]. Only 10~20% of bearings can reach their design life time, according to statistics. They usually cannot work properly due to fatigue, wear, strain, electrical corrosion, fracture, and gluing, etc. Therefore, the bearing is taken as an example to verify the science of the proposed method in this article.

3.1. Experiment Description

The full life data of bearings collected on the PRONOSTIA experimental platform (Figure 2), specially designed and implemented by the AS2M department of the French FEMTO-ST Institute, are used. The PRONOSTIA experimental platform consists of three main parts: the rotating part, the loading part (where radial force is applied on the test bearing), and the measuring part.

![Figure 2. PRONOSTIA test platform of bearing.](image)

The operating conditions of the bearing are measured by the radial force exerted on the bearing, the rotational speed of the bearing, and the torque on the bearing, with a sampling frequency of 100 Hz. The amount of degradation of the bearing is measured by vibration and temperature sensors, and the vibration signal is measured by two placements at a 90-degree angle on the horizontal and vertical axes. The sampling frequency is 25.6 kHz, sampling every 10 s, and each sampling duration of 0.1 s, including 2560 sample points. The temperature sensor is located in the hole near the outer ring of the bearing, with a sampling frequency of 10 Hz and 600 points per minute.

The test dataset used for training in this article is shown in Table 1.

### Table 1. Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Operating Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning set</td>
<td>Bearing1_1 Bearing1_2 Bearing1_4 Bearing1_5 Bearing2_1 Bearing2_4 Bearing3_1</td>
</tr>
<tr>
<td>Test set</td>
<td>Bearing1_6 Bearing1_7 Bearing2_5 Bearing3_3</td>
</tr>
</tbody>
</table>

3.2. Feature Extraction

Calculate the entropy of the bearing vibration acceleration data and align the temperature data with the vibration acceleration data. The Tsallis entropy with \( q = 2 \), kurtosis, Shannon entropy, and Renyi entropy were extracted from the same full-life of the bearing vibration signal.

As shown in Figure 3, it can be seen that Shannon entropy and Renyi entropy are not monotonic, and they also fluctuate greatly during the full-life signal period. The trend of the kurtosis value is not obvious enough, and there are some values larger than the failure value, where the sample number ranges from 450 to 900. Those cannot accurately express the performance of the bearing degradation. The Tsallis entropy exhibits monotonic characteristics during the whole life cycle, but also begins to degrade when the bearing performance is not very well (sample number = 1500 in Figure 3) and the numerical stability of the Tsallis entropy is good with bearing degradation.
### Table 1. Dataset Operating Conditions

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning set</td>
<td>Bearing1_1</td>
<td>Bearing2_1</td>
</tr>
<tr>
<td></td>
<td>Bearing1_2</td>
<td>Bearing2_4</td>
</tr>
<tr>
<td></td>
<td>Bearing1_4</td>
<td>Bearing1_5</td>
</tr>
<tr>
<td>Test set</td>
<td>Bearing1_6</td>
<td>Bearing3_3</td>
</tr>
<tr>
<td></td>
<td>Bearing1_7</td>
<td></td>
</tr>
</tbody>
</table>

### 3.2. Feature Extraction

Calculate the entropy of the bearing vibration acceleration data and align the temperature data with the vibration acceleration data. The Tsallis entropy with $q = 2$, kurtosis, Shannon entropy, and Renyi entropy were extracted from the same full-life of the bearing vibration signal.

As shown in Figure 3, it can be seen that Shannon entropy and Renyi entropy are not monotonic, and they also fluctuate greatly during the full-life signal period. The trend of the kurtosis value is not obvious enough, and there are some values larger than the failure value, where the sample number ranges from 450 to 900. Those cannot accurately express the performance of the bearing degradation. The Tsallis entropy exhibits monotonic characteristics during the whole life cycle, but also begins to degrade when the bearing performance is not very well (sample number = 1500 in Figure 3) and the numerical stability of the Tsallis entropy is good with bearing degradation.

![Figure 3. Feature comparison.](image)

The bearing temperature signal is resampled to make its length the same as the vibration signal; the maximum value is extracted, and the three groups of signals are normalized respectively to transform it into a dimensionless expression. Taking Bearing1_1 as an example, its characteristic expression is shown in Figure 4.

![Figure 4. Normalized features of bearing in condition 1.](image)

### 3.3. The Automatic State Number Determination of the Markov Model

The automatic partition method is performed on the feature data of condition 1-bearing 1 in the dataset. The input feature is [Tsallis 1, Tsallis 2, Temperature], and the result is shown in Figure 5.

