

Research Article

Fault Diagnosis Strategy for Wind Turbine Generator Based on the Gaussian Process Metamodel

Dongmei Zhang ¹, Jun Yuan ², Jiang Zhu,³ Qingchang Ji ¹, Xintong Zhang,⁴
and Hao Liu ⁵

¹School of Automation and Communication Engineering, Hebei University of Water Resources and Electric Engineering, Cangzhou, Hebei, China

²China Institute of FTZ Supply Chain, Shanghai Maritime University, Shanghai, China

³Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, China

⁴Powerchina Hebei Electric Power Design & Research Institute Co., Ltd., Shijiazhuang, China

⁵Vestas Wind Technology (China) Co., Ltd., Shanghai, China

Correspondence should be addressed to Jun Yuan; yuanj@shmtu.edu.cn

Received 25 August 2019; Revised 18 December 2019; Accepted 31 December 2019; Published 21 January 2020

Academic Editor: Yang Li

Copyright © 2020 Dongmei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To facilitate continuous development of the wind power industry, maintaining technological innovation and reducing cost per kilowatt hour of the electricity generated by the wind turbine generator system (WTGS) are effective measures to facilitate the industrial development. Therefore, the improvement of the system availability for wind farms becomes an important issue which can significantly reduce the operational cost. To improve the system availability, it is necessary to diagnose the system fault for the wind turbine generator so as to find the key factors that influence the system performance and further reduce the maintenance cost. In this paper, a wind farm with 200 MW installed capacity in eastern coastal plain in China is chosen as the research object. A prediction model of wind farm's faults is constructed based on the Gaussian process metamodel. By comparing with actual observation results, the constructed model is proved able to predict failure events of the wind turbine generator accurately. The developed model is further used to analyze the key factors that influence the system failure. These are conducive to increase the running and maintenance efficiency in wind farms, shorten downtime caused by failure, and increase earnings of wind farms.

1. Introduction

Nowadays, global climate warming has caused frequent occurrences of climatic anomaly, extreme climate, and major natural disasters, thus bringing serious challenges to sustainable development. To solve the significant increase of energy demands and greenhouse gas emissions, energy transformation and development is a big problem which is highly concerned by the international society. As the representative of renewable energy sources, wind power is the renewable energy power generation technology which grows the most quickly in the world. The large-scaled commercialization of wind power has been formed gradually. Wind power generation is becoming a mature technology with cost competitiveness [1–6]. The Renewables 2019 Global Status

Report (GSR) which is released by “Renewable Energy Policy Network for the 21st Century (REN21)” states that, in 2018, the newly installed capacity of wind power in the world reached 51 GW and the accumulative installed capacity was 591 GW. In the past one decade, China occupied the leading role in global wind power development. In the global newly installed capacity of wind power in 2018, China's market share accounted for 41%, manifested by 21.1 GW of the newly installed capacity of wind power and 209.5 GW of the accumulative installed capacity. China has become the country with the largest wind power scale in the world [7].

However, the rapid development of wind power generation in China depends on the Feed-in-Tariff (FIT) mechanism to some extent. Currently, fixed electric charge policy is the main supporting mechanism to renewable

energy sources in China, but the FIT mechanism cannot offer sustainable capital supports to the renewable energy source project with increasing scale. It is imagined that wind power development in China has to get rid of dependence on FIT basically from 2020 to 2022 [8]. On May, 2018, the National Energy Administration hereby regulated the new centralized onshore wind power projects and offshore wind power projects both shall configure and determine the on-grid price [9].

To help wind power industrial development cope with the challenge of FIT reduction successfully, optimizing management of the wind farms is another effective strategy except for reducing cost per kilowatt hour of the electricity generated by WTGS by technological innovation. Facing with annual increase of installed capacity in wind farms and running time of generators, owners of the wind farms propose higher requirements on enhancing maintenance decision management of wind turbine generators, controlling operation and maintenance cost, and protecting safe sound operation of the generators [10, 11]. Since the natural environment of the wind farm is generally tough, the operation health state of the wind turbine generator fluctuates violently. As an important part of the power production system in the wind farm, operation fault of the wind turbine generator may influence the whole production link significantly. Therefore, how to enhance operation availability of the wind turbine generator to increase grid generation of the wind farm becomes extremely important.

In this aspect, major influencing factors that influence operation availability have to be recognized firstly. There are various influencing factors such as wind, blade, rotor, bearing, generator, gearbox [12], pitch, and so on. They may bring about different types of faults, which include blade faults, generator faults, gearbox faults, pitch machine faults, and yaw system faults [13]. When gear and bearing goes wrong, it could lead to gearbox faults and reduce the generator accuracy. If the yaw counter times out, the wind turbine generator would be unable to get best wind direction in time, which is not conducive to power generation efficiency. We can also analyze the pitch angle to predict the faults of the pitch mechanism to prevent blade damage and unstable power generation. These faults have to be predicted and diagnosed accurately by scientific methods, thus enabling to formulate targeted strategies. However, it is found from actual maintenance practices of wind turbine generators that there is a high nonlinear relationship between the turbine fault and relevant factors. Traditional linear model is difficult to analyze the relationship between the turbine fault and relevant factors. Hence, it is very important to construct a nonlinear model for turbine fault analysis.

