Research on the Deviation Analysis of Utilization Hours in Wind Power Plants Based on Regulatory Data Fusion and Feature Extraction & Processing Technologies

CHENG Xiao¹, LI Ping¹, SUN Li-shi¹, WANG Bing², CONG Lin², YU Yang³*

¹State Grid Electric Power Company, Beijing 100031, China;
²State Grid Integrated Energy Service Group Co. Ltd, Beijing 100052, China;
³Beijing Kedong Electric Power Control System Co. Ltd

*yuyang7@sgepri.sgcc.com.cn

Abstract: The continuous elevation of wind power consumption pressure has proposed higher requirements for the technical means, which comprehensively support the wind power consumption. The power generation level of a wind power plant is measured through the utilization hours at present, but the factors influencing the utilization hours are scarcely analyzed and quantified. In order to accurately figure out the reasons for the different utilization hours of wind power plants located in the same area, the total power prediction, power plant data, standalone data, resources, etc. were summarized, and the contribution degrees of five dimensions—resources, equipment availability, AGC command balance degree, AGC command following performance, and power generation capacity—to the utilization hours were analyzed and quantified through the data fusion and feature extraction, in an effort to find the main factors influencing the utilization hours in different power plants. The customized measures were proposed according to the conclusions in order to increase the utilization hours for the power plants. The derived computing method was applied to the wind power plant data acquired by a company in a province, followed by the difference analysis of the utilization hours. It was found that in Mulans sectional area, the reason for partially low utilization hours in Jifeng and Xunfeng Wind Power Plants was resource factor, that for Daheishan and Songshan Wind Power Plants lied in equipment availability, and that for Jufeng Wind Power Plant was possibly equipment following performance or equipment power generation capacity. In addition, the AGC command balance degrees in all target power plants were at low levels.

1. Introduction
According to “January-June national statistical data of electric power industry released by National Energy Administration”, the national accumulative installed capacity of 6,000 KW or above wind power plants reached 192.69 million, newly added installed capacity was 9.09 million kilowatts, and the proportion of wind power installation approached 10.5%. The average number of wind power utilization hours was 1,133 h, which was much lower than that (1,824 h) of power generation equipment. The consumption problem of wind power, the third major installed power source following thermal power and hydropower, has become increasingly severe. Although the average number of utilization hours is relieved somehow, the overall energy loss rate is still presenting a rising tendency with the increase of installed capacity, along with continuous growth of clean energy...
consumption pressure, thus imposing higher requirements for the technical means, which comprehensively support the clean energy consumption. The equipment utilization level of new energy power plants has always aroused high attention within the industry, and the number of utilization hours of power generation equipment is rightly a critical index. Therefore, it is essential to analyze the reasons for the high or low utilization hours of wind power plants, and find the main influence factors, so as to take pertinent measures and increase their utilization hours.

Under normal circumstances, the power generation levels of power plants will present a certain regional regularity [1]. For instance, the number of utilization hours of power generation equipment tends to be high in the area with electricity shortage. However, the utilization hours vary from power plant to power plant in the same area, and even the difference may be different. From the angle of electric energy production, the primary influence factor of utilization hours may be inferred as power rationing [2], followed by resource factor, AGC command, power generation capacity of equipment itself, etc. However, how to quantify the contribution degrees of the factors to utilization hours, and further infer the main reasons for the high/low utilization hours remains to be further probed. Given this, the quantification method for the influence factors of utilization hours in wind power plants is mainly discussed based on the multi-source data fusion and feature extraction technologies.

2. Overview
As an important index in the new energy operation, the utilization level of new energy sources has aroused high attention from the energy supervision department, state grid corporation, new energy power generation enterprises, and all sectors of society, especially in 2018, National Development and Reform Commission and National Energy Administration printed and distributed A Notice of Clean Energy Consumption Action Plan (2018-2020), where they proposed the index requirements for the new energy consumption, which rapidly elevated the importance and urgency of accurate mastery of new energy utilization level.

The number of utilization hours is a significant index used to measure the power generation level of a wind power plant. As the wind power generation is restricted by wind power consumption, the electric energy production does not grow synchronously with the installed capacity. From 2014 to 2016, the utilization hours were declining year by year, and the regional differences were apparent [3]. Therefore, it is necessary to conduct a detailed analysis of the reasons for the differences between wind power plants, between sections, and between regions in the aspect of utilization hours.

The influences of five dimensions—resources, equipment availability, AGC command balance degree, AGC command following performance, and power generation capacity—on the utilization hours in wind power plants were analyzed. The following measurement indexes were proposed: average wind velocity, average availability, AGC command ratio, AGC command following performance ratio, and power generation capacity percentage of the power plants, and the mean values of the sections were taken as the reference values to analyze the contribution degrees of the factors to the utilization hours.

3. Data Fusion and Feature Extraction
1) Data fusion
When it comes to multi-source data fusion, the computer technology is used to automatically analyze, synthesize, and dominant the multi-source information under certain criteria, the consistent explanation and description of the measurement objects are obtained to complete the required tasks, so the system can acquire better performance than its constituent parts.

The electric power data involves five links: generation, transmission, transformation, distribution, and utilization, where the power grid regulation data derives from various multi-source heterogeneous systems such as OMS, EMS, and PMS, each of which stores mass data with different features [4]. The data are mainly divided into model data and operating data, including equipment parameters, measured data, operating events, statistical data, etc., which are classified through labels,
and managed and saved by centering on the power grid and equipment, etc. The acquired data can be automatically associated according to different features, e.g. domain thesaurus and synonym thesaurus, followed by structured storage.

For analyzing the indexes used to quantify the influence factors of wind power utilization hours, the data from dimensions of power plant and standalone device should be acquired. The data required by the dimension of power plant, and those by the dimension of standalone device, are shown in Table 1 and Table 2, respectively.

| Data item               | Data granularity | Data source                                      |
|------------------------|------------------|-------------------------------------------------|
| Power plant ID         |                  | OMS, wind field information                     |
| Name of power plant    |                  | OMS, wind field information                     |
| Installed capacity     |                  | OMS, wind field information                     |
| Electric power generation |              | