Article

The Prediction Method on the Early Failure of Hydropower Units Based on Gaussian Process Regression Driven by Monitoring Data

Huade Huang ¹ , Aisong Qin ¹ , Hanling Mao ^{1,*}, Jiahe Fu ², Zhenfeng Huang ¹, Yi Yang ², Xinxin Li ¹ and He Huang ¹

¹ School of Mechanical Engineering, Guangxi University, Nanning 530004, China; huadeh@163.com (H.H.); qinaisong66@163.com (A.Q.); zfhuan@gxu.edu.cn (Z.H.); lixinx@gxu.edu.cn (X.L.); hh1092038510@163.com (H.H.)

² Songcun Power Plant of Nanning Traffic Assets Management Co. Ltd., Nanning 530021, China; jiahe_fu@163.com (J.F.); yangyigxu@163.com (Y.Y.)

* Correspondence: maohl79@gxu.edu.cn

Abstract: The hydropower units have a complex structure, complicated and changing working conditions, complexity and a diversity of faults. Effectively evaluating the healthy operation status and accurately predicting the failure for the hydropower units using the real-time monitoring data is still a difficult problem. To this end, this paper proposes a prediction method for the early failure of hydropower units based on Gaussian process regression (GPR). Firstly, by studying the correlation between different monitoring data, nine state parameters closely related to the operation of hydropower units are mined from the massive data. Secondly, a health evaluation model is established based on GPR using the historical multi-dimensional monitoring information and fault-free monitoring data at the initial stage of unit operation. Finally, a condition monitoring directive based on the Mahalanobis distance (MD) is designed. The effectiveness of the proposed method is verified by several typical examples of monitoring data of a hydropower station in Guangxi, China. The results show that, in three cases, the abnormal conditions of the unit are found 2 days, 4 days and 43 days earlier than those of regular maintenances respectively. Therefore, the method can effectively track the change process of the operation state of hydropower units, and detect the abnormal operation state of hydropower units in advance.

Keywords: hydropower units; multi-dimensional monitoring information; correlation analysis; Mahalanobis distance; gaussian process regression model; state evaluation

Citation: Huang, H.; Qin, A.; Mao, H.; Fu, J.; Huang, Z.; Yang, Y.; Li, X.; Huang, H. The Prediction Method on the Early Failure of Hydropower Units Based on Gaussian Process Regression Driven by Monitoring Data. *Appl. Sci.* **2021**, *11*, 153. <https://dx.doi.org/10.3390/app11010153>

Received: 1 December 2020

Accepted: 22 December 2020

Published: 25 December 2020

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

Copyright: © 2020 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

Hydropower is the key renewable and clean energy in pursuit of economic decarbonization. The authors [1] reviewed a lot of innovative technologies in the field of hydropower, such as the digital implementation innovation technology that improves efficiency by 12%, and discuss the cost of standardization, best practices, and innovative technologies. The hydropower station is a key part of hydropower generation. Hydropower units are the core equipment of hydropower stations. Due to their complex internal structure and the influence of mechanical, electrical, and hydraulic factors during operation, hydropower units have a high failure rate and long maintenance time. The performance degradation of hydropower units has the property of gradual change [2]. At present, the state monitoring of hydropower units is mainly based on their vibration signal to extract the fault characteristics that can characterize the operation state. Methods such as EMD [3,4], VMD [5,6], and wavelet analysis [7,8] are commonly used to decompose the vibration signal to remove the high-frequency noise in the original signal, to extract the typical characteristic frequency corresponding to different faults of the units. A reliability centered maintenance

(RCM) concept for hydropower equipment is proposed, which uses failure mode and effect analysis (FMEA) technology to identify maintenance critical components [9–12]. Some authors [11] applied this method to the radial bearing fault detection of Kaplan turbine, to make maintenance plan before the failure of radial bearing, improve the availability of bearing and reduce the maintenance cost. However, the existing diagnosis technology has difficulty meeting the requirements of engineering applications due to the fewer faults with the label of the unit and the abnormal information contained in a signal that is not enough to fully evaluate the unit status [13,14]. The hydropower unit system has the characteristics of large-scale uncertainty, high nonlinearity, and strong correlation, so it is difficult to establish an accurate mechanism model. Many sensors are installed in hydropower units to collect many data closely related to the operation of the units. Based on the regression-based models of monitoring variables reflecting the state of the unit system, the abnormal monitoring of the unit can be realized by setting the threshold value [15]. Therefore, the monitoring system of the hydropower station is used to store many state parameters closely related to the operation status of hydropower units, and the realization of the hydropower unit state evaluation method has become a research hotspot [16–18]. Based on the swing data, the authors [19] integrated the vibration signal preprocessing, feature selection and prediction methods into a multi-objective optimization framework, and proposed an aggregated empirical wavelet transform-Gram-Schmidt orthogonal-multi-salp-objective-swarm-algorithm-kernel extreme learning machine (AEWT-GSO-MOSSA-KELM) model for vibration tension prediction of hydropower, and the predicted MAE of this model is 0.65 and the calculation time is 95.716 s, which are better than EEMD based, PCA based, SVR based, SSA based and MOPSO based models in the same operating environment. Zhou et al. [14] used vibration information and swing information to build a health state model of hydropower units. Based on the comprehensive consideration of the bearing vibration, active power, and working head information of the units, a condition parameter degradation assessment, and prediction model was proposed to evaluate and forecast hydropower units [13,20,21]. Based on the least square support vector machine model, the Y-direction vibration data of the bearing and the monitoring data of the draft tube were predicted to realize the diagnosis of potential faults [22].

