Study and Application of Wind Turbine Power Loss Assessment based on GD-BIN Algorithm

Yangyan Zhan*, Minqiang Zhou, Danyuan Ren
Zhejiang Windey Co., Ltd, Key Laboratory of Wind Power Technology of Zhejiang Province, Hangzhou, 310012, China

*Corresponding author e-mail: zhanyy@chinawindey.com

Abstract. In order to accurately assess the lost production of each wind turbine, a fusion method, which is the combination of Gaussian Distribution and bins of method (GD-BIN), was established. Firstly, complete the first cleaning of operation data according to the operation characteristics and status logs of the unit. Further, GD-BIN algorithm is employed to establish the Wind Power Curve Model (WPCM), which can reflect the real performance of each unit. Thus, the operation data of the unit is delivered to WPCM to calculate theoretical power and lost power at each statistical moment. To verify the accuracy, it is applied in a wind farm in Northeast China and is compared to the other three current methods in the industry. Result shows that the proposed algorithm could effectively estimate lost production with a high accuracy.

1. Introduction
The mapping relationship between wind speed and output power of wind turbine, called WPCM is an important index. Not only can it reflect the actual output performance of each unit, but also provide important data support for analyzing a series problems, such as the design and location of unit, the evaluation of real performance, the lost production's evaluation, etc [1]. However, Influenced by terrain, air density, turbulence and other factors, the WPCM of unit in different wind farms are different, and the WPCM between different units in same wind farm are also different [2], especially in complex terrain condition, the differences between units are greater.

Numerous approaches have been applied to model WPCM, such maximum value method, bins of method, maximum probability method, box method, etc. Based on Supervisory Control And Data Acquisition (SCADA) data, Wanjie [3] studied the performance of unit, the fluctuation and intermittence of wind speed, and established power performance model and lost production evaluation model, which is due to wind power curtailment. Jiang Wenling [4] proposed a method to calculate the lost energy of wind power curtailment with Nacelle Wind Speed (NWS). Song Yingwei [5] corrected the wind speed with power coefficient function Cp, and introduced it into the wind energy conversion formula to calculate the theoretical output power, and this method is so dependent on measured wind speed that the accuracy of calculation is low. Wang Zheng [6] analyzed the historical data of wind farm, and proposed a non-parametric regression theoretical power calculation model without complete dispatching control instructions and operation records.

In order to accurately assess lost production, a hybrid assessment model of each unit which based on GD-BIN is presented. The statistics of operational data from SCADA are used to model WPCM.
Delivered the actual running data of each unit into WPCM, the quantitative assessment of the unit's lost production id realized.

2. Study on calculation of lost production

2.1. WPCM based on GD-BIN algorithm.

According to Lyapunov's central limit theorem, the wind power distribution obeys the normal distribution [7] with enough sample data in each bin. Therefore, the proposed algorithm applied GD in each bin to remove the influence of abnormal outliers and applied bins of method to model WPCM, which can better reflect output performance of each unit in normal operation and realize quantitative assessment the differences between any two units no matter how complex the operation environment of the units are. And with WPCM and operation data of each unit, we can calculate theoretical power under specific operating condition of each unit. The steps of the proposed algorithm are as follows:

Step1: Collect and preprocess the original data. Original operation data shall be collected continuously and include: status logs, wind speed, power, air temperature, air pressure, etc. And it need to be preprocessed to 10min statistical by downsampling. The 10min statistical sets include: maximum value, minimum value, mean value, standard deviation and combination of unit status;

Step2: Abnormal operation data cleaning: To ensure that only normal operation data sets are used to build WPCM, the 10min statistical data contain below states need to be eliminated, such as beyond the operation conditions, degradation of measuring equipment, start-stop operation (including fault stop, operation and maintenance, etc.) ;

Step3: Model WPCM based on GD-BIN: the data sets obtained in step 2 is divided into N bins with a width of 0.5m/s according to wind speed value. And GD is adopted to power in each bin to eliminate the outliers, then use BIN to calculate the characteristic value to build model.