There are various nonlinear models which are extensively used, including artificial neural network (ANN) [14–16], support vector machine (SVM) [17, 18], Gaussian process (GP) metamodel [19–21], and adaptive network-based fuzzy inference system (ANFIS) [22]. As a nonlinear dynamic system, the neural network can map nonlinear functions and has been widely utilized in pattern recognition and classification, such as distinguishing a variety of factors to different types of faults. Back propagation neural

network is the most commonly used networks, and it is an agile tool to settle complex identification problems in fault diagnosis of the wind turbine. It has good abilities of parallel distributed processing, self-learning, associative memory, and precisely nonlinear pattern recognition [16]. However, there are many types of faults in operation of the wind farm, which make the network structure difficult to determine, or there are too many nodes; it is therefore inconvenient to train this model. SVM is a universal learning method, evolved from statistical learning theory. Nonlinear SVM has been used to appraise the deviation of prediction to reflect the potential risk factors in the short-term wind power. The research shows that wavelet transform-support vector machine (WT-SVM) models have better accuracy than the traditional radical basis function (RBF) SVM models due to their multiresolution features [12]. But, the choice of the optimal wavelet base is confusing. The Gaussian process metamodel has been widely used in many different fields due to its good mathematical characteristics and flexibility [23, 24]. Although the Gaussian process metamodel has also been applied in system fault analysis, research on its applications in wind turbine generator fault analysis is still at the initial stage. In this paper, the Gaussian process metamodel is constructed for prediction of the wind turbine generator fault. This model is used to analyze relationships between different factors and turbine fault and predict the probability of turbine fault occurrence under different states. The main contributions of this paper can be summarized as follows:

- (i) Identify the main fault factors that may cause system failure for different fault systems and categorize the fault levels based on the system failure time.
- (ii) Develop Gaussian process metamodels for different fault systems to represent the relationship between fault factors and system failure time. The developed models are further used to predict the fault levels for different fault systems.
- (iii) Analyze the effects of different fault factors on the system failure time so as to find the sensitive factors that need to be carefully maintained.
- (iv) Conduct a real case study to illustrate the performance of the proposed method and analyze the fault events of the wind turbine generator.

The rest of this paper is organized as follows: the description of the wind turbine system is given in Section 2. The development of the Gaussian process metamodel and its application to fault diagnosis of the wind turbine generator are given in Section 3. A case study is given in Section 4, and the conclusion is provided in Section 5.

2. Wind Turbine System

Wind turbine system is composed of several different subsystems which make the system quite complex. The five main subsystems include driving system, electrical control system, pitch system, yawing system, and cooling system. The failure of any subsystem may make the whole system

shutdown. The description of these subsystems is given as below based on the double-fed induction generator.

The driving system, which is used to raise the rotation speed through the main gearbox, usually includes main shaft, gearbox, and coupling. As the driving ratio is quite high, the output speed of the gearbox will be raised to more than 100 times. If an error happens in any component of this system, it may lead to vibration or wheel gear damage.

The electrical control system is used to monitor all the signals of the wind turbine and adjust the grid frequency. This system, especially the converter system, is the key part of the wind turbine. This is because only the frequency adjusted by the converter system meets grid's requirement, the electric power can then be transported from the wind turbine to the main grid stably. Any error of this system may make the wind turbine fail to be cut-in.

The pitch system is used to catch the wind energy in an optimization way. By the intelligent control according to the different wind speeds and directions, the pitch system can be optimized to catch the maximal wind energy. To support the optimization of control, there are many protection sensors in the pitch system. If some important sensors are broken, it may lead to the shutdown of the wind turbine or being out of speed. Therefore, the maintenance of sensors in the pitch system is quite important during the operation.

Yawing system is used to detect the wind direction. The yawing system usually contains four or six yawing motors and some sensors. When the wind direction changes, the motors will work and adjust the wind turbine to the optimal direction where the pitch system can capture the highest wind energy. To trigger this motion, the signal has to be successfully sent from the wind vane. If the wind vane or one of these motors has error, the wind turbine could not find the correct wind direction, and the wind turbine will stop.

Cooling system has two types: water cooling system and air cooling system. By contrast, the water cooling is usually more efficient than air cooling. The function of this system is to decrease the temperature of the electrical control system and the generator. Once the running temperature is higher than the system setting, the wind turbine will stop to avoid the persistent increase of temperature and protects the internal devices.

3. Fault Diagnosis with a Gaussian Process Metamodel

Gaussian process metamodel is a statistical approximation of the original system based on the input and output data. When using the metamodel, the original system can be viewed as a black-box model. The metamodel is used to analyze relationships in the system without understanding on the internal structure of the system. To apply the Gaussian process metamodel for fault diagnosis, the input includes the fault factors (the factors that cause the fault) such as the wind power mismatch failure, pitch charger failure, and cable switch failure. Specifically, each fault factor has its standard setting value, and the input values to the model can be represented as the percentage of deviation from the standard setting value. The output of interest in the

study is the occurrence of different fault levels under different fault systems. As the fault level depends on the failure time, the essential output of interest is the failure time of each fault. After the identification of the input and output in the system, the data for these input and output are collected from real system observations. Next, the details of Gaussian process metamodel development and the application for fault diagnosis are provided.

Let $y(x)$ be the system output for the input x . In our case study, x represents the deviation from standard setting value for the considered fault factors, and $y(x)$ represents the failure time for the wind turbine system. Specifically, four fault systems are considered for the wind turbine system, including the pitch system, cooling system, converter, and others. For each fault system, x can represent the corresponding fault factors, and $y(x)$ can represent the failure time for this fault system. For instance, in the cooling system, x denotes the deviation values for low inlet valve temperature and generator temperature comparison failure, and $y(x)$ denotes the failure time for this cooling system. A Gaussian process metamodel can be developed for each fault system. Therefore, there are four Gaussian process metamodels that can be developed for four fault systems. As model development procedure is the same for different fault systems, the output $y(x)$ and the corresponding input x are taken as examples to show the model development.