However, in the above research, the selection of state parameters often depends on human experience, and the correlation between state parameters is not fully considered. Additionally, since the existing monitoring system may not be perfect, it is easy to form an information island, leading to a lot of redundancy between monitoring data [23]. In other fields, many researchers have applied support vector machine (SVM), deep learning, and other machine learning methods to similar research and achieved many results [24–26]. However, none of the models perform well in all cases, e.g., slow training speed for a large sample set, sensitivity to noise, and difficult to determine parameters. With the increasing number of layers and nodes of the neural network model, the larger the data set needs to be trained, the higher the complexity of the model and the training process is difficult to understand. Compared with support vector machine and neural network, the model established by the Gaussian process has fewer parameters that can adaptively deduce super parameters and realize noise separation according to data characteristics, which is very suitable for data processing with noise. In order to address the shortcomings mentioned above, this paper proposes a method of health assessment of hydropower units based on Gaussian process regression (GPR). GPR, as a new machine learning method based on Bayesian theory and statistical learning theory, has achieved good results in other fields of health monitoring [27–29]. Firstly, the paper uses the Pearson correlation coefficient (PCC), maximum information coefficient (MIC), and grey correlation degree (GCD) to select the appropriate operation state parameters from the massive data stored in the hydropower unit's system to construct the input eigenvector. Then the relevant multi-dimensional information is fully used to establish the health assessment of the hydropower units based on the GPR model using the historical data of no-fault records at the beginning of the unit operation. Finally, a condition monitoring directive (CMD) based on the MD is designed to

evaluate the operation state of the unit in real-time. In this paper, the basic process of the ideal framework of hydropower unit health assessment based on GPR is as follows. Firstly, in the off-line phase, for the problem of monitoring parameter selection, the correlation and redundancy between monitoring parameters are analyzed by using the historical monitoring data, and the parameter variables with strong correlation are selected to further remove the redundant variables. Then, based on the above selected monitoring parameters, the health assessment model based on GPR model is constructed by using the data without fault records during the initial operation of hydropower units, and the model is used as the benchmark model for online evaluation. Finally, in the on-line stage, combined with the online monitoring data, an index for evaluating the operation status of hydropower units is designed based on MD, and the reasonable alarm threshold is determined, to realize the on-line condition monitoring of hydropower units, as shown in Figure 1.

Figure 1. The proposed health assessment framework for hydropower units.

This paper is organized as follows. Section 2 introduces the correlation analysis methods for mining the appropriate input variables by using the correlation among health parameters. Section 3 describes the process of GPR modeling and the method of designing state evaluation indicators. Section 4 presents results from case studies, and Section 5 summarizes the conclusions.

2. Input Variables Selection

Extracting the state parameters closely related to the operation of hydropower units from the monitoring system of hydropower units is the premise to realize the health assessment of hydropower units. The correlation analysis methods such as PCC, GCD, and MIC can solve this problem well. The detailed descriptions of these three correlation analysis methods are as follows.

The PCC is a commonly used correlation analysis method, which can characterize the linear degree of tightness between the two variables [30]. In this paper, from the perspective of linear correlation, the PCC is used to identify the strong linear correlation monitoring parameters with the operation status of hydropower units, and to remove some redundant monitoring parameters with similar functions.

2.1. The GCD

Under the actual conditions, the operation condition of hydropower units is complex, and the monitoring data have the characteristics of high noise and strong randomness, which lead to many nonlinear relationships among the monitoring parameters except a linear relationship. In this paper, the GCD is used to identify the nonlinear characteristics among these monitoring parameters. The GCD mainly measures the relationship between each factor and behavior output or index according to the similarity of each correlation factor curve, which is suitable for extracting the relationship between small samples and poor information state [31].

Assuming the time series of active power is $y = \{y_1, y_2, \dots, y_n\}$, and the time series of monitoring parameters $x_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$, where i represents the i th monitoring parameter. The new sequences are obtained using dimensionless mean processing as Equations (1) and (2). The GCD of y and x_i can be described by Equation (3), and the specific calculation process is shown in Reference [31].

$$x'_{ik} = \frac{x_{ik}}{\frac{1}{n} \sum_{a=1}^n x_{ia}} \quad (1)$$

$$y'_k = \frac{y_k}{\frac{1}{n} \sum_{a=1}^n y_a} \quad (2)$$

$$R(x'_i, y') = \frac{1}{n} \sum_{k=1}^n \zeta[x'_{ik}, y'_k] \quad (3)$$

In Equation (3), $k = 1, 2, \dots, n$, $R(x'_i, y')$ grey relational grade represents the level of correlation between the time series of active power and the time series of monitoring parameters.

2.2. The MIC

Although the GCD can find out the non-linear relationship between the monitoring parameters, some of them still have many non-linear correlations and cannot be expressed by specific formulas. Therefore, this paper uses MIC to further explore the correlation between them. The MIC is a correlation algorithm that does not need to make any assumptions about the data distribution to evaluate the function and statistical relationship between variables. Given enough samples, it can effectively capture a wide range of relationships without being limited to specific relationship types [32]. The calculation process is as follows.

- (1) The formula for calculating the maximum mutual information value of two random variables X and Y for any given partition is Equation (4).

$$I(D, X, Y) = \sum_{x \in X, y \in Y} p(X, Y) \log \left(\frac{p(X, Y)}{P(X)P(Y)} \right) \quad (4)$$

- (2) Normalizing the above formula to make its value in the interval $[0, 1]$.

$$M(D)_{x,y} = \frac{I(D, X, Y)}{\log(\min\{x, y\})} \quad (5)$$

- (3) In the ordered pair data set D with the known sample n , the MIC of the two variables x and y in this set D is defined as Equation (6).

$$MIC(D) = \max_{xy < B(n)} \{M(D)_{x,y}\} \quad (6)$$

In Equation (6), $xy < B(n)$ ($B(n) = n^a$), n is the data scale to limit the mesh size to divide the region, and the value of a constant a can be set according to experience or scale.

MIC can detect the correlation whether two variables are functionally related, hyperfunction related, or even not related to any functional relationship. However, some statisticians believe that MIC can also make the illusion of correlation between two uncorrelated variables [33,34]. Therefore, to avoid the introduction of irrelevant variables due to the defects of MIC, this paper combines CGD theory to mine the nonlinear relationship between variables.

