Research on Wind Deviation Detection Based on DENCLUE Abnormal Working Condition Filtering

Xianlong Yin¹*, Yijie Shi¹, Xinda Xu², Yongchun Duan², Yaowu Jia², Gang Chen¹, Xinyue Zhang¹, Feifei Yin³,a

¹State Power Investment Corporation Research Institute, Beijing 102209, China
²CLP International New Energy Holding Co., Ltd, Shanghai, 200086, China
³Innovation Center for Industrial Big Data, Beijing 102208, China
³yinfeifei@k2data.com.cn

*Corresponding author’s e-mail: yinxianlong@spic.com.cn

Abstract. Aiming at the problem that the wind vane of wind turbines has a deviation to the wind, which damages the power generation efficiency of the unit, a filtering method of abnormal working conditions based on DENCLUE density clustering is proposed, which mainly included two stages of working condition screening and calculation of the angle of the wind deviation. Firstly, the density clustering algorithm based on DENCLUE is used to filter the working conditions of the data and to filter out the working condition data of abnormal power generation. Then, the data is further processed, including data smoothing and wind speed binning. Furthermore, according to the relationship between the output power of the unit and the yaw angle of the wind deviation, a regression model is established to obtain the wind deviation angle of the unit. Finally, the method is verified through the actual operation of SCADA data. The results show that after filtering the working conditions, more stable output can be obtained and the performance of the unit can be obviously improved.

1. Introduction

Wind turbines are mechanical equipment that convert wind energy into electrical energy. Accurate measurement of wind direction is the premise of efficient use of wind energy. The wind vane is usually used to measure the wind direction, and the yaw control system adjusts the wind according to the yaw error to achieve the efficient utilization of wind energy [1]. However, currently, the wind vane installed on the unit has varying degrees of measurement deviation. On the one hand, when leaving the factory, the wind vane deviations are caused by installation and measurement and other human factors; on the other hand, the unit has been operating in a complex and harsh environment for a long time, due to imperfect maintenance in the later stage, the wind direction measurement is inaccurate. Correcting the wind vane to the wind deviation in time can not only improve the utilization rate of wind energy by the unit and increase the electric energy output of the entire wind farm, but also reduce the wear between large components and extend the service life of the unit [2].

At present, relevant scholars have carried out research on the deviation of the wind vane from multiple perspectives such as radar measurement [3, 4], vibration analysis [5], and historical operation data [6, 7]. For example, in reference [3], by installing a lidar anemometer, a wind turbine yaw-to-wind deviation analysis method was proposed to solve the problem of low accuracy of wind caused by the yaw-to-wind deviation. Although using lidar can measure the wind direction more accurately, the cost
of each lidar is more than one hundred thousand yuan, which is not suitable for large-scale use. In reference [5], the wind deviation is calculated by using engine room vibration signal, and the wind deviation is calculated by analyzing and calculating SCADA historical data, and verified it by drawing the wind speed power curve under different vibration conditions. In reference [6], through the analysis of the relationship between power and wind speed in SCADA data, a method for calculating the deviation angle of wind based on quantile regression is proposed, and the corresponding system design scheme is given. Although the above-mentioned methods based on SCADA data mining the feature information of a large amount of data to establish the wind deviation analysis model. However, because the working conditions of the wind turbine are complex and changeable, such as the power limit of the unit, the start of the unit, and the shutdown of the unit, etc., whether these abnormal working conditions can be filtered has a great influence on the further analysis of the wind deviation.

Aiming at the high cost of installing lidar anemometers for units and the complex and changeable operating conditions of units, an analysis method for wind deviation based on abnormal operating conditions of DENCLUE [8] algorithm is proposed. Firstly, the density clustering algorithm based on DENCLUE is used to filter the abnormal conditions of the data. Then the data was further processed, including data smoothing and wind speed binning. Finally, the regression method is used to calculate the wind deviation angle of the unit.

2. Abnormal working condition filtering based on DENCLUE
DENCLUE is an unsupervised clustering method based on kernel density estimation, which improved the performance of the algorithm based on grid cells. The core idea of the algorithm is to use the method of kernel density estimation (KDE) [9] to calculate the influence value between data sets, and to estimate the KDE of this data by superimposing the influence value of one data point with other data. A surface space is formed by the kernel density estimation value of each point. The local maximum value of the surface is called the density attraction point of a cluster. Finally, the abnormal working condition data is filtered through the set threshold. This method filtered the abnormal conditions of the data through unsupervised density clustering, which provides guarantee for the subsequent calculation of the wind deviation angle. The specific implementation process is as follows.