With the Gaussian process model, $y(x)$ can be hypothesized as the Gaussian process, which is specified by its mean function $m(x)$ and its variance function $\Sigma(x, x)$. The mean function $m(x)$ can be hypothesized as a constant in most real applications. The variance function $\Sigma(x, x)$ describes spatial correlation among different input points, and the corresponding covariance function can be expressed in different forms [25]. The comparative study of different covariance functions is performed by Pandit and Infield [26]. Here, the commonly used Gaussian covariance function is adopted due to its flexibility. Specifically, the Gaussian covariance can be expressed as follows:

$$\Sigma(x, x) = \tau^2 R(x, x), \quad (1)$$

where τ^2 is an unknown variance and $R(x, x)$ is the correlation function. The correlation function can be further expressed as

$$R(x, x) = \exp\left(-\frac{\|d\|}{\theta}\right), \quad (2)$$

where θ is an unknown parameter and $\|d\|$ is the Euclidean distance between two input points.

Hypothesis based on Gaussian process can construct the Gaussian process metamodel by using the system input and output data. The predictive function of the model output can be further deduced. If y^* is the output value under the unknown input setting x^* , the predictive function of y^* can be expressed as

$$y^* | y \sim N\left(m(x) + \Sigma_*^T \Sigma^{-1} (y - m(x)), \Sigma_{**} - \Sigma_*^T \Sigma^{-1} \Sigma_*\right), \quad (3)$$

where y is the actually observed data, $m(x)$ is the mean function, Σ is the variance function between the observed points, Σ_* is the variance function between the unobserved point and observed point, and Σ_{**} is the variance of the unobserved point. The form of the variance function is given by equation (1). It can be known from this predictive function that the prediction of the unobserved point follows the normal distribution with the given mean and variance. The predictive mean is $m(x) + \Sigma_*^T \Sigma^{-1} (y - m(x))$, which can be used to represent the expected failure time at input setting x^* . The predictive variance is $\Sigma_{**} - \Sigma_*^T \Sigma^{-1} \Sigma_*$, which can be used to measure the predictive error of the failure time. One advantage of the Gaussian process metamodel is that it can easily carry out uncertainty assessment so as to provide more robust results [27]. In this paper, the observed failure time is treated as deterministic, and there are not enough data to measure the noise of the observations. Therefore, the predictive variance only accounts for the spatial uncertainty due to the model development. The extension to consider the stochastic model can be found in the research studied by Pandit and Infield [28].

There are several unknown parameters in the developed Gaussian process metamodel and the derived predictive function, such as τ^2 and θ . These parameters are often treated as constant using the estimated values from estimation measures such as the maximum likelihood estimation method [29]. However, there are uncertainties for these estimations, and these uncertainties may significantly influence the model accuracy and predictive performance [30]. To account for these uncertainties, a Bayesian estimation method is proposed which is easy for uncertainty analysis. Let ξ denote the unknown parameters that need to be estimated and $g(\xi)$ denote the prior for these parameters. Different priors can be assigned such as noninformative prior or conjugate prior [31]. In this paper, conjugate prior is assumed for unknown parameters due to its mathematical convenience. That is, Gaussian process mean is assumed to be conditional normal, variance τ^2 is assumed to be inverse Gamma distribution, and parameter θ is assumed to be Gamma distribution. Given the likelihood that the observed data are normally distributed conditional on the unknown parameters, the posterior distribution of the unknown parameters can be derived as the following equation based on the Bayesian theory:

$$g(\xi | y) \propto g(\xi)L(y | \xi), \quad (4)$$

where $L(y | \xi)$ denotes the likelihood function. The posterior distribution can then be used to make inference about the unknown parameters. For instance, the unknown parameters can be estimated as posterior mean or mode. The uncertainties of these parameters can also be taken into account by integrating out these parameters using the numerical integration methods (e.g., Markov chain Monte Carlo method) with respect to the posterior distribution.

Given the estimated parameters, the developed model can be used for further investigation. For instance, the developed metamodel can be used to analyze and predict the occurrence of different fault levels under different fault

systems. The details of applying the proposed method in wind power system fault diagnosis are illustrated in the case study.

4. Case Study

In this case, the fault occurrence of the wind turbine generator is studied based on the Gaussian process metamodel. Four Gaussian process metamodels are developed for four fault systems, respectively, including the pitch system, cooling system, converter, and others. With the developed metamodels, the relationship between the identified fault factors and the occurrence of different fault levels for different fault systems is analyzed. Furthermore, fault occurrences are predicted in operation process of the wind farm using the developed metamodels.

4.1. Case Description. Here, the fault of the wind turbine generator in the wind farm with 200 MW installed capacity in eastern coastal plain in China is analyzed. Fault data of 100 wind turbine generators with 2 MW unit capacity in this wind farm are selected between April 2015 and March 2017; that is, the period of fault data starts from the beginning of operation and lasts 2 whole years. Data are collected every 15 min of the turbine sensor and then sent to the server in the central control room. This case mainly discusses relationships of fault factors with fault type and fault level based on the data collected by the turbine sensor. When analyzing fault occurrence of the wind farm, fault types are classified according to major functional systems in which fault occurs (Table 1), mainly including the pitch system, cooling system, converter, and others (e.g., cable switch and torque comparison). Under different fault types, the system output considered in the prediction model is the failure time. Then, the fault levels can be identified based on the predicted failure time. Four fault levels are divided according to failure time: level I (continuous failure time of unit generator ≥ 7 days), level II (failure time = 4–6 days), level III (failure time = 2–3 days), and level IV (failure time ≤ 1 day).