3. Health Assessment Model of Hydropower Units Based on GPR

3.1. GPR Theory

Gaussian process (GP) is a machine learning method based on statistical learning theory and Bayesian theory, which has good adaptability to deal with complex regression problems such as high dimensionality, small sample sizes, and nonlinearity [35]. The statistical properties of GP are determined by their mean function $m(x)$ and covariance function $k(x, x')$ [36]. Gaussian prior is composed of mean function and variance function, and its process can be described as follows.

$$y = f(x_n) + \xi_n \quad (7)$$

where x is the input, y is the output, $f(x_n)$ is a hidden function, and ξ_n is noise. Given the data $D = [(x^*, y^*)]$, the joint distribution of training set y and y^* can be obtained according to the Bayesian principle.

$$\bar{y}^* | X, y, X^* \sim N[\bar{y}^*, \text{cov}(\bar{y}^*)] \quad (8)$$

$$\bar{y}^* = K(x^*, X) \left(K(X, X) + \sigma_n^2 I \right)^{-1} y \quad (9)$$

$$\text{cov}(\bar{y}^*) = k(x^*, x^*) - K(x^*, X) \left[K(X, X) + \sigma_n^2 I \right]^{-1} K(X, x^*) \quad (10)$$

where y^* and $\text{Cov}(y^*)$ are fitting values and variances, respectively.

The super parameters set of the GPR model is one of the key factors affecting the model prediction. In this paper, the square exponential covariance function is selected as the covariance function described as follows:

$$K(x_i, x_j; \Theta) = \sigma_f^2 \exp \left[-\frac{1}{2} (x_i - x_j)^T M (x_i - x_j) \right] + \sigma_n^2 \delta_{ij} \quad (11)$$

In Equation (11), σ_f^2 is the signal variance. $M = \text{diag}(l^{-2})$ is a symmetric matrix with y per parameters, l is the characteristic length scale parameter, and $\Theta = (l, \sigma_f^2, \sigma_n^2)$ is usually called the super parameter.

After the super parameters of the GPR model is obtained, the model is established.

3.2. Online Evaluation Using CMD

The health assessment model of hydropower units reflects the complex relationship between the health status of hydropower units and many influencing factors in normal operation. When the hydropower units deteriorate, their original relationship will inevitably change, resulting in their operation state deviated from the original health state. Therefore, to identify the deterioration or failure state of hydropower units, it is necessary to find a method to quantify the degree of deviation between the two states.

The MD [37] was proposed by Mahalanobis, an Indian statistician. It is a method to calculate the similarity of two sample sets. In other fields, it is often used as an anomaly detection index and has achieved good results [38,39]. This method is very suitable to measure the distance between the characteristic vector and the healthy operation state space. The MD is defined as Equation (12).

$$d_i = \sqrt{(X_i - u)^T C^{-1} (X_i - u)} \quad (12)$$

In Equation (12), $X_i = [x_1, x_2, \dots, x_m]$ is the i th m -dimension observation vector. $u = [u_1, u_2, \dots, u_n]$ is the mean vector, C is the covariance matrix.

Assuming that the characteristic vector of the current operating state for the hydropower units is X_i and considering the influence of the historical state on hydropower units at the current time, the CMD is defined as Equation (13).

$$CMD(t) = \frac{d(t) + d(t-1) + \dots + d(t-1+n)}{n} \quad (13)$$

In Equation (13), n is the sliding window and $d(t)$ is the MD between the current time vector X_i and the health assessment model.

4. Example Verification

4.1. Description of Collection Data Type

To verify the effectiveness of the above methods, this paper takes the bulb tubular hydropower units, with a rated power of 30.77 MW, which was put into use in May 2015, in Guangxi Province of China as the research object. The general structure and some sensors installation positions of the hydropower unit studied are shown in Figure 2. The Monitoring System of NaRI Open-2000 installed in this power station is designed and manufactured by NARI Group Co., Ltd. (Nanjing, China). The function of the Open-2000 monitoring system is to realize the centralized monitoring and control, recording and management of the whole power station, and remote monitoring of the system. Its data acquisition system can collect real-time data from various field control units and external data links, complete the database refresh, and analyze, process and calculate the collected data to form the data needed for monitoring and management, and store these data in the database. The monitoring parameters collected by the acquisition system include electrical parameters such as current, voltage, and power, mechanical parameters such as vibration, travel, displacement, guide vane opening, water level, flow, pressure, and other hydraulic parameters, and thermal parameters such as tile temperature, oil temperature, winding temperature, etc. In this paper, 107 kinds of parameters are selected from the above monitoring parameters for analysis. Some of the original data of some monitoring parameters are shown in Table 1. It can be seen from Table 1 that the data of monitoring are discrete data, which cannot be analyzed by traditional diagnosis method based on the continuous signal. The main types of sensors used are shown in Table 2. Figure 3 shows the installation position of the RTZ bearing bush temperature sensors for collecting the temperature of water guide bearing bush, and basic dimensions parameters of turbine model are shown in Table 3.

Table 1. Part of the original data of some of 107 measuring parameters on 9 October 2015.

Monitoring Parameters	Units	Time				
		8:00:00	8:10:00	8:20:00	8:30:00	8:40:00
Temperature of No.11 forward thrust bearing bush	°C	48.52	48.47	48.46	48.42	48.33
Temperature of No.12 forward thrust bearing bush	°C	48.59	48.65	48.55	8.56	48.51
Unit speed	r/min	78.88	78.88	79.10	78.88	78.79
The swing of oil receiver in X direction	μm	64.39	120.91	87.60	61.38	86.50
Vibration of stator frame in the X direction	μm	38.87	33.34	33.80	28.81	30.26
Vibration of stator frame in the Y direction	μm	40.97	29.42	14.38	18.21	32.44
Vibration of combined bearing in the X direction	μm	12.95	8.73	11.35	11.03	11.32
Vibration of combined bearing in the Y direction	μm	3.85	3.50	2.86	3.60	3.45
Vibration of water guide bearing in the X direction	μm	39.33	28.71	47.77	40.69	41.12
Vibration of water guide bearing in the Y direction	μm	20.94	17.49	14.50	16.18	19.55

Figure 2. The structure of the bulb tubular hydropower units and the location of some sensors.