![Figure 5. Result of data segmentation.](image)
3.3. The Automatic State Number Determination of the Markov Model

The automatic partition method is performed on the feature data of condition 1-bearing 1 in the dataset. The input feature is \([\text{Tsallis 1}, \text{Tsallis 2}, \text{Temperature}]\), and the result is shown in Figure 5.

Figure 5. Result of data segmentation.

Therefore, the segmentation results of Bearing1_1 are \([0, 360], [361, 1800], [1801, 2740], [2741, 2803]\). That is, the degradation process of the bearing can be divided into four stages, which can be considered as the running-in stage, the normal operation stage, the degradation failure stage, and the complete failure stage. It is consistent with the actual use, which fully verifies the validity and scientific rationality of the multivariate automatic partition algorithm.

The complete failure stage is regarded as the complete failure state; the running-in stage, the normal operation stage, and the degenerate failure stage are denoted as the state 1, the state 2, and the state 3 in this article.

Different bearings have different degradation trajectories. Samples are aligned based on the slope-based dynamic time warping (DTW) method to obtain more accurate training data, according to the change trend of the degradation amount, not the degradation amount. Taking the alignment process of Bearing1_1 and Bearing1_2 as an example, the optimal path for slope alignment, the slope alignment result and sample alignment results are shown in Figures 6–8, respectively.

Figure 6. Best path of slope alignment.

Figure 7. Result of slope alignment.
The complete failure stage is regarded as the complete failure state; the running-in stage, the normal operation stage, and the degenerate failure stage are denoted as the state 1, the state 2, and the state 3 in this article.

Different bearings have different degradation trajectories. Samples are aligned based on the slope-based dynamic time warping (DTW) method to obtain more accurate training data, according to the change trend of the degradation amount, not the degradation amount. Taking the alignment process of Bearing1_1 and Bearing1_2 as an example, the optimal path for slope alignment, the slope alignment result and sample alignment results are shown in Figures 6–8, respectively.

Figure 6. Best path of slope alignment.

Figure 7. Result of slope alignment.

Figure 8. Result of DTW on sample data.

The degradation trends of Bearing1_1 and Bearing1_2 are basically the same, as shown in Figure 8. Based on the segment location of Bearing1_1, the segment location of Bearing1_2 is ([1, 289], [290, 392], [393, 844], and [845, 871]).

However, not every bearing will go through these four degradation stages. In the tested samples, as shown in Figure 9, the Bearing2_4 fails directly without obvious degradation stage. The break-in stage of Bearing2_4 is [1, 402], the degradation failure stage is [403, 740], and the complete failure stage is [741, 751]. It can be concluded that the slope-based dynamic time warping method performed very well for samples that do not fully conform to the degenerate process. It avoids the negative impact caused by confusion of different degradation stages during Markov model training. The segmentation result of dataset is listed in Table 2.
The degradation trends of Bearing1_1 and Bearing1_2 are basically the same, as shown in Figure 8. Based on the segment location of Bearing1_1, the segment location of Bearing1_2 is ([1, 289], [290, 392], [393, 844], and [845, 871]). However, not every bearing will go through these four degradation stages. In the tested samples, as shown in Figure 9, the Bearing2_4 fails directly without obvious degradation stage. The break-in stage of Bearing2_4 is [1, 402], the degradation failure stage is [403, 740], and the complete failure stage is [741, 751]. It can be concluded that the slope-based dynamic time warping method performed very well for samples that do not fully conform to the degenerate process. It avoids the negative impact caused by confusion of different degradation stages during Markov model training. The segmentation result of dataset is listed in Table 2.

Table 2. Segmentations result of dataset.

<table>
<thead>
<tr>
<th>Bearing Number</th>
<th>State 1 Running-in Stage</th>
<th>State 2 Normal Operation Stage</th>
<th>State 3 Degradation Stage</th>
<th>State 4 Complete Failure Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bearing1_1</td>
<td>[1, 360]</td>
<td>[361, 1800]</td>
<td>[1801, 2740]</td>
<td>[2741, 2803]</td>
</tr>
<tr>
<td>Bearing1_2</td>
<td>[1, 289]</td>
<td>[290, 392]</td>
<td>[393, 844]</td>
<td>[845, 871]</td>
</tr>
<tr>
<td>Bearing1_4</td>
<td>[1, 329]</td>
<td>[330, 1084]</td>
<td>[1085, 1376]</td>
<td>[1377, 1428]</td>
</tr>
<tr>
<td>Bearing1_5</td>
<td>[1, 612]</td>
<td>[613, 1020]</td>
<td>[1021, 2421]</td>
<td>[2422, 2463]</td>
</tr>
<tr>
<td>Bearing1_7</td>
<td>[1, 540]</td>
<td>[541, 1980]</td>
<td>[1981, 2210]</td>
<td>[2211, 2259]</td>
</tr>
<tr>
<td>Bearing2_1</td>
<td>[1, 660]</td>
<td>[661, 817]</td>
<td>[818, 874]</td>
<td>[875, 977]</td>
</tr>
<tr>
<td>Bearing2_4</td>
<td>[1, 402]</td>
<td>-</td>
<td>[403, 740]</td>
<td>[741, 751]</td>
</tr>
<tr>
<td>Bearing2_5</td>
<td>[1, 566]</td>
<td>[567, 2254]</td>
<td>-</td>
<td>[2255, 2311]</td>
</tr>
<tr>
<td>Bearing3_1</td>
<td>[1, 246]</td>
<td>[247, 490]</td>
<td>-</td>
<td>[491, 517]</td>
</tr>
<tr>
<td>Bearing3_3</td>
<td>[1, 134]</td>
<td>[135, 294]</td>
<td>[295, 423]</td>
<td>[424, 436]</td>
</tr>
</tbody>
</table>