4.2. Fault Analysis. In this case, 347 fault events for the wind turbine generator took place in the wind farm from 2015 to 2017. The failure events of the wind turbine generator are sent to the server as the analog signal type, which is then transformed to digital and consequently showed as the error code in the interface screen. According to the error code, engineers analyze these errors during the internal system firstly and later go for the onsite checking. Therefore, the input values of fault factors and the failure time and fault levels for the corresponding fault system are recorded, and these data can be used to develop the Gaussian process metamodel for different fault systems. Statistics on occurrence frequency of four fault types are shown in Figure 1. Clearly, occurrences of the fault in the pitch system are significantly higher than rest three fault types.

According to fault classification standards in this case, statistics on occurrence frequency of different levels of faults for the whole system (including four fault systems) in this

TABLE 1: Major fault factors of the wind farm.

Fault system	Fault factors	Factor number
Pitch system	Wind power mismatch failure	1
	Loss of pitch converter signal indicating the ready state	2
	Pitch position comparing fault	3
	Pitch charger failure	4
	Pitch position sensor anomaly failure	5
	Pitch safety chain failure	6
Cooling system	Low inlet valve temperature	7
	Generator temperature comparison failure	8
Converter	Emergency shutdown of converter	9
	Closing failure of 1U3_breaker	10
Others	Cable switch failure	11
	Torque comparison failure	12

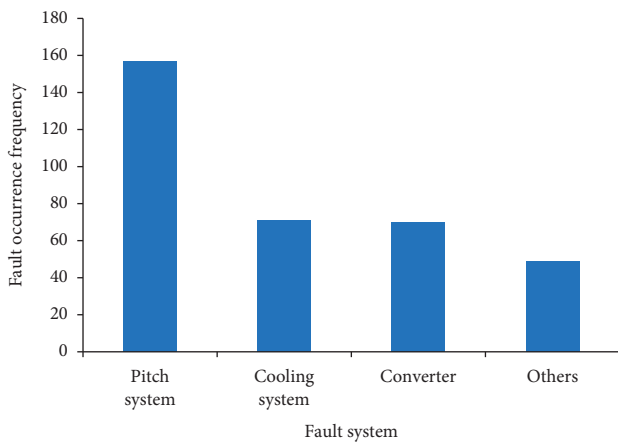


FIGURE 1: Occurrences of four fault types.

wind farm are carried out. Results are shown in Figure 2. The occurrence frequency of level III and level IV is higher, which is more than 3 times higher than that of level I and level II.

Based on the data obtained from the wind farm from April 2015 to March 2017, four Gaussian process metamodels are developed for four fault systems. The developed metamodels are further used for prediction and compared with the observed data. Different fault levels for different fault systems of the wind farm are predicted by the developed Gaussian process metamodel. The occurrence frequency predicted by the model and the actual observed occurrence frequency are shown in Figures 3–6. Obviously, all four fault types have unique distribution characteristics on different fault levels. In the pitch system, cooling system, and converter, the occurrence frequency of level III and level IV is more than 3 times higher than that of level I and level II. However, no significant differences of occurrence frequency have been observed among different fault levels. Although the occurrences of four fault levels predicted by the model are not agreeing with actual observation values, the difference is very small. Based on further analysis on root-mean-square error (RMSE) of model prediction of four fault levels (Table 2), no significant difference exists between model prediction and actual observation. On the contrary, root-mean-square errors (RMSEs) of four fault systems also

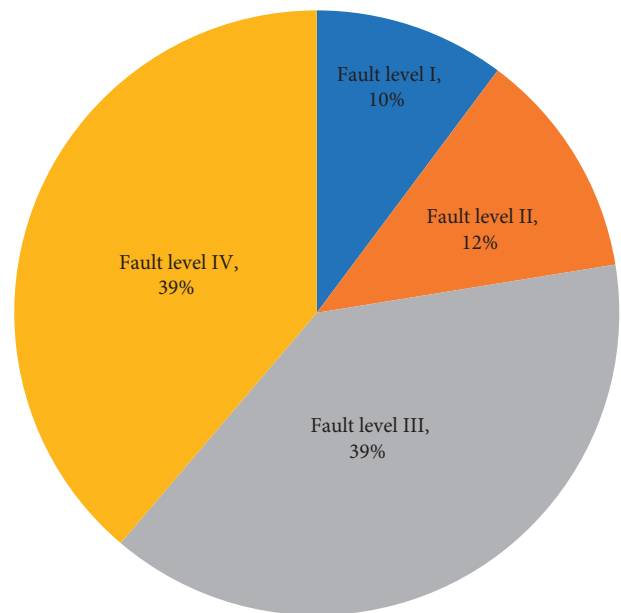


FIGURE 2: Occurrences of different fault levels.

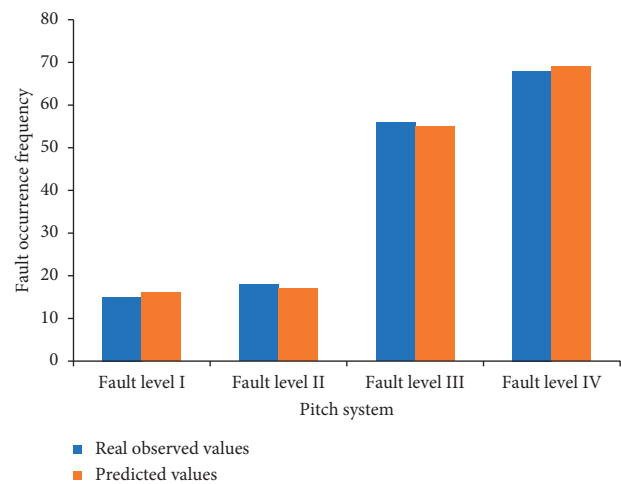


FIGURE 3: Observation and model prediction of fault occurrences in the pitch system.