Table 2. The main sensors used in the monitoring system.

Model	Frequency Response Range	Collection Data Type
Dp-VH-06-0400 low frequency vibration sensors	0.5 Hz–200 Hz	Vibration data
811100-50-09-01 key phasor sensors	0–10 KHz	Key phase data
BENTLY3300XL sensors	0–100 kHz	Obtain the swing data
SZCB-01 magnetoelectric speed sensors	50–5000 Hz	Get the spindle speed
PM388 piezoresistive pressure sensors	07 Mpa	Draft tube outlet pressure

Table 3. Basic dimension parameters of turbine model.

Parameters	Number of Guide Vanes	Number of Runner Blades	Turbine Diameter (mm)	Hub Ratio
	16	4	6700	0.38

4.2. Feature Vector Selection

This paper uses the monitoring data of no-fault records with a sampling interval of 10 min in the first fourth months of its operation to complete feature selection. The active power information of the hydropower units is used to evaluate the operation state of the hydropower units. Therefore, when selecting the monitoring parameters, the parameters closely related to the active power are mainly selected. According to the above correlation analysis methods, the correlation between all simulated monitoring parameters (107 in total) of the monitoring system and the output active power of the hydropower units are calculated. Some results are shown in Table 4, and the vibrations mentioned in this paper are measured vibration.

Figure 3. The installation position of RTZ bearing bush temperature sensors for collecting the temperature of water guide bearing bushe.

Table 4. Correlation analysis results of the partial condition monitoring parameters using maximum information coefficient (MIC), Pearson correlation (PCC), and grey correlation degree (GCD).

Condition Monitoring Parameters	The Correlation Analysis Methods		
	MIC	PCC	GCD
Swing Y of the water guide bearing	0.22	−0.03	0.72
Vibration of oil receiver in the X direction	0.21	0.39	0.88
Vibration of oil receiver in the Y direction	0.21	0.43	0.92
Vibration of stator frame in the X direction	0.15	0.30	0.88
Vibration of stator frame in the Y direction	0.26	0.49	0.86
Vibration of combined bearing in the X direction	0.20	0.37	0.92
Vibration of combined bearing in the Y direction	0.36	0.57	0.93
Vibration of water guide bearing in the X direction	0.20	0.36	0.83
Vibration of water guide bearing in the Y direction	0.20	0.37	0.93
Temperature of No.1 forward thrust bearing bush	0.35	0.45	0.73

From Table 4, the correlation between each monitoring parameter and the active power parameter is quite different. To reduce the influence of weakly correlated or unrelated state parameters on the next step of modeling, the following Equations (14) and (15) are used to select the monitoring parameters closely related to the operation state of the unit, and these monitoring parameters are taken as the input eigenvectors of the initial selection.

$$f(i) = \frac{1}{3}[M(i) + p(i) + G(i)] \quad (14)$$

$$g = \frac{1}{N} \sum_{i=1}^N f(i) \quad (15)$$

In Equations (14) and (15). N is the monitoring parameters of 107 measuring points collected from the hydropower station monitoring system. $M(i)$ is the MIC value of the i th monitoring parameter. $P(i)$ is the PCC value of the i th monitoring parameter. $G(i)$ is the GCD value of the i th monitoring parameter. When $f(i) > g$, it is considered that the i th parameter is closely related to the unit operation state specifying it as a preliminary input feature.

According to the above methods, the input eigenvectors with 41 parameters, i.e., the vibration of the oil receiver in the Y direction, the vibration of the stator frame in the Y direction, the vibration of the combined bearing in the Y direction, the vibration of the water guide bearing in the Y direction, the temperatures of the thrust bearing bushes from No.1 to No.12, the stator coil temperatures from No.1 to No.12, air cooler hot air temperatures from No.1 to No.6, stator core temperatures from No.1 to No.12, and the unit flow, are preliminarily obtained. To reduce the influence of redundant variables, the above parameters are further analyzed. The time-domain variation curves of some monitoring parameters are shown in Figure 4.

Figure 4. Time-domain diagram of from 1 to 6 stator coil temperatures.

From Figure 4, these monitoring parameters are highly correlated, indicating that they have the same roles in reflecting the operation state of the unit. Therefore, one of the parameters is selected as the input feature. Through linear correlation analysis, nine parameters, including the vibration of oil receiver, the stator vibration in the Y direction, the vibration of combined bearing in the Y direction, the vibration of water guide bearing in the Y direction, the temperature of No.1 forward thrust bearing bush, the temperature of No.1 stator coil, the temperature of No.1 air cooler hot air temperature, the temperature of No.1 stator core and unit flow rate, are selected as final input eigenvectors.

4.3. The Establishment of the Health Assessment Model for Hydropower Units

The health model of the hydropower unit is established by using ten monitoring parameters in the initial three months of operation. Each monitoring parameter has 1200 groups of data, and the data sampling time is 10 min. The first 1000 sets of data are used for training the GPR model, and the remaining 200 sets of data are used for model verification. The validation effect of the GPR model is shown in Figure 5a. The predicted curve by the GPR model is in good agreement with the actual monitoring data. To highlight the superiority of the GPR model, the back propagation neural network (BPNN), and the radial basis function neural network (RBFNN) model are used for comparison. BPNN can realize the nonlinear mapping between input samples and output samples, and has the characteristics of self-organization and self-learning, which makes it very effective in nonlinear approximation. In addition, the potential relationship between these data can be summarized through the automatic learning process. So far, there is no definite method to determine the number of nodes in the hidden layer of BPNN. Homik points out that the number of hidden layers should be between $\sqrt{2m+1}$ and $2m+n$, Hecht Nielsen