3.4. Reliability Evaluation

In order to verify the validity and accuracy of the state division of mechanical devices of rail transit trains proposed in this article, Bearing1_1, Bearing1_2, Bearing2_1, Bearing2_4, and Bearing3_1 datasets are used for training, and other bearing data are used for testing.

In the time-varying Markov process model, it is assumed that the state transition probability obeys the probability density function of mixed Gaussian distribution, and the probability distribution function of state sojourn time obeys the probability density function of a single Gaussian distribution. Based on the automatic partition method, the state number of the time-varying Markov model is 4, that is $N = 4$.

The maximum number of iteration steps is set to be 1000, and the stop criteria of the error convergence is $1 \times 10^{-5}$. The initial matrix of mutual conversion between each state is shown in Table 3. The average sojourn time in each state can be obtained based on training datasets; its result is shown in Table 4. A 4-state, time-varying Markov process reliability and life prediction model is obtained.
Table 3. Initial transformation matrix.

<table>
<thead>
<tr>
<th>States</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>0.8712</td>
<td>0.1112</td>
<td>0.0173</td>
<td>0.0003</td>
</tr>
<tr>
<td>State 2</td>
<td>0.0000</td>
<td>0.7496</td>
<td>0.2511</td>
<td>0.0993</td>
</tr>
<tr>
<td>State 3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.7028</td>
<td>0.2972</td>
</tr>
<tr>
<td>State 4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 4. Average sojourn time in each state.

<table>
<thead>
<tr>
<th>States</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average sojourn time</td>
<td>413.8</td>
<td>739.5</td>
<td>383.9</td>
<td>44.3</td>
</tr>
</tbody>
</table>

The trained time-varying Markov model is updated when a new monitoring data sample is arrived. The bearing staying in state 1 for 10 h is taken as an example. After feature extraction of the monitoring data sample, the data is sent into the trained time-varying Markov model, and state transition matrix is updated at this time. At this time, the state transition matrix is shown in Table 5, while the mean and variance of the state sojourn time are shown in Table 6.

Table 5. State transformation matrix when bearing stays in state 1 for 10 h.

<table>
<thead>
<tr>
<th>States</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 1</td>
<td>0.8695</td>
<td>0.1128</td>
<td>0.0174</td>
<td>0.0003</td>
</tr>
<tr>
<td>State 2</td>
<td>0.0000</td>
<td>0.7496</td>
<td>0.2511</td>
<td>0.0993</td>
</tr>
<tr>
<td>State 3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.7028</td>
<td>0.2972</td>
</tr>
<tr>
<td>State 4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 6. The mathematical expectation and variance of sojourn time when bearing stays in state 1 for 10 h.

<table>
<thead>
<tr>
<th>States</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectation of sojourn time</td>
<td>390.8451</td>
<td>739.5547</td>
<td>383.93594</td>
<td>44.3543</td>
</tr>
<tr>
<td>Variance of sojourn time</td>
<td>2.9458</td>
<td>1.5942</td>
<td>1.5026</td>
<td>1.4358</td>
</tr>
</tbody>
</table>

The bearing reliability at current time is \( R(i = 1, st_1 = 10) = 1 - 0.0003 = 0.9997 \).

To demonstrate the adaptability of the proposed model to changes in time, we also show the model changing parameters when the bearing stays in state 2 for 40 h. At this time, the state transition matrix is shown in Table 7, while the mean and variance of the state sojourn time are shown in Table 8.

Table 7. State transformation matrix when bearing stays in state 2 for 40 h.