2.1. KDE-based data density estimation
DENCLUE algorithm used the influence function to model the neighborhood data, and the influence function is shown as equation (1).

\[ f(x, y) = \rho e^{-\frac{d^2(x, y)}{2\sigma^2}} \]  

where \( d(x, y) \) is the distance between the two data, which is generally Euclidean distance; \( \rho \) is the data influence amount of the point; \( \sigma \) is the bandwidth of the smooth function, which represents the influence of the point on the surrounding data. The kernel function generally used is the Gaussian function with Euclidean distance.

For the given sample set \( S \) containing \( n \) data points, the density function of any data point is shown in equation (2).

\[ f_{\text{Gauss}}^S(x) = \sum_{i=1}^{n} e^{-\frac{(x-x_i)^2}{2\sigma^2}} \]  

2.2. Density attraction point calculation
The density attraction point can be determined by the Hill climbing method. When the kernel function is continuously differentiable at each sample point, the hill climbing process can be used to determine the density attraction point \( x^* \) as follows:
$$x^0 = x; \quad x^i = x^{i-1} + \delta \frac{\nabla \hat{f}(x^{i-1})}{|\nabla \hat{f}(x^{i-1})|}$$  \hspace{1cm} (3)$$

$$\nabla f(x) = \sum_{i=1}^{n} (x_i - x) e^{-\frac{(x-x_i)^2}{2\sigma^2}}$$  \hspace{1cm} (4)$$

where $d(x, y)$ is the parameter that controls the convergence speed of the algorithm.

In order to avoid the influence of noise points on the calculation of density attraction points, DENCLUE introduces the threshold value $\xi$. If $\hat{f}(x^*) < \xi$, take this point as the noise point and delete it.

2.3. Abnormal working condition filtering based on density attraction point

According to the above, the density function and density attraction points $x^*$ of the set S have been obtained. Define the density attraction points and the points attracted to it as clusters. The data set D whose data density value is less than the threshold $\xi$ is discarded. Data D contains abnormal data of a variety of operating conditions, such as unit power curtailment, unit start-up, turbulence impact, etc. On the one hand, a large number of these data exist in the unit operation data, affecting the accuracy of the results based on the data statistical algorithm; on the other hand, compared with the data of healthy operation, the abnormal condition data only accounts for a small proportion, and its data density is far lower than that of healthy operation data.

The scatter diagram of a wind farm based on density clustering before and after abnormal working conditions were shown in Fig.1. It can be concluded that the data can be deeply cleaned by density clustering based on DENCLUE, and some data generated by abnormal conditions can be filtered out.

![Sample scatter plots before and after processing based on DENCLUE density clustering](image)

**Fig.1** Sample scatter plots before and after processing based on DENCLUE density clustering

3. Wind deviation detection algorithm

3.1. Algorithm principle

Wind turbines are mechanical equipments that convert natural wind energy into electrical energy. During normal operation of the wind turbines, the blades of the wind turbines are blown by the wind to produce mechanical energy, which is then converted into electrical energy by the generator. The power expression of the unit is as shown in equation (5) [10].

$$P = \frac{1}{2} \rho C_p(\lambda, \beta) A v^3 (\cos \phi)^3$$  \hspace{1cm} (5)$$

where...
where $P$ is the output power of the unit; $\rho$ is the atmospheric density; $C_{t}(\lambda, \beta)$ is the wind energy utilization coefficient; $\lambda$ is the tip speed ratio; $\beta$ is the pitch angle; $A$ is the swept area of the wind wheel; $v$ is the wind speed; $\varphi$ is the yaw error.

According to the equation (5), the output power of the unit is proportional to the cube of the cosine of the angle of the wind deviation. As the absolute value of the wind deviation increases gradually, the output power decreases. When the absolute value of the wind deviation angle is $10^\circ$, the output power is reduced by 4.48%; when the absolute value of the wind deviation angle is $15^\circ$, the output power is reduced by 9.87%. Therefore, when the static deviation of the wind vane of the unit is $0^\circ$, the generating power of the unit is the highest. However, when the wind vane was deviated from the wind, the blades will always be unable to align the wind accurately at all times, which seriously damaged the power generation efficiency of the unit. Based on the above theoretical basis, the wind deviation of the unit could be analyzed.

