International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2024): 7.741

Research on Monitoring Method of Wind Turbine Gearbox Operation Status based on SSA-Informer

Yuzhi Wang

North China electric Power University, School of Control and Computer Engineering, Beijing, 102206, China

Abstract: ***In view of the complexity and redundancy of the operating data brought by the harsh environment of the wind turbine, and the difference of the operating state of the wind turbine in the harsh environment. In this paper, singular spectrum analysis is used to decompose the normal operation data with high temperature correlation of the gearbox, and the serial signals of different time scales are obtained. Then, the generation power of the wind turbine is predicted by Informer model, which provides fast and convenient monitoring for the wind turbine gearbox.***

Keywords: wind power, forecasting; deep learning, ensemble learning

1. Introduction

As China's economic development model is shifting to new and clean energy, wind power generation technology can avoid environmental pollution while creating energy. The western and northern parts of Gansu, most of the Qinghai-Tibet Plateau, the Northeast, the central and eastern parts of Inner Mongolia, and parts of Xinjiang are all areas with abundant wind resources, which are conducive to wind power generation [1]. However, with the widespread application of wind power generation, some problems have also been exposed about wind turbines. In practice, wind turbines are generally installed at the draught, such as mountains, wilderness, and the sea, so wind turbines are affected by harsh environments and irregular wind directions[11]. Once there is a problem with the wind turbine, its maintenance difficulty and maintenance cost are relatively large, so it is of practical significance to do a good job of monitoring the status of the wind turbine at ordinary times.

The supervisory control and data acquisition (SCADA) system records many core parameters of wind turbines (such as wind speed, temperature, oil temperature, vibration, etc.)[9-15]. Since the SCADA system has complex and changeable high-latitude data, which can fully reflect the different operating states of wind turbines. It is possible to use SCADA system data for wind turbine status monitoring and fault warning [2]. Literature [3] uses the mutual information method and the long-term short-term memory network to establish a bearing temperature prediction model, and then uses the sliding window method to effectively warn the gearbox bearing temperature. Literature [4] determines the input of the model through the gray relational degree method, then determines the predicted value by training the long short-term memory (LSTM) model, and finally uses the 3 δ criterion to calculate the early warning threshold to realize the monitoring of the gearbox operation status and failure warning. Literature [5] uses gray correlation degree analysis to determine the input variables of the model, and then determines the predicted value by training the long short-term memory (LSTM) model, and then calculates the absolute value of the residual between the actual value and the predicted value, and use the probability distribution fitting method to set the early warning threshold, so as to realize the early warning of wind turbine failure.

Although the above literature has proposed some effective methods, there are still some defects[6], such as failing to fully consider the complexity of the original data and the redundancy between features caused by the harsh environment of the wind turbine, and the operating status of wind turbines varies greatly; the determination of the early warning threshold lacks rationality and flexibility. In view of this shortcoming, this paper uses singular spectrum analysis to process the data with high temperature correlation of wind power gearbox, and obtains sequence signals of different time scales, and then uses the Informer model to predict[7], thereby improving the predictive ability of the model, and providing wind turbine gear Box monitoring provides quick and easy access.

2. Related Work

2.1 Singular Spectrum Analysis

The SCADA data contains a lot of complex information, and the method of empirical mode decomposition to process SCADA data is likely to cause modal aliasing; while the wavelet analysis method is easily interfered by white noise. Singular spectrum analysis is a signal analysis method combining probability theory and mathematical statistics, which is very suitable for the processing of nonlinear time series [5].

Let a signal matrix be $X=[x_1,x_2,\ldots,x_N]$ with length N. Singular spectrum analysis extracts the trend, shock and noise components of the time series through the four steps of embedding, SVD decomposition, grouping and focusing averaging. The specific steps are:

(1) Embed operations. Set the length of the window to L, and satisfy 2≤L≤N, then the trajectory matrix is:

$$Y=\begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} \tag{2.1}$$

In the formula: K=N-L+1, all elements on the diagonal are equal.

(2) Break down operations. Decompose the trajectory matrix into n components, and rewrite the trajectory matrix as follows:

$$Y=\sum_{i=1}^{n}U_i\sqrt{YY_i}V_i \tag{2.2}$$

In the formula, $S=YY^T$; n is equal to the number of positive eigenvalues of S ; U_i is the left eigenvector of S; V_i is the right eigenvector of S .