Taking bin \( j \) as an example, the specific method is as following:

a) Calculate statistical characteristic values in bin \( j \), include: mean values of wind speed \( V_{\text{mean},j} \), mean values of power \( P_{\text{mean},j} \), standard deviation of power \( P_{\text{std},j} \). If the number of available sampled data in bin \( j \) is more than 30 min, step b) shall be carried out; otherwise, step c);

b) Apply GD to each bin to complete secondary cleaning of abnormal values in data, and only 95% data are used to calculate theoretical power value in bin \( j \). Therefore, the power within \( \left(P_{\text{mean},j} \pm 1.96 \times P_{\text{std},j}\right) \) are selected to calculate the mean power \( P'_{\text{mean},j} \), and the power out of range are considered as outliers, which will not be took into calculation;

c) If the number of available data in bin \( j \) is less than 30min, calculate the median value of power \( P_{\text{mid},j} \) as the representative value of theoretical power in bin \( j \), and record it as \( P'_{\text{mean},j} \);

Basing on data sets \( \left\{ \left(V_{\text{mean},j}, P'_{\text{mean},j} \right), j = 1, ..., N \right\} \) each unit’s WPCM will be fitted. With the obtained WPCM and operation data at each statistical time, we can estimate theoretical power \( P_r \), which multiply with sampling time is the theoretical production in statistical time.

2.2. Calculation of Lost Production.

With theoretical power \( P_t \) in previous section and the operation power \( P_r \) at each computing cycle, lost production \( E_{LT} \) will be calculated as the following equation (1):

\[
E_{LT} = \sum_{i=1}^{T} \left( P_t - P_r \right) \times dt
\]
where, $P_i$ and $P_{iP}$ is operation power and theoretical power of each unit at moment $i$; $dt$ is the computation cycle, in this paper, it is 10min; $T$ is the Stat Period; $E_{LT}$ is the lost production in Stat Period $T$.

3. **Case Verification**

To demonstrate the accuracy of the proposed method, it is implemented to a wind farm with its operation data in northeast China, and there are 33 units with a 88m diameter and 1500kW capacity.

3.1. **WPCM Verification**

GD-BIN is applied to all 33 units' operation data and WPCM of each unit will be get. In the implementation phase, comparing the differences between original data and data after proposed method. Took D19 as example, and the comparison is shown in Fig.1 and Fig2.

![Fig. 1 Comparison between original data and data processed with GD-BIN of D19](image)

The blue points in Fig.1 are operation power of D19. In left figure, points in orange cycle indicate that D19 has grid constrain, points in green cycle means that D19 in a bad performance, and points in red cycle indicate that D19 is at a state of start-stop, which are all classified as abnormal data. The proposed method is used to process operation data and the result is shown in right figure. Comparing the original data and processed data in Fig.1, we can know that the outlier operation data are almost removed and only normal data are left to model WPCM of D19.

3.2. **Quantitative assessment of Lost production**

To illustrate the accuracy, the proposed method is compared with the other three comment methods in the industry, include: Benchmarking units(record as BM), Mean Values of whole wind farm(record as MV) and Theoretical Simulation of unit (record as TS). The comparison of the four methods, which uses the error between actual power and theoretical power, is shown in Fig.2. Apparently, GD-BIN kernel performs best, TS second, and BG worst.
Fig. 2 Power deviation comparison of four methods

The mean absolute error (MAE), the root mean square error (RMSE) and correlation coefficient (R²) are used frequently in quantitative evaluation of methods. The three metrics presented in equation (2)-(4) will be used to evaluate the performance of the four models in Fig.2 (see Table 1).

\[
RMSE(P, \hat{P}_i) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - \hat{p}_n)^2} \tag{2}
\]

\[
MAE(P, \hat{P}_i) = \frac{1}{n} \sum_{i=1}^{n} |p_i - \hat{p}_n| \tag{3}
\]

\[
R^2(P, \hat{P}_i) = 1 - \frac{\sum_{i=1}^{n} |p_i - \hat{p}_n|}{\sum_{i=1}^{n} |p_i - \bar{p}_n|} \tag{4}
\]

Where \( p_i \) is the actual power and \( \hat{p}_n \) is the theoretical power.