thinks that the number of nodes is $2m+1$ [40], and Some scholars think that the number of nodes is $\sqrt{m+n}+a$, where a is a constant between 1 and 10 [41]. In these methods, m and n are the number of nodes in the input layer and the output layer respectively. Therefore, according to these empirical formulas, the number of nodes in the hidden layer should be between 4 and 19. Considering the complexity of monitoring parameters, the transfer functions *logsig* and *purelin* are used as the transfer functions of hidden layer and output layer. *Tainglm* adaptive learning algorithm is used to complete network training and prediction, and is programmed and run in MATLAB 2019a environment. The main parameters of the network are as follows: the target error is 6.5×10^{-4} , the learning rate is 0.005, the maximum number of calculations is 5000, and the additional momentum factor is 0.90. The test error of typical network performance with different number of hidden layer nodes is shown in Table 5. Therefore, the best prediction model of BPNN is $9 \times 19 + 1$, and the research results of the BPNN model is seen in Figure 5b. The research results of the RBFNN model are seen in Figure 5c. The comparison between the results obtained with the GPR model to the results of BPNN and RBFNN models is showed in Table 6.

Figure 5. The prediction results of different models, (a) Gaussian process regression (GPR) model, (b) BPNN model, (c) radial basis function neural network (RBFNN) model.

Table 5. The back propagation neural network (BPNN) model performance of different numbers of nodes in the hidden layer.

Numbers of Nodes in Hidden Layer								
	12	13	14	15	16	17	18	19
σ_{MAE}	6.6522%	7.8118%	9.3786%	8.2751%	94.2632%	6.90822%	7.6217%	6.6455%

Table 6. Quantitative results predicted by GPR model, BPNN and RBFNN models.

Models	GPR Model	BPNN Model	RBFNN Model
σ_{MAE}	2.1158%	6.6455%	5.1104%

In this paper, Equation (16) is used to quantitatively test the above prediction results.

$$\sigma_{MAE} = \frac{1}{N_1} \sum_{i=1}^{N_1} \left| \frac{\bar{y}(i) - y(i)}{y(i)} \right| \quad (16)$$

In Equation (16), σ_{MAE} is the absolute value of average relative error. N_1 is the number of test samples $\bar{y}(i)$ is the predicted value and $y(i)$ is the real value of the monitoring data. According to Table 6, the σ_{MAEs} of the GPR model, the BPNN model, and the RBFNN model are 2.1158%, 6.6455%, and 5.1104%, respectively. Obviously, the predicted curve by BPNN model diverges largely from the actual monitoring data. For the monitoring data with strong randomness and high noise of hydropower units, compared with BPNN and RBFNN, the covariance function is used to describe the random distribution of the monitoring data of hydropower units in GPR modeling, and the noise of the data can be effectively identified and separated. Therefore, the GPR model is very suitable for the health assessment model of hydropower units. In the field of condition monitoring, there are many research results on determining the alarm threshold of equipment abnormal operation by analyzing the monitoring data collected by equipment status [38,39,42]. To determine the alarm threshold value of the state index, this paper selects another 2400 groups of data in the fault-free operation time to determine the distribution law of the CMD. The results are shown in Figure 6. It can be seen from Figure 6 that the deviation range between the normal operation of hydropower units and the health model is stable at 0.6 or below. Considering that the hydropower station is easily affected by runoff and undertakes the task of peak load regulation and frequency regulation, the alarm threshold is determined as 0.8. This threshold is considered as a universal value in the operation of the unit. When the value of CMD exceeds 0.8, it indicates that the current state of hydropower units deviates from the health model, which means that the current operating state of hydropower units deviates from the normal state, and there may be potential risks or faults. Several typical cases from the historical monitoring data of this hydropower units are used to verify the effectiveness of the proposed method.

Figure 6. The distribution of the condition monitoring directive (CMD).

4.4. Case Analysis

Through several planned maintenances, the staff of the hydropower station found that the flange plate of the blade of the motor unit leaked oil, the floating bearing bush of the

oil receiver was seriously worn, the oil leakage was large, and the connecting parts of the guide vane mechanism of the unit were broken, and the faults occurred on 29 December 2015, 5 February 2018 and 23 January 2019, respectively. During these periods, the sensors did not fail and did not cause abnormal monitoring data. Therefore, the change rule of the operation state of the hydropower units is studied mainly before and after these faults. To highlight the engineering application of the method proposed, the law of amplitude value change in the time domain of relevant monitoring parameters such as the vibration of oil receiver in the Y direction, the vibration of combined bearing in the Y direction, vibration of water guide bearing in the Y direction, temperature of forwarding thrust bearing bush and temperature of stator core before and after the fault is studied.

Case 1 focuses on the data set from 2:40 am on 19 December 2015, to 7:30 on 3 January 2016, case 2 on the data set from 19:00 on 25 January 2018, to 10:50 on 14 February 2018, and case 3 on the data set from 10:30 am on 5 December 2018, to 8:30 am on 26 January 2019. The average time interval for these three data sets is 10 min. The results are shown in Figures 7–12. In these pictures, the black dotted line indicates the time when the CMD curve exceeds the alarm threshold, and the red represents the time when the unit faults were found in the planned inspection.

Figure 7. The changing curve of the CMD for hydropower units in case 1.

Figure 8. The changing curve of correlation monitoring parameters of hydropower units in case 1: (a) vibration of oil receiver in the Y direction, (b) vibration of combined bearing in the Y direction, (c) vibration of water guide bearing in the Y direction, (d) temperature of forwarding thrust bearing bush, (e) temperature of the stator core.

Figure 9. The changing curve of the CMD for hydropower units in case 2.

Figure 10. The changing curve of correlation monitoring parameters of hydropower units in case 2: (a) vibration of oil receiver in the Y direction, (b) vibration of combined bearing in the Y direction, (c) vibration of water guide bearing in the Y direction, (d) temperature of forwarding thrust bearing bush, (e) temperature of the stator core.