3.2. Data processing

3.2.1. Data smoothing

In order to eliminate data noise, the method of Exponentially Weighted Moving-Average (EWMA) [12] is used for smoothing, which is shown in equation (6).

$$v_i = \beta v_{i-1} + (1-\beta)\theta_i$$  \hspace{1cm} (6)

3.2.2. Wind speed bin

It can be analyzed from the equation (5) that the wind speed is closely related to the power output. In order to obtain the relationship between the wind deviation $\varphi$ and the power accurately, the wind speed is divided into bins, which is completed in the following two steps.

- In order to eliminate the influence of the change of the blade angle on the relationship between the power and the deviation to the wind when the rated wind speed is reached, the data from the cut-in wind speed to the rated wind speed is selected for analysis. At this stage, the output power of the unit is directly related to the input wind energy. Generally, the range of wind speed is $3\text{m/s} \sim 12\text{m/s}$.

- The wind speed data determined in the previous step is processed into bins and divided into several wind speed intervals, which is shown in equation (7).

$$BIN^{(i)} = \{ P^{(i)}, \alpha^{(i)} | v^{(i)} < v < v^{(i+1)} \}$$

\hspace{1cm} (7)

where $v^{(i)}$ is the lower limit of wind speed in the $i$-th interval; $v^{(i+1)}$ is the upper limit of wind speed in the $i$-th interval; the step size of each interval is $0.5\text{m/s}$.

3.2.3. Regression fitting

According to the power output expression formula of the unit, it is known that the third power of the cosine of the wind deviation angle is proportional to the output power. Therefore, $\cos^3 \theta$ is selected to perform regression fitting on the output power in each wind speed bin.

$$P^{(i)} = \lambda^{(i)} \cos^3 (\alpha^{(i)} - \alpha_d^{(i)})$$

\hspace{1cm} (8)

where $\lambda^{(i)}$ is the proportional coefficient of the $j$-th wind speed bin, and $\alpha_d^{(i)}$ is the wind deviation angle of the $j$-th wind speed bin.

The parameters $\lambda^{(i)}$ and $\alpha_d^{(i)}$ of each wind speed bin are calculated by regression fitting with the least square method, where $\alpha_d^{(i)}$ is the wind speed deviation of the wind speed bin. Then, the data of the $M$ wind speed bins are fitted by weight, and finally a wind deviation angle is obtained.
\[ a_0 = \sum_{j=1}^{M} \omega^j a_j \]  
\[ \omega^j = \frac{(v_j')^3}{\sum_{j=1}^{M} (v_j')^3} \]  

where the choice of weight \( \omega^j \) takes into account the nonlinear relationship between wind speed and power.

4. Experimental Verification

4.1. Data Description

The SCADA data of 10 units of a wind farm in Jiangsu are selected to calculate the wind deviation. The time interval of all data is 7s, and the deviation angle is calculated with a period of 15 days. The rated power of all units is 1500kW, the cut-in wind speed is 4m/s, and the rated wind speed is 12m/s.

4.2. Comparison of Stability of Different Schemes

The wind deviation analyzed in this paper belongs to the static deviation of the wind vane, which is relatively stable within a certain period of time. Table 1 compares the stability of the results using DENCLUE anomaly filtering method with those without anomaly filtering method by comparing the standard deviation of the wind deviation over a period of time.

Table 1. Comparison of stability of different schemes.

| Unit number | N001 | N002 | N003 | N004 | N005 | N006 | N007 | N008 | N009 | N010 |
|-------------|------|------|------|------|------|------|------|------|------|------|
| Scheme 1    | 3.28 | 3.24 | 2.56 | 3.74 | 2.67 | 1.76 | 3.37 | 2.93 | 5.05 | 3.83 |
| Scheme 2    | 2.13 | 1.81 | 2.15 | 2.25 | 2.32 | 1.58 | 2.69 | 2.03 | 4.31 | 2.42 |

As can be seen from the above table, the variance of the filtering results based on DENCLUE anomaly conditions is smaller and the results are more stable over a period of time. It is consistent with the fact that the static deviation is relatively stable. It can be proved to a certain extent that the density-based aggregation method can better screen the data conditions and obtain more robust results.

4.3. Result Verification

Based on the analysis method proposed in this paper, the wind deviation angle of 10 units is calculated, and the results are evaluated as normal, slight, and serious according to the deviation angle.