<table>
<thead>
<tr>
<th>State Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.6512</td>
<td>0.3291</td>
<td>0.1179</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.7028</td>
<td>0.2972</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
without manual intervention, and the state number is obtained by taking full advantage of the hidden Markov model (HMM) algorithm, the prediction result of Bearing1_1 is presented in Figure 10. Because the state transition probability matrix is not changing over time in practice, it is only updated when a new monitoring data sample has arrived. The shorter the interval of condition monitoring data, the smoother the prediction curve would be. In addition, the proposed algorithm becomes more and more accurate as the bearing performance gradually degrades; that is, the results of predicted life are more accurate in the later stage. It should be emphasized that there is also a slight step-descent characteristic as the red line shown in Figure 10. Because the state transition probability matrix is not changing over time in practice, it is only updated when a new monitoring data sample has arrived. The shorter the interval of condition monitoring data, the smoother the prediction curve would be.

<table>
<thead>
<tr>
<th>States</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectation of sojourn time</td>
<td>-</td>
<td>665.9154</td>
<td>383.9326</td>
<td>44.3427</td>
</tr>
<tr>
<td>Variance of sojourn time</td>
<td>-</td>
<td>1.9425</td>
<td>1.5026</td>
<td>1.4358</td>
</tr>
</tbody>
</table>

The bearing reliability at current time is \( R(i = 2, st_1 = 40) = 1 - 0.1179 = 0.8821 \).

For the bearing used in this experiment, its effective life is the sum of sojourn time in state 1, 2, and 3. When the bearing reaches state 4, it has completely failed. Compared with the hidden Markov model (HMM) algorithm, the prediction result of Bearing1_1 is shown in Figure 10. The lifetime prediction of the HMM presents an obvious step-descent characteristic. Owing to the stable state transition matrix of HMM, the remaining useful life would not change when device performance is detected into a certain state. This characteristic makes the HMM seriously inconsistent with the actual situation.

![Figure 10. Result of remaining useful life prediction.](image)

In contrast, the predicted lifetime of the proposed algorithm in this paper has higher accuracy than HMM, which is in line with the gradual decline with time, since the state transition probability matrix has been designed to update over time theoretically. In addition, the proposed algorithm becomes more and more accurate as the bearing performance gradually degrades; that is, the results of predicted life are more accurate in the later stage. It should be emphasized that there is also a slight step-descent as the red line shown in Figure 10. Because the state transition probability matrix is not changing over time in practice, it is only updated when a new monitoring data sample has arrived. The shorter the interval of condition monitoring data, the smoother the prediction curve would be.

4. Conclusions

The proposed automatic partition time-varying Markov model has two main contributions on real time reliability evaluation, based on the Markov process. Firstly, the state number of Markov processes is selected scientifically on the basis of the device’s performance degradation stages. This work is realized automatically, based on the characteristics of the degradation curve, with full use of the multi-source monitoring data. It is done without manual intervention, and the state number is obtained by taking full advantage of multidimensional information, instead of by only a single threshold. This work builds a good foundation for a time-varying Markov model. The other main contribution is the time-varying Markov model. It describes the timely changing degradation process of device performance. The polymorphic real-time reliability assessment and remaining useful life prediction are achieved in this procedure. From the mathematic expectation and variance of sojourn time when device stays in state 2 for 40 h.

<table>
<thead>
<tr>
<th>States</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectation of sojourn time</td>
<td>-</td>
<td>665.9154</td>
<td>383.9326</td>
<td>44.3427</td>
</tr>
<tr>
<td>Variance of sojourn time</td>
<td>-</td>
<td>1.9425</td>
<td>1.5026</td>
<td>1.4358</td>
</tr>
</tbody>
</table>
useful life prediction are achieved in this procedure. From the experiment, the proposed algorithm demonstrates a strong capability for data changes and great potential in accurately performing degradation prediction. This model would be more accurate if the monitoring data points are provided more densely.

There are still some aspects left for improvement of this work. A good feature is the basis of the whole model. In this work, the feature was manually designed on the vibration signal. Its quality depends entirely on human experience, and this step is also time-consuming. In addition, the assessments procedure could be simplified in the future. Some migration algorithms may be considerable approaches to this problem.

Author Contributions: L.K. contributed to the conceptualization, methodology, formal analysis, and writing. B.C. designed and performed the research and approved the final manuscript. Y.C. analyzed the data, interpreted the results, and revised the manuscript. Y.Q. worked on conceptualization, formal analysis, and final writing review. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Beijing Mass Transit Railway Operation Corp. Ltd., grant number 2021000581000002.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

7. Nejad, R.M.; Liu, Z.; Ma, W.; Berto, F. Reliability analysis of fatigue crack growth for rail steel under variable amplitude service loading conditions and wear. *Int. J. Fatigue* 2021, 152, 106450. [CrossRef]