Figure 11. The changing curve of the CMD for hydropower units in case 3.

Figure 12. The changing curve of correlation monitoring parameters of hydropower units in case 3: (a) vibration of oil receiver in the Y direction, (b) vibration of combined bearing in the Y direction, (c) vibration of water guide bearing in the Y direction, (d) temperature of forwarding thrust bearing bush, (e) temperature of the stator core.

In case 1, when the unit fails, the amplitudes of each monitoring parameter of the hydropower unit are far lower than the alarm threshold set by the monitoring system, for example, it can be seen from Figure 8 that the vibration amplitude of the oil receiver in the Y direction is in the range of 20 to 40 μm , and the temperature value of monitoring parameters is lower than 100 $^{\circ}\text{C}$. It can be seen from Tables 7 and 8 that only when the vibration amplitude of the oil receiver in the Y direction is greater than 250 μm , or the temperature value of the monitoring parameter is greater than 100 $^{\circ}\text{C}$, the first level alarm of the monitoring system will be triggered. The proposed CMD curve exceeded the alarm threshold of 0.8 at 21:30 on 27 December 2015, which can monitor the fault state of the unit in advance.

Table 7. Vibration swing alarm threshold of monitoring system of hydropower station.

Measuring Point	Rotational Speed (r/min)	Zone A Qualified (μm)	Zone B Level 1 Alarm (μm)	Zone C Unit Shutdown (μm)
Swing of bearing	79	0–260	260–530	>530
Vibration of oil receiver and bearing	79	0–250	250–350	>350
Vibration stator frame	79	0–350	350–600	>600
Vibration of water guide bearing and combined bearing	79	0–120	120–170	>170

Table 8. Temperature alarm threshold of the monitoring system of hydropower station.

Measuring Point	Rotational Speed (r/min)	Zone A Qualified (°C)	Zone B Level 1 Alarm (°C)	Zone C Unit Shutdown (°C)
Temperature of stator core	79	0–100	100–105	>105
Temperature Stator coil	79	0–100	100–105	>105
Temperature of forwarding thrust bearing bush	79	0–100	100–105	>105

To reduce the influence of accidental factors and verify the universality of the proposed method, case 2 and case 3 are further used for analysis. It can also be seen from in case 2 and case 3 that when the unit fails, the amplitude of each monitoring parameter of the hydropower unit is far lower than the alarm threshold of the monitoring system. Although with the change of time, the vibration amplitude of the oil receiver in the Y direction fluctuates in the range of 20 to 50 μm , and the amplitude value changes greatly, it is still far lower than 250 μm . It can be seen from case 2 that the CMD curve exceeded the alarm threshold of 0.8 at 5:50 on 31 January 2018, and the fault status of the unit was monitored in advance. In case 3, the CMD curve also exceeded the alarm threshold of 0.8 at 20:20 on 11 December 2018, and the fault status of the unit was also monitored in advance. It can be seen from the above cases that the fault alarm and early warning means of the hydropower station monitoring system are relatively single, and the alarm setting value of the hydropower station monitoring system is mainly fixed value.

However, in the process of equipment operation, with the continuous change of monitoring status with environmental conditions and operating conditions, the collected monitoring point data amplitude does not exceed the set value, but the equipment has failed. Therefore, the role of hydropower station monitoring system in engineering application is limited. Although the deterioration or failure information of hydropower units must be reflected in the massive data collected from each monitoring point, the above failures were found in the scheduled maintenance. This also shows that it is difficult for hydropower station staff to master the operation status of units in real-time through the information reflected by these massive data. In view of the problems existing in these engineering applications, this paper studies how to extract the operation status information from the massive monitoring data. According to the method proposed in this paper, the health model of hydropower units in normal operation is established, this model reflects the complex relationship between the operation status of hydropower units and multiple monitoring parameters, and monitoring index based on MD is set to measure the change degree of these complex links. From the above cases, the CMD curve is abnormal ahead of the maintenance time of the staff. In order to explore the change trend of the CMD curve after maintenance, with the first maintenance of the unit on 18 February 2016, we selected the data of the unit from 5:30, 16 February 2016 to 13:40, 19 February 2016, using the method proposed in this paper for analysis, and the results are shown in the Figure 13. It can be concluded from the Figure 13 that after the maintenance of the unit, the CMD curve drops to below 0.8. Therefore, a conclusion is drawn that when the unit fails, the state of the unit at this time will inevitably lead to the change of the original structure information relationship among the parameters, and GPR model captures the subtle relationship of these changes and the evaluation and monitoring index CMD is more sensitive than the existing monitoring and alarm system to catch the abnormal operation of the unit, which can effectively find the potential failure of the unit, send out the early warning information, and give the decision-maker plenty of time to make the maintenance plan in advance.

Figure 13. The change curve of CMD of hydropower unit after maintenance was completed in case 1.

5. Conclusions

In this paper, a state evaluation method of hydropower units based on GPR theory is proposed. In this paper, the PCC, GCD, and MIC analysis methods are used to mine 41 monitoring parameters which are closely related to the operation of hydropower units from 107 monitoring parameters. To reduce the influence of redundant parameters on the modeling, PCC is used to analyze the correlation among 41 monitoring parameters. Nine monitoring parameters are obtained and used as input eigenvectors. Secondly, to verify the performance of the GPR model, this paper uses the classic BPNN for comparative analysis, and the results show that the GPR model is superior to the BPNN model, and BPR model. Finally, a monitoring index based on MD is designed to measure the deviation between the abnormal state of the unit and the state of the health model. In the process of verifying the effectiveness of the proposed method through several fault examples, it is compared with the analysis results of the existing monitoring system of hydropower stations. The results show that the proposed method can effectively track the operation state change process of the hydropower units and identify the potential fault of the unit in advance. The proposed method mainly uses the information rich data stored in the monitoring system of hydropower station, which is usually easy to be ignored. In addition, the method mainly uses the data stored in the system and does not need additional equipment to obtain, so it reduces the cost of research. Compared with the existing research methods, this method has the advantage of simplicity in application. The method proposed in this paper can detect the faults of the unit before the regular inspection and maintenance, and the shortest time is two days ahead of the regular inspection and maintenance. During this period, the crew members can make the best maintenance decisions, reduce the time of unit operation with faults, bring down the incidence of accidents, and lower the overall maintenance cost.