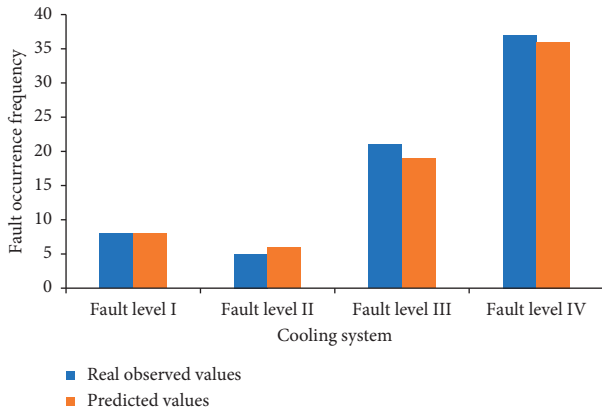


FIGURE 4: Observation and model prediction of fault occurrences in the cooling system.

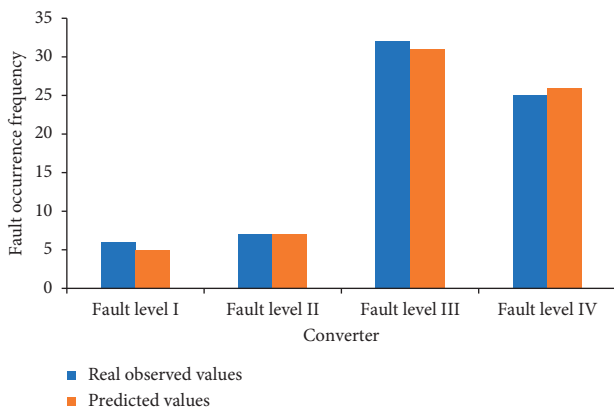


FIGURE 5: Observation and model prediction of fault occurrences in the converter.

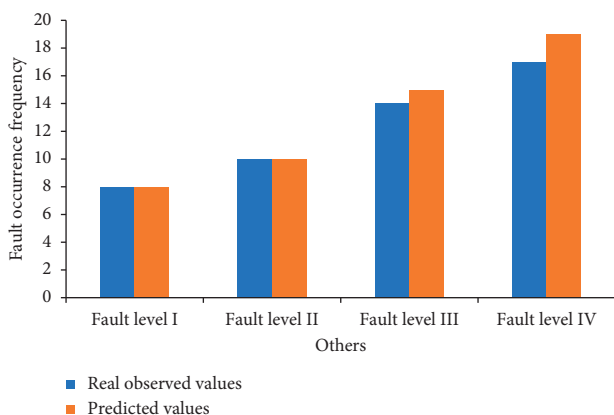


FIGURE 6: Observation and model prediction of fault occurrences in others.

reflect the prediction accuracy of the proposed model. The proposed model has the highest prediction accuracy in the pitch system followed by the cooling system, converter, and others successively. This order is consistent with data size difference among four fault levels in this case. Hence, it could be speculated that increasing fault data size of research objects could increase

prediction accuracy of the model effectively. Noted that the Gaussian process metamodel may not perform well if the data size is too small or it is too large. If the data size is small, the developed model may not be accurate. If the data size is too large, the model can be computationally expensive. In this case study, the data size is appropriate for the model development.

The performance of the developed Gaussian process metamodels is further compared with the neural network, where a two-layer feedforward neural network with adaptive parameters and bias units is used. The RMSEs of four fault systems using the neural network are given in Table 2. The results indicate that the RMSEs using the neural network are similar to the values using the Gaussian process metamodels. The t -test indicates that there is no significant difference between the two methods. Therefore, the two models have similar predictive performance using the obtained training data. The mean absolute error (MAE) and the R -square are also computed to assess the performance of the models. The results as shown in Table 2 indicate that the Gaussian process metamodels are slightly better than the neural networks. Furthermore, the data obtained from July 2017 to September 2017 are used to validate the developed models. The RMSEs for different fault systems using different models are given in Table 2. The results indicate that the Gaussian process metamodels have smaller RMSE values, and the test results show that the difference is significant. Therefore, the Gaussian process metamodels have better predictive performance than the neural network using the validation data in this case study. The MAE and R -square are also computed, and the results in Table 2 show that the Gaussian process metamodels have better performance. As a result, the constructed Gaussian process substitution model could predict fault type and fault level of the wind turbine generator accurately. In addition, compared with the neural network, the Gaussian process metamodels are easier for implementation, and they can account for the uncertainties in the prediction.

The developed Gaussian process metamodel is further used to analyze the effects of deviation of different fault factors on the system failure time. The results are shown in Figure 7. It can be seen that the failure time is longer with larger input deviation for most factors, except factor 6 pitch safety chain failure. The reason is that the observed failure time is quite long for a single event with accepted deviation, and the average failure time for the event with larger input deviation is short. Factors 3, 6, and 12 have relatively larger failure time compared with other factors. Therefore, they are more sensitive to the fault level and have larger probability to cause high-level fault. The probability of different fault levels are further calculated when the deviation of each factor is below -5% , within $-5\% \sim 5\%$, and above 5% . The results are given in Table 3. The results show that when the input deviation is within $-5\% \sim 5\%$, there is usually a smaller probability to get high-level fault (longer failure time). When the input deviation becomes larger (below -5% or above 5%), the probability to

TABLE 2: Root-mean-square errors (RMSE), mean absolute error (MAE), and R -square of prediction of four fault types.