(3) Group operations. Divide the L components into m disjoint groups $I_1, I_2, ..., I_n$, where $I=(i_1, i_2, ..., i_t)$. Then the matrix Y_i is defined as

$$Y_i = Y_{i1} + Y_{i2} + ... + Y_{it} \tag{2.3}$$

and the trajectory matrix is expressed as:

$$Y = Y_{I1} + Y_{I2} + ... + Y_{It} \tag{2.4}$$

(4) Diagonal averaging operation. Use the following formula to convert the trajectory matrix into a sequence $y_1, y_2, ..., y_n$.

$$y_k \begin{cases} \frac{1}{k}\sum_{m=1}^{k} y^*_{m,k-m+1} & for\ 1 \le k < min(L,K) \\ \frac{1}{min(L,K)}\sum_{m=1}^{min(L,K)} y^*_{m,k-m+1} & for\ min(L,K) \le k \le max(L,K) \\ \frac{1}{N-k+1}\sum_{m=k-max(L,K)+1}^{N-max(L,K)+1} y^*_{m,k-m+1} & for\ max(L,K) < k \le N \end{cases}$$

2.2 Exponentially Weighted Moving Average

Due to the characteristics of wind resources, the operating status of wind turbines varies greatly in practice, and the distribution of residual values in the model is also very divergent, which does not conform to the assumption of normal distribution[8]. The Exponentially Weighted Moving Average (EWMA) is used to determine the RMSE threshold and monitor its trend change. This method is suitable for monitoring the trend change of continuous data [16]. EWMA controls the first statistic Z_t :

$$Z_t = (1-\lambda)R_{et} + \lambda Z_{t-1} \tag{2.5}$$

In the formula: λ is the weight of the historical threshold to the current statistics, λ∈(0,1], in this paper, λ is 0.8; is the RMSE at time t; when t=1, is the oil tank of the wind turbine gearbox The mean value of the RMSE of the temperature residual over a period of time in the normal state.

The mean and variance of the EWMA statistic are:

$$\begin{cases} \mu_{Zt} = \mu_{\mathrm{Re}} \\ \sigma^2_{Zt} = \frac{\sigma^2_{\mathrm{Re}}}{n_s}\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}] \end{cases} \tag{2.6}$$

In the formula: μ_{Re} is the mean value of the RMSE of the gearbox oil tank temperature residual; σ^2_{Re} is the variance of the gearbox oil tank temperature residual over a period of time; n_s is the sampling length.

The EWMA control diagram of the wind turbine gearbox oil tank temperature is based on the threshold value of the gearbox oil tank temperature residual at time t:

$$U(t) = \mu_{\mathrm{Re}} + k\sigma_{\mathrm{Re}}\sqrt{\frac{\lambda[1-(1-\lambda)^{2t}]}{(2-\lambda)n_s}} \tag{2.7}$$

where k is a coefficient.

3. Predictive model building

3.1 Predictive model building

The SSA-Informer model is established according to the feature selection characteristics and parameter nonlinear characteristics. The specific steps of the model implementation are as follows:

(1) Preprocess the SCADA data generated by the wind turbine according to the normal state of the wind turbine. The first step is to eliminate the interference of irrelevant data such as shutdown data and power quality module monitoring data whose active power is less than 0kW. The second step is to conduct a correlation analysis on the SCADA data in the normal operating state, and select parameters with a high correlation with the oil tank temperature of the wind turbine gearbox as input data.

(2) Build a predictive model. The main structure of the prediction model is the SSA module and the Informer module. The singular spectrum analysis is a signal analysis method combining probability theory and mathematical statistics. The number of layers of the Informer module is 2, and finally a fully connected layer with output dimensions of 16 and 1 is added. Select Relu for the activation function of all structures, select the mean absolute error (MAE) as the loss function of the model, and select the Adam optimizer that uses the adaptive learning rate to speed up the convergence of the model.

(3) Train the prediction model. Input the preprocessed SCADA data of the normal operating state of the wind turbine into the prediction model, and then compare the loss of the verification set and the training set to judge whether the model is overfitting or underfitting, and determine the number of iterations.

3.2 Monitoring and early warning model

Based on the above theoretical part, this paper establishes the SSA-ARIMA-Informer model to monitor and warn the temperature of the wind turbine gearbox oil pool. The following figure shows the monitoring of the wind turbine gearbox.