The method in this paper is suitable for the monitoring data analysis of a hydropower station in Nanning, Guangxi, China. However, if it is to be extended to other power stations or similar equipment, some improvements should be made in combination with the research object to improve the generalization ability of the method. For example, the model parameters are timely updated according to the healthy evolution of the system. Secondly, most of the monitoring data studied in this paper are the normal operation data of hydropower units, with less fault data and less fault types with labels, which cannot accurately locate and identify the fault types. Only when the data is complete can the fault location be realized by using the existing fault diagnosis methods. The above is also the focus of our follow-up work.

Author Contributions: Formal analysis, H.H. (Huade Huang), A.Q. and X.L.; investigation, H.H. (He Huang) and J.F.; funding acquisition, H.M., and Z.H.; methodology, H.H. (Huade Huang) and

Y.Y.; supervision, H.M. and X.L.; validation, H.H. (He Huang); writing—original draft, H.H. (Huade Huang); writing—review and editing, H.M. and A.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Science and Technology Base and Talents Special Project of Guangxi Province (grant No. AD19259002), the Natural Science Foundation of Guangxi (Grant Nos. 2018GXNSFAA180116, and 2016GXNSFAA380119), the National Natural Science Foundation of China (Grant No. 51365006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available because the data is company confidential and should not be shared with unauthorized individuals.

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

References

1. Quaranta, E.; Bonjean, M.; Cuvato, D.; Nicolet, C.; Dreyer, M.; Gaspoz, A.; Rey-Mermet, S.; Boulicaut, B.; Pratalata, L.; Pinelli, M.; et al. Hydropower Case Study Collection: Innovative Low Head and Ecologically Improved Turbines, Hydropower in Existing Infrastructures, Hydropeaking Reduction, Digitalization and Governing Systems. *Sustainability* **2020**, *12*, 8873. [\[CrossRef\]](#)
2. Tran, V.; Pham, H.; Yang, B.-S.; Nguyen, T. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mech. Syst. Signal Process.* **2012**, *32*, 320–330. [\[CrossRef\]](#)
3. Delechelle, E.; Lemoine, J.; Niang, O. Empirical mode decomposition: An analytical approach for sifting process. *IEEE Signal Process. Lett.* **2005**, *12*, 764–767. [\[CrossRef\]](#)
4. Lu, S.; Wei, J.; Bao, H.; Xue, Y.; Ye, W. The dynamic hydropower troubleshooting information based on EMD multi-scale feature entropy extraction. *Int. J. Mob. Commun.* **2017**, *15*, 677–692. [\[CrossRef\]](#)
5. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
6. Zhang, X.; Miao, Q.; Zhang, H.; Wang, L. A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery. *Mech. Syst. Signal Process.* **2018**, *108*, 58–72. [\[CrossRef\]](#)
7. Li, P.; Kong, F.; He, Q.; Liu, Y. Multiscale slope feature extraction for rotating machinery fault diagnosis using wavelet analysis. *Measurement* **2013**, *46*, 497–505. [\[CrossRef\]](#)
8. Miceli, R.; Gritli, Y.; Tommaso, A.D.; Filippetti, F.; Rossi, C. Vibration signature analysis for monitoring rotor broken bar in double squirrel cage induction motors based on wavelet analysis. *Compel Int. J. Comput. Math. Electr.* **2014**, *33*, 1625–1641. [\[CrossRef\]](#)
9. Niu, G.; Pecht, M. A framework for cost-effective and accurate maintenance combining CBM RCM and data fusion. In Proceedings of the 2009 8th International Conference on Reliability, Maintainability and Safety, Chengdu, China, 20–24 July 2009; City University of Hong Kong: Hong Kong, China, 2009.
10. Cheng, Y.; Chung, T.S.; Yu, C.W.; Chung, C.Y.; Zeng, M.; Sun, X. Application of reliability-centered stochastic approach and FMECA to conditional maintenance of electric power plants in China. In Proceedings of the 2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies, Hong Kong, China, 5–8 April 2004; IEEE: New York, NY, USA, 2004.
11. Silva, D.; Carazas, F.; Souza, G. Method to select instrumentation for hydraulic turbines in retrofitting process. In Proceedings of the 2009 International Conference on Computers & Industrial Engineering, Troyes, France, 6–9 July 2009; IEEE: New York, NY, USA, 2009.
12. Kumar, R.; Singal, S.K. Cost effective, reliable and accurate maintenance model for SHP plants using RCM. In Proceedings of the 2014 International Conference on Reliability Optimization and Information Technology (ICROIT) Faridabad, NCR, India, 6–8 February 2014; IEEE: New York, NY, USA, 2014.
13. An, X.; Pan, L.; Yang, L. Condition parameter degradation assessment and prediction for hydropower units using Shepard surface and ITD. *Trans. Inst. Meas. Control* **2014**, *36*, 1074–1082. [\[CrossRef\]](#)
14. Zhou, J.; Shan, Y.; Liu, J.; Xu, Y.; Zheng, Y. Degradation Tendency Prediction for Pumped Storage Unit Based on Integrated Degradation Index Construction and Hybrid CNN-LSTM Model. *Sensors* **2020**, *20*, 4277. [\[CrossRef\]](#)
15. Si, X.-S.; Wang, W.; Hu, C.-H.; Zhou, D.-H. Remaining useful life estimation—A review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [\[CrossRef\]](#)
16. Fu, W.; Wang, K.; Li, C.; Li, X.; Li, Y.; Zhong, H. Vibration trend measurement for a hydropower generator based on optimal variational mode decomposition and an LSSVM improved with chaotic sine cosine algorithm optimization. *Meas. Sci. Technol.* **2018**, *30*, 015012. [\[CrossRef\]](#)
17. Fu, W.; Zhou, J.; Zhang, Y.; Zhu, W.; Xue, X.; Xu, Y. A state tendency measurement for a hydro-turbine generating unit based on aggregated EEMD and SVR. *Meas. Sci. Technol.* **2015**, *26*, 125008. [\[CrossRef\]](#)