Fault type	Pitch system	Cooling system	Converter	Others	
Training data	RMSE for GP	0.79	0.82	0.84	0.98
	RMSE for neural network	0.79	0.83	0.86	0.99
	MAE for GP	0.75	0.77	0.78	0.91
	MAE for neural network	0.76	0.80	0.77	0.93
	R -square for GP	0.87	0.89	0.90	0.93
	R -square for neural network	0.85	0.88	0.88	0.90
Validation data	RMSE for GP	0.89	0.92	0.91	1.07
	RMSE for neural network	1.02	1.13	1.09	1.21
	MAE for GP	0.83	0.86	0.84	1.02
	MAE for neural network	0.88	0.89	0.87	1.09
	R -square for GP	0.83	0.86	0.87	0.89
	R -square for neural network	0.81	0.84	0.86	0.87

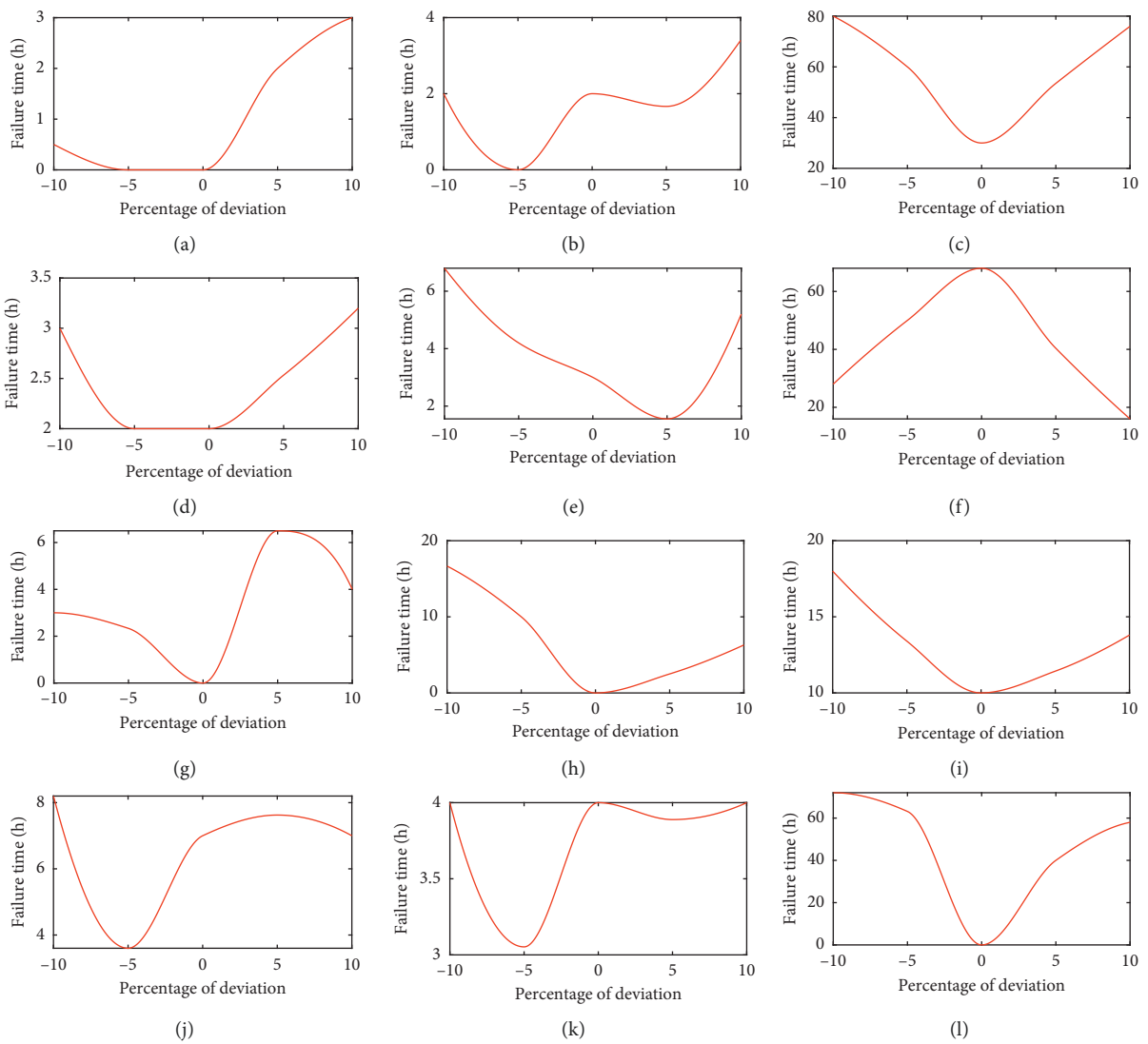


FIGURE 7: Predicted failure time per event with different input deviations for different factors. (a) Factor 1. (b) Factor 2. (c) Factor 3. (d) Factor 4. (e) Factor 5. (f) Factor 6. (g) Factor 7. (h) Factor 8. (i) Factor 9. (j) Factor 10. (k) Factor 11. (l) Factor 12.

TABLE 3: Predicted fault probabilities with different input deviations for different factors.