Table 2. Calculation result of wind deviation.

| Unit number | N001 | N002 | N003 | N004 | N005 | N006 | N007 | N008 |
|-------------|------|------|------|------|------|------|------|------|
| Deviation angle \( /^{(\circ)} \) | -2.6 | 12.3 | 4.7  | -5.4 | 2.4  | -9.8 | -2.3 | 1.4  |
| Deviation rating | Normal | Serious | Slight | Slight | Normal | Serious | Normal | Normal |

It can be seen from the above table that most of the units have a normal and slight deviation to the wind, and only N002 and N006 have larger deviation angles. In order to further verify the effectiveness of the method, the wind vanes of the two units were re-calibrated, and after a period of operation, the wind speed and power curves of the two periods before and after were compared.
Fig. 2 Wind speed power curve comparison before and after correction

(a) Unit N002

(b) Unit N006

From the Fig. 2, it can be concluded that after re-calibrating the wind vane, the wind speed power curve had been significantly improved, and the power generation efficiency and power generation of the unit also have been optimized.

5. Conclusion

This paper proposed a regression method based on the DENCLUE algorithm to filter out data density clustering conditions to solve the problem of static wind vane deviation. According to the density of the data, special working conditions such as unit power limit and unit start-up are filtered to improve data quality; then, the preprocessed data is used to construct a regression model to obtain the angle of deviation to wind. With the method in this paper to carry out actual calculations on a wind farm in Jiangsu, through the comparison, it is verified that the proposed method is more stable in a similar period of time. And by recalibrating the wind vane, it is proved that the proposed method can increase the wind energy utilization rate of the unit and electricity revenue.

This paper mainly filtered the operating conditions of the unit from the perspective of data density, but it is necessary to repeatedly adjust the threshold of the method for the features of each unit. In future research, more attention should be paid to the time series relationship of the data, and various operating conditions of the unit should be accurately judged according to the dynamic changes of the time series data for the next model analysis.

Acknowledgments

This study is sponsored by the project “Research on remote centralized monitoring and diagnosis system of wind turbine in wind farm” of China Power International Holding Limited.

References

[1] Bu, F., Huang, W., Hu, Y., Xu, Y., Shi, K., & Wang, Q. (2009, September). Study and implementation of a control algorithm for wind turbine yaw control system. In 2009 World Non-Grid-Connected Wind Power and Energy Conference (pp. 1-5). IEEE.
[2] Jeong, M. S., Kim, S. W., Lee, I., Yoo, S. J., & Park, K. C. (2013). The impact of yaw error on aeroelastic characteristics of a horizontal axis wind turbine blade. Renewable energy, 60, 256-268.
[3] Xiaoyu, W. A. N. G., Tongguang, D. I. N. G., Nianzong, B. A. I., Xin, G. A. O., Bingkun, X. U., & XinJian, F. E. N. G. (2019). Error Analysis and Correction Method of Wind Direction Meter for Wind Turbine. Distributed Energy Resources, 4(6), 57-62.
[4] Weixin, Y. A. N. G., Peng, S. O. N. G., Lei, C. H. E. N., Peng, G. U. O., Yang, C. U. I., Xiaogang, L. I. (2020). Influences of wind turbine yaw's static deviation on power generation performance and the optimization method. Renewable Energy Resources, 38(05), 680-684.
[5] Xin, G. A. O.(2019). Wind Turbine Vibration Data Analysis and Yaw Correction. Journal of Engineering for Thermal Energy and Power, 34(06), 172-177.

[6] Chuang, L. I., Chunhua, T. I. A. N., Jiayang, L. I. U., Pengfei, C. U. I., Wei, J. I. A. N. G. (2020). Research and application on yaw error detection of wind turbine yaw system. Renewable Energy Resources, 38(05), 620-624.

[7] Jinhui, X. U., Peng, G. U. O., Peng, S. O. N. G., Yu, L. I. U.(2019). Analysis and research on static deviation of the yaw system of wind turbines. Huadian Technology, 41(07), 1-5+9.

[8] Hinneburg, A., & Keim, D. A. (1998). An efficient approach to clustering in large multimedia databases with noise.

[9] Terrell, G. R., & Scott, D. W. (1992). Variable kernel density estimation. The Annals of Statistics, 1236-1265.

[10] Pedersen, T. F., Gjerding, S., Enevoldsen, P., Hansen, J. K., & Jørgensen, H. K. (2002). Wind turbine power performance verification in complex terrain and wind farms.

[11] Hunter, J. S. (1986). The exponentially weighted moving average. Journal of quality technology, 18(4), 203-210.