18. Hu, X.; Li, C.; Tang, G. A Hybrid Model For Predicting The Degradation Trend Of Hydropower Units Based On Deep Learning. In Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 25–27 October 2019; IEEE: New York, NY, USA, 2019.
19. Zhou, K.; Zhang, J.; Shan, Y.; Ge, M.; Ge, Z.; Cao, G. A Hybrid Multi-Objective Optimization Model for Vibration Tendency Prediction of Hydropower Generators. *Sensors* **2019**, *19*, 2055. [\[CrossRef\]](#) [\[PubMed\]](#)
20. An, X.; Pan, L. Characteristic parameter degradation prediction of hydropower unit based on radial basis function surface and empirical mode decomposition. *J. Vib. Control* **2015**, *21*, 2200–2211. [\[CrossRef\]](#)
21. An, X.; Yang, L.; Pan, L. Nonlinear prediction of condition parameter degradation trend for hydropower unit based on radial basis function interpolation and wavelet transform. *Arch. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2015**, *229*, 1–9. [\[CrossRef\]](#)
22. Qiao, L.; Chen, Q. Forecasting Models for Hydropower Unit Stability Using LS-SVM. *Math. Probl. Eng.* **2015**, *2015*, 1–9. [\[CrossRef\]](#)
23. Yang, B.; Han, T.; Kim, Y. Integration of ART-Kohonen neural network and case-based reasoning for intelligent fault diagnosis. *Expert Syst. Appl.* **2004**, *26*, 387–395. [\[CrossRef\]](#)
24. Błaszczyszki, J.; de Almeida Filho, A.T.; Matuszyk, A.; Szelag, M.; Słowiński, R. Auto loan fraud detection using dominance-based rough set approach versus machine learning methods. *Expert Syst. Appl.* **2021**, *163*, 113740. [\[CrossRef\]](#)
25. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 865–873. [\[CrossRef\]](#)
26. Wahab, O.; Bentahar, J.; Otrok, H.; Mourad, A. Resource-Aware Detection and Defense System against Multi-Type Attacks in the Cloud: Repeated Bayesian Stackelberg Game. *Ieee Trans. Dependable Secur. Comput.* **2019**, *1*. [\[CrossRef\]](#)
27. Avendaño-Valencia, L.D.; Chatzi, E.N.; Tcherniak, D. Gaussian process models for mitigation of operational variability in the structural health monitoring of wind turbines. *Mech. Syst. Signal Process.* **2020**, *142*, 106686. [\[CrossRef\]](#)
28. Chandrasekhar, K.; Stevanovi, N.; Cross, E.J.; Dervilis, N.; Worden, K. Gaussian Processes for Structural Health Monitoring of Wind Turbine Blades. In Proceedings of the 12th International Workshop on Structural Health Monitoring, Standord, CA, USA, 10–12 September 2019; Standord University: Stanford, CA, USA, 2019.
29. Kong, D.; Chen, Y.; Li, N. Gaussian process regression for tool wear prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 556–574. [\[CrossRef\]](#)
30. Nicewander, J.L.R.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66.
31. Shi, J.; Ding, Z.; Lee, W.J.; Yang, Y. Hybrid Forecasting Model for Very-Short Term Wind Power Forecasting Based on Grey Relational Analysis and Wind Speed Distribution Features. *IEEE Trans. Smart Grid* **2014**, *5*, 521–526. [\[CrossRef\]](#)
32. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [\[CrossRef\]](#)
33. Simon, N.; Tibshirani, R. Comment on “Detecting Novel Associations In Large Data Sets” by Reshef Et Al, Science Dec 16, 2011. *arXiv* **2014**, arXiv:1401.7645.
34. Kinney, J.B.; Atwal, G.S. Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3354–3359. [\[CrossRef\]](#)
35. Seeger, M. Gaussian processes for machine learning. *Int. J. Neural Syst.* **2004**, *14*, 69–106. [\[CrossRef\]](#)
36. Cai, H.; Jia, X.; Feng, J.; Li, W.; Hsu, Y.-M.; Lee, J. Gaussian Process Regression for numerical wind speed prediction enhancement. *Renew. Energy* **2020**, *146*, 2112–2123. [\[CrossRef\]](#)
37. Maesschalck, R.D.; Jouan-Rimbaud, D.; Massart, D.L. The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 1–18. [\[CrossRef\]](#)
38. Bangalore, P.; Tjernberg, L.B. An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings. *IEEE Trans. Smart Grid* **2015**, *6*, 980–987. [\[CrossRef\]](#)
39. Niu, G.; Singh, S.; Holland, S.W.; Pecht, M. Health monitoring of electronic products based on Mahalanobis distance and Weibull decision metrics. *Microelectron. Reliab.* **2011**, *51*, 279–284. [\[CrossRef\]](#)
40. Deng, X.J.; Kang, T.; Wang, D.S. Application of BP neural network in predicting the cement materials performance. *J. Chem. Pharm. Res.* **2014**, *6*, 1681–1688.
41. Zheng, J.; Zhang, Q.; Wang, X. The prediction of backfilling drill-hole’s life based on BP neural networks. *J. Xiangtan Norm. Univ. Nat. Sci. Ed.* **2008**, *30*, 40–44.
42. Wang, Y.; Miao, Q.; Duan, T.; Tsui, K.L.; Pecht, M.G. Online Anomaly Detection for Hard Disk Drives Based on Mahalanobis Distance. *IEEE Trans. Reliab.* **2013**, *62*, 136–145. [\[CrossRef\]](#)