Factors	Input deviation	Predicted fault probability			
		Level I	Level II	Level III	Level IV
1	Below -5%	0.089	0.206	0.412	0.293
	Within -5%~5%	0.012	0.157	0.314	0.517
	Above 5%	0.135	0.283	0.397	0.185
2	Below -5%	0.039	0.212	0.508	0.241
	Within -5%~5%	0.008	0.196	0.437	0.359
	Above 5%	0.102	0.231	0.529	0.138
3	Below -5%	0.236	0.317	0.354	0.093
	Within -5%~5%	0.194	0.286	0.318	0.202
	Above 5%	0.214	0.287	0.319	0.18
4	Below -5%	0.028	0.203	0.408	0.361
	Within -5%~5%	0.007	0.167	0.325	0.501
	Above 5%	0.037	0.221	0.398	0.344
5	Below -5%	0.183	0.249	0.307	0.261
	Within -5%~5%	0.142	0.236	0.297	0.325
	Above 5%	0.169	0.253	0.310	0.268
6	Below -5%	0.209	0.216	0.238	0.337
	Within -5%~5%	0.318	0.243	0.217	0.222
	Above 5%	0.179	0.208	0.297	0.316
7	Below -5%	0.174	0.213	0.301	0.312
	Within -5%~5%	0.132	0.184	0.293	0.391
	Above 5%	0.169	0.224	0.277	0.33
8	Below -5%	0.201	0.247	0.289	0.263
	Within -5%~5%	0.134	0.201	0.237	0.428
	Above 5%	0.235	0.254	0.283	0.228
9	Below -5%	0.191	0.226	0.257	0.326
	Within -5%~5%	0.110	0.168	0.265	0.457
	Above 5%	0.187	0.243	0.255	0.315
10	Below -5%	0.125	0.231	0.274	0.37
	Within -5%~5%	0.106	0.196	0.357	0.341
	Above 5%	0.162	0.264	0.299	0.275
11	Below -5%	0.136	0.244	0.315	0.305
	Within -5%~5%	0.099	0.217	0.263	0.421
	Above 5%	0.142	0.239	0.334	0.285
12	Below -5%	0.203	0.367	0.201	0.229
	Within -5%~5%	0.156	0.248	0.302	0.294
	Above 5%	0.211	0.319	0.243	0.227

get high-level fault becomes larger. Therefore, it is important to make the input deviation within control in order to improve the turbine generator system's performance.

5. Conclusions

This paper studies wind turbine generator fault events by the Gaussian process metamodel. The Gaussian process metamodels are developed to predict the failure time and the different fault levels of different fault systems given the fault factors. In a specific case study, relationships among different relevant factors in the fault event are analyzed, and the fault prediction model in the wind farm is constructed. According to comparison between model prediction results and actual observation results, the proposed model can predict fault level for different fault types in the wind turbine generator accurately. However, model prediction accuracy is related with data size in the case study. Usually, the model is more accurate

with larger data size. However, it also depends on the accuracy of the data. If the data obtained are not accurate, the model may perform even worse with more data.

The developed model is further used to analyze the effects of different factors on the fault occurrence. The results indicate that some factors have higher impacts on the fault level which have larger probability to get high-level fault, including factor pitch position comparing fault, pitch safety chain failure, and torque comparison failure. These factors should be maintained more appropriately in order to reduce the maintenance cost. The fault factor safety chain failure would cause high-level fault with smaller input deviation. This is due to the data obtained from the wind farm. A small deviation of some input factors may cause longer failure time. It should be noted that the observed failure time may also be influenced by other factors such as the reaction time of the engineers. Therefore, it is important to identify the failure time that is

caused by the input fault factors and further consider the noise in the observations.

In general, large deviations of the fault factors often cause longer failure time and high-level fault. Hence, it is necessary to keep the fault factors under control. The developed Gaussian process metamodel provides an efficient way to analyze the impacts of fault factors without conducting experiments on real systems. In addition, using of the Gaussian process metamodel in operation management of the wind farm can predict fault events that are going to happen accurately so that maintenance staff can adopt effective protection measures or formulate maintenance schemes in advance to protect safe sound operation of wind turbines. These are conducive to reduce failure time caused by the fault and increase power generation of the wind turbine generator, as well as increase earnings of wind farms.

In this paper, the Gaussian process metamodel is developed to analyze the system fault of the wind turbine generator for one wind farm. A potential future work is to apply this method to other wind farms to further validate the applicability of the proposed method. Another possible future work is to compare the proposed method with other metamodel-based methods such as the artificial neural network. The comprehensive comparison is needed to study the advantages of different metamodel-based methods on fault diagnosis for the wind turbine generator.

Data Availability

The “Fault data of 100 wind turbine generators with 2 MW unit capacity in this wind farm from 2015 to 2017” used to support the findings of this study are included within the supplementary information file.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by Self-Financing Project of Hebei Provincial Science and Technology Plan (16214319) and the National Natural Science Foundation of China (Grant no. 71804108).

Supplementary Materials

The supplementary material is the fault data, which are selected between April 2015 and March 2017, of 100 wind turbine generators with 2 MW unit capacity in a wind farm in eastern coastal plain in China. The fault data include failure name, deviation of sensor data, probability of deviation, failure frequency, failure time, loss of electricity, failure reason, and so on. (*Supplementary Materials*)