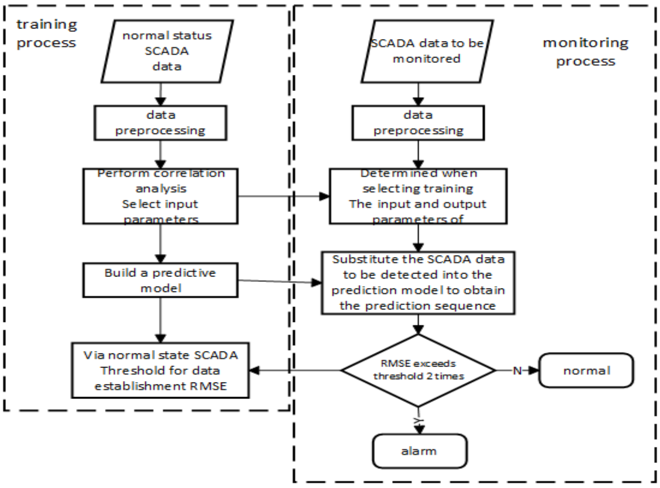

Figure 1: Monitoring and early warning model

4. Experiments

4.1 Data Sources

This paper selects the data collected by the SCADA system with a total of 12960 sampling points in three groups of a wind farm, and collects data every minute. The time span of data #1 collected by the SCADA system is one month, from 0:00:00 on September 1, 2019 to 23:50:00 on September 30, 2019; data #2 and data #2 come from different Gearbox, the time span is one month, the time is from September 1, 2019 0:00:00 to September 30, 2019 23:50:00; data #3 and data #1 come from the same gearbox, the time span It is one month, from 0:00:00 on June 1, 2020 to 23:50:00 on June 30, 2020. In this paper, the first 3456 sample points are selected as the training set data, and the last 864 sample point data are selected as the test set data to verify the performance of the model.

4.2 Data preprocessing and input parameter determination

The SCADA data of the wind turbine is prepared to record all the states of the wind turbine, including normal operation, fault, shutdown and maintenance. Therefore, some irrelevant data will also be recorded in the SCADA data. In order to prevent the irrelevant data from affecting the prediction accuracy of the model, it is necessary to eliminate the abnormal data in the SCADA data.

Table 1: Correlation coefficient between gearbox oil pool temperature and some parameters of wind turbine

Wind Turbine Parameters	correlation coefficient
UNIT.CI_Gearbox Output Shaft Temperature	0.9059
UNIT.CI_Gearbox Inlet Oil Temperature	0.8947
UNIT.CI_Gearbox Input Shaft Temperature	0.8687
UNIT.CI_Main Bearin gGearbox Side Temperature	0.6816
UNIT.CI_Generator Bearing Temperature B	0.6613
SCADAUSE.CI_Rotor Speed	0.6304
SCADAUSE.CI_Pcs Measured Generator Speed	0.6285
UNIT.CI_GeneratorWindingTemperatureV1	0.5006
UNIT.CI_Wind Speed	0.4816
UNIT.CI_Main Bearing Rotor Side Temperature	0.4640

According to Table 1, finally select the gearbox output shaft shaft temperature, gearbox inlet oil temperature, gearbox input shaft shaft temperature, main shaft gearbox side temperature, generator bearing B temperature, impeller speed 1, and generator speed (pcs).

4.3 Statistical Analysis of SSA-Informer Model Prediction Residuals

According to the analysis of the normal operation data of the wind turbine in the previous part, the oil tank temperature of the gearbox of the wind turbine is drawn in Figure 2. It can be seen from the figure that under normal conditions, the temperature of the oil tank of the wind power gearbox varies within 30-80 ° C, and the range of variation is relatively large. Therefore, it may be wrong to directly set the early warning threshold for the oil tank temperature of the wind power gearbox to judge the working status of the wind turbine. Alarm situation. A comprehensive analysis of it with other state parameters of wind turbines will help improve the accuracy of model monitoring and improve the efficiency of wind turbine management.

The SSA-Informer model is used to predict the temperature of the oil tank of the wind turbine gearbox under normal operating conditions. By adjusting the hyperparameters such as the learning rate and the number of iterations, the training loss and verification loss of the training set are obtained, as shown in Figure 3. According to Figure 3, the learning rate is set to 0.0016, and the learning decay rate is 0.001. After multiple trainings, the number of iterations is determined to be 100, and the number of batches is 64. The residual time sequence diagram of the wind power gearbox oil tank temperature prediction is obtained as shown in Figure 4. . It can be seen from Figure 4 that most of the residual data are distributed between -6 and 6° C, but there are also points with large deviations. Therefore, the operating status of the wind turbine cannot be judged based on the abrupt residual. It is necessary to further judge the operation of the wind turbine based on the residual analysis index state.