The error statistics of D19 and the whole wind farm is shown in Table 1 and Table 2.

| Method | Power error | Lost Production[kWh] |
|--------|-------------|----------------------|
|        | RMSE  | MAE  | R²     | Actual Production | Theoretical Production | Relative Error |
| BG     | 179.21 | 128.10 | 0.8601 | 140217          | 160086              | 15.58%        |
| BM     | 177.94 | 128.18 | 0.8601 | 140217          | 158367              | 13.37%        |
| TS     | 80.27  | 57.73  | 0.9759 | 140217          | 151155              | 7.80%         |
| GD-BIN | 45.83  | 34.19  | 0.9908 | 140217          | 143326              | 2.21%         |
Table 2. Error statistics of whole wind farm

| Method | Power error | Lost Production[MWh] | Actual Production | Theoretical Production | Relative Error |
|--------|-------------|----------------------|-------------------|------------------------|----------------|
|        | RMSE        | MAE                  | R²                |                        |                |
| BG     | 132.69      | 91.52                | 0.8896            | 14738                  | 16294          | 110.56%        |
| BM     | 127.07      | 87.45                | 0.8948            | 14738                  | 16071          | 109.05%        |
| TS     | 75.05       | 52.92                | 0.9692            | 14738                  | 15751          | 106.87%        |
| GD-BIN | 38.67       | 27.63                | 0.9909            | 14738                  | 15034          | 102.01%        |

The error statistics of D19 in Table 1 illustrates that GD-BIN has the smallest value of RMSE and MAE, nearly 1/2 of TS, and 1/4 of BM and BG. This means the error of lost production with BG and BM is 4 times bigger than proposed method, and error of TS is also 2 times bigger. Comparing actual production and theoretical production, the ranking of lost production is GD-BIN>TS>BG>BM, and the value of R² is bigger than 0.99, and the error of lost production is also the most smallest than the other three methods. The proposed method is applied to all 33 units in the wind farm and evaluate the mean error statistics of the whole farm, and the results is shown in Table 2. From Table 2, we can get the accuracy ranking of 4 methods is GD-BIN>TS>BG>BM, which is same as single D19. Therefore, the proposed method in the paper can be effectively used to calculate the lost production, and the calculation accuracy is high.

4. Conclusion
The paper proposed a fusion GD-BIN model by combing Gaussian Distribution and bins of method, which applied to assess the lost production of each unit, and it was verified to be effective by operation data of each unit. For one thing, the model provide a new method can compare and analysis the comprehensive performance of unit, not only different units of the wind farm, but also the units of different wind farms, and it is applicable for wind farms with complex terrain. For another, the algorithm can realize the quantitative evaluation of the lost production of each unit which provide the theoretical basis for the subsequent improvement.

Acknowledgments
This work was financially supported by Key research and development (R&D) projects in Zhejiang Province (2019C01050).

References
[1] Qi Xia, An Lei. Comparative study of multiple power curve modeling methods [J], 2018, 36 (04): 580-585.
[2] WANG Mingjun, GAO Yuansheng. An analysis of influence factors on wind turbine’s actual power curve [J]. Wind Energy, 2013 (4): 74-79.
[3] Wang Jie. Wind Turbines Running Condition Assessment Method Researched by SCADA Data [D]. North China Electric Power University,2014.
[4] JIANG Wenling, Feng Shuanglei, Sun Yong. Study on Energy Loss Calculation During Wind Power Curtailment Based on Wind Speed Measured by Turbine Nacelle Anemometers [J]. Power System Technology,2014,38 (03):647-652.
[5] Song Yingwei, Wan Jie, Liu Yan, et al. Wind power integration curtailment electricity assessment based on improved wind speed [J]. Electrotechnical Application, 2014, 33 (23):42-45.
[6] Wang Zheng, Liu Chun, Feng Shuanglei, et al. The Wind Farm Theoretical Power Calculation Method Research Based on Non-Parameter Regression [J]. Power System Technology,2015,39 (8): 2148-2153.
[7] Liu Huaxin, Yuan Yiming, Zhou Pei, et al. Normal cloud model of wind turbine state evaluation based on Fusion Theory. Acta Energiae Solaris Sinica, 2018, 39 (10):2891-2900.