References

- [1] F. Blaabjerg and K. Ke Ma, “Future on power electronics for wind turbine systems,” *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 3, pp. 139–152, 2013.
- [2] P. W. Carlin, A. S. Laxson, and E. B. Muljadi, “The history and state of the art of variable-speed wind turbine technology,” *Wind Energy*, vol. 6, no. 2, pp. 129–159, 2003.
- [3] J. F. Conroy and R. Watson, “Low-voltage ride-through of a full converter wind turbine with permanent magnet generator,” *IET Renewable Power Generation*, vol. 1, no. 3, pp. 182–189, 2007.
- [4] E. Echavarria, B. Hahn, G. van Bussel et al., “Reliability of wind turbine technology through time,” *Journal of Solar Energy Engineering*, vol. 130, no. 3, Article ID 031005, 2008.
- [5] M. Jureczko, M. Pawlak, and A. Mężyk, “Optimisation of wind turbine blades,” *Journal of Materials Processing Technology*, vol. 167, no. 2-3, pp. 463–471, 2005.
- [6] J. I. Lewis, “Technology acquisition and innovation in the developing world: wind turbine development in China and India,” *Studies in Comparative International Development*, vol. 42, no. 3-4, pp. 208–232, 2007.
- [7] REN21, *Renewables 2019 Global Status Report*, REN21, Paris, France, 2019.
- [8] R. C. H. Cheng and W. Holloand, “Sensitivity of computer simulation experiments to errors in input data,” *Journal of Statistical Computation and Simulation*, vol. 57, no. 1–4, pp. 219–241, 1997.
- [9] R. C. H. Cheng and W. Holland, “Two-point methods for assessing variability in simulation output,” *Journal of Statistical Computation and Simulation*, vol. 60, no. 3, pp. 183–205, 1998.
- [10] J. A. Andrawus, J. Watson, M. Kishk, and A. Adam, “The selection of a suitable maintenance strategy for wind turbines,” *Wind Engineering*, vol. 30, no. 6, pp. 471–486, 2006.
- [11] L. A. Soriano, W. Yu, and J. D. J. Rubio, “Modeling and control of wind turbine,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 982597, 13 pages, 2013.
- [12] B. Lu, Y. Li, X. Wu, and Z. Yang, “A review of recent advances in wind turbine condition monitoring and fault diagnosis,” in *Proceedings of the IEEE Power Electronics and Machines in Wind Applications*, pp. 1–7, IEEE, Lincoln, NE, USA, 2009.
- [13] A. Kusiak and W. Li, “The prediction and diagnosis of wind turbine faults,” *Renewable Energy*, vol. 36, no. 1, pp. 16–23, 2011.
- [14] G.-F. Fan, W.-S. Wang, C. Liu, and H.-Z. Dai, “Wind power prediction based on artificial neural network,” in *Proceedings of the CSEE*, vol. 34, pp. 118–123, CNKI, Beijing, China, 2008.
- [15] G. N. Kariniotakis, G. S. Stavrakakis, and E. F. Nogaret, “Wind power forecasting using advanced neural networks models,” *IEEE Transactions on Energy Conversion*, vol. 11, no. 4, pp. 762–767, 1996.
- [16] S. Yang, W. Li, and C. Wang, “The intelligent fault diagnosis of wind turbine gearbox based on artificial neural network,” in *Proceedings of the International Conference on Condition Monitoring And Diagnosis*, pp. 1327–1330, IEEE, Beijing, China, 2008.
- [17] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, “Short-term wind-power prediction based on wavelet transform-support vector machine and statistic-characteristics analysis,” *IEEE Transactions on Industry Applications*, vol. 48, no. 4, pp. 1136–1141, 2012.
- [18] D. T. Bui, B. Pradhan, O. Lofman et al., “Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and naive bayes models,” *Mathematical*

- Problems in Engineering*, vol. 2012, Article ID 974638, 26 pages, 2012.
- [19] N. Chen, Z. Qian, and X. Meng, "Multistep wind speed forecasting based on wavelet and Gaussian processes," *Mathematical Problems in Engineering*, vol. 2013, Article ID 461983, 8 pages, 2013.
- [20] M. Seeger, "Gaussian processes for machine learning," *International Journal of Neural Systems*, vol. 14, no. 2, pp. 69–106, 2004.
- [21] Z.-H. Xiong, J.-C. Zhang, and H.-H. Shao, "GP-based soft sensor modeling," *Acta Simulata Systematica Sinica*, vol. 17, no. 4, pp. 793–800, 2005.
- [22] B. Chen, P. C. Matthews, and P. J. Tavner, "Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6863–6876, 2013.
- [23] J. Yuan and S. H. Ng, "A sequential approach for stochastic computer model calibration and prediction," *Reliability Engineering & System Safety*, vol. 111, pp. 273–286, 2013.
- [24] R. K. Pandit, D. Infield, and A. Kolios, "Comparison of advanced non-parametric models for wind turbine power curves," *IET Renewable Power Generation*, vol. 13, no. 9, pp. 1503–1510, 2019.
- [25] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*, Springer, New York, NY, USA, 2003.
- [26] R. K. Pandit and D. Infield, "Comparative analysis of Gaussian process power curve models based on different stationary covariance functions for the purpose of improving model accuracy," *Renewable Energy*, vol. 140, pp. 190–202, 2019.
- [27] C. Hultquist, G. Chen, and K. Zhao, "A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests," *Remote Sensing Letters*, vol. 5, no. 8, pp. 723–732, 2014.
- [28] R. K. Pandit and D. Infield, "Using Gaussian process theory for wind turbine power curve analysis with emphasis on the confidence intervals," in *Proceedings of the 6th International Conference on Clean Electrical Power (ICCEP)*, pp. 744–749, IEEE, Santa Margherita Ligure, Italy, 2017.
- [29] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 1, pp. 3–36, 2011.
- [30] Y. Li, Z. Yang, G. Li, and D. Zhao, W. Tian, "Optimal scheduling of an isolated microgrid with battery storage considering load and renewable generation uncertainties," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1565–1575, 2019.
- [31] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian Analysis*, vol. 1, no. 3, pp. 515–534, 2006.