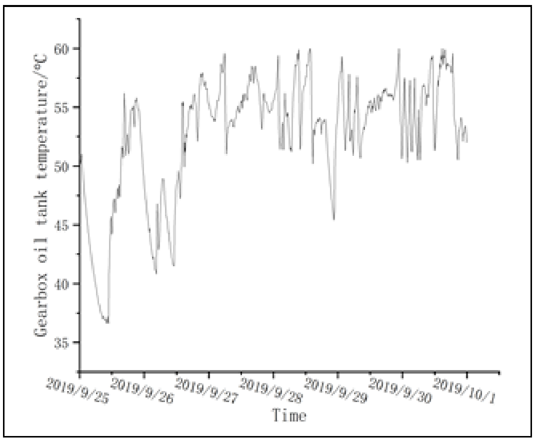

Figure 2: Gearbox oil tank temperature

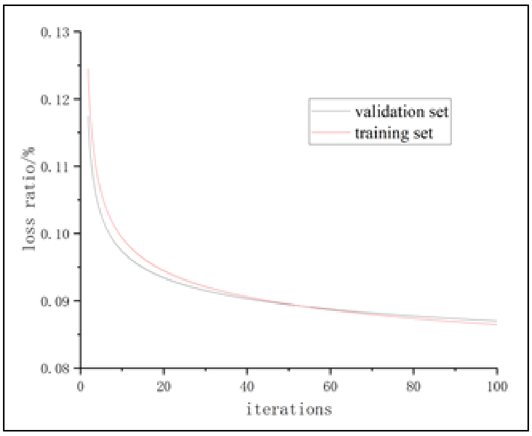

Figure 3: Loss ratio

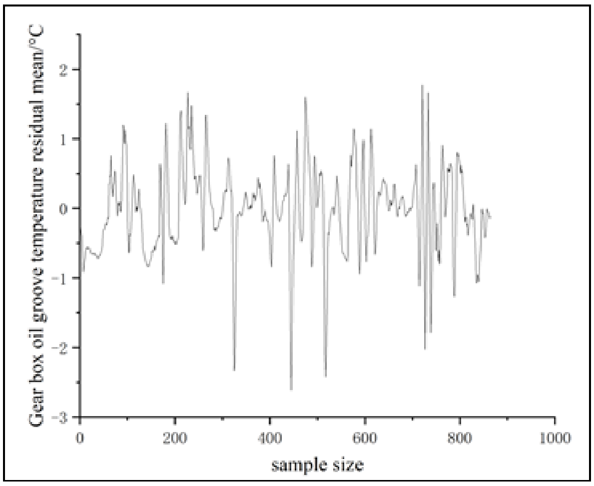

Figure 4: Gearbox oil tank temperature residual mean

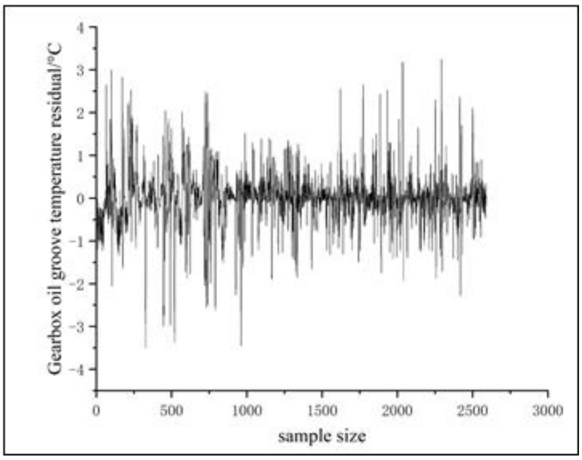

Figure 5: Gearbox grove temperature residual

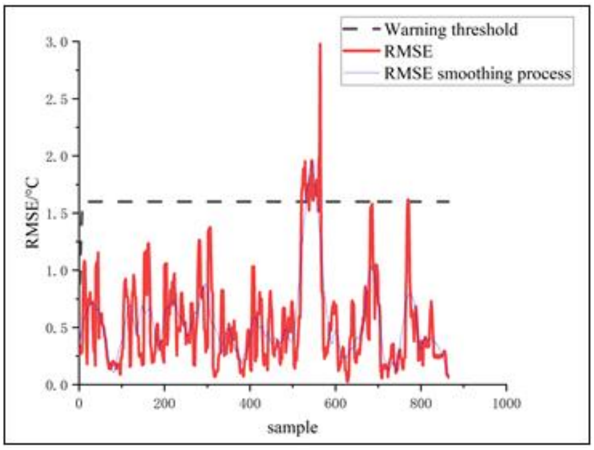

Figure 6: RMSE

4.4 Residual Error Analysis of Abnormal State of Wind Turbine

The SCADA data is screened, and the screened data is input into the prediction model to obtain the predicted value of the target parameters, and the residual sequence of the gearbox bearing temperature is calculated to obtain the mean and variance with a sliding window of 100.

The mean of the normal state fluctuates up and down around 0° C, and the variance of the normal state. It shows that the data volatility is small. It can be seen from the figure that when the wind turbine is in an abnormal state, the variance of the residual error fluctuates greatly. From this, it can be judged that the wind turbine has failed. However, to accurately determine the fault in advance requires further analysis of the residual.

Bring the sample data into the EWMA. After debugging, when the coefficient k=20 is selected, the warning threshold is higher than the RMSE curve.

5. Equations

In this paper, SCADA wind power data is deeply mined and SSA-Informwer model is studied. According to the oil temperature of many parts of the wind turbine, the relationship between wind power and oil temperature of many parts is found. Thus, it can predict the power of the wind turbine under normal conditions, and according to the power of the wind turbine under normal conditions. Taking the predicted normal power of wind turbine as the parameter of early warning of wind turbine can improve the management efficiency of wind turbine and improve the benefit of wind power generation.

References

[1] LIN B Q. China's electric power development achievements during the '13th Five-Year Plan' period and prospects for the '14th Five-Year Plan'［J］. China electric power,2020(12):22-23.
[2] WANG C，LI D Z. Fault early warning of fan gearbox bearing based on LSTM network［J]. Electric power science and engineering，2020，36（9）：40-45.
[3] HUANG R Z，TANG B P，YANG Y N，et al.Condition monitoring of wind turbine gearbox based on LSTM neural network fusing SCADA data［J］.Acta energiae solaris sinica,2021,42(1):235-239.
[4] WEI S R，ZHANG X，FU Y，et al.Early fault warning and diagnosis of offshore wind DFIG based on

GRA-LSTM-Stacking model［J］.Proceedings of the CSEE，2021，41（7）：2373-2382.

[5] Andrew R. Liddle, Information criteria for astrophysical model selection, Monthly Notices of the Royal Astronomical Society: Letters, Volume 377, Issue 1, May 2007, Pages L74–L78, https://doi.org/10.1111/j.1745-3933.2007.00306.x

[6] Akaike, Hirotsugu. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974, 19 (6): 716–723.

[7] YANG G Q，ZHANG K，WANG D Y，et al. Multi-mode fusion ultra-short-term photovoltaic power prediction algorithm based on envelope clustering［J］. Electric power automation equipment，2021，41(2)：39-46.

[8] HU S，XIANG Y，SHEN X D，et al. Wind power prediction model considering meteorological factor and spatial correlation of wind speed［J］. Automation of electric power systems，2021，45（7）：28-36

[9] ANDRADE J R.，BESSA R J. Improving renewable energy forecasting with a grid of numerical weather predictions [J]. IEEE transactions on sustainable energy，2017，8(4)：1571-1580.

[10] U Qian，QIAN Zheng，NIE Zhipeng．Research on filling method of abnormal and missing data of wind turbines[J．Electrical Measurement & Instrumentation，2020，57(23)：1-8(in Chinese)．

[11] ZHU Qianwen，YE Lin，ZHAO Yongning，et al．Methods for elimination and reconstruction of abnormal power data in wind farms[J]．Power System Protection and Control，2015，43(3)：38-45(in Chinese)．

[12] Yalian，LIANG Kunxin，SU Yongxin，et al．Fault warning of wind turbine components based on SCADA data[J]．Wireless Internet Technology，2018，15(13)：122-127(in Chinese)．

[13] ZHAO Yongning，YE Lin，ZHU Qianwen．Characteristics and processing method of abnormal data clusters caused by wind curtailments in wind farms[J]．Automation of Electric Power Systems，2014，38(21)：39-46(in Chinese)．

[14] YAN Jie，ZHANG Hao，LIU Yongqian，et al．Uncertainty estimation for wind energy conversion by probabilistic wind turbine power curve modelling[J]．Applied Energy，2019，239：1356-1370．

[15] ZHANG Yingmei，YANG Genghuang，HAO Xiayi，et al．Research on identification and processing method for abnormal data of residential electric power consumption[C]//Proceedings of the 2019 IEEE 3rd International Electrical and Energy Conference．Beijing：IEEE，2019：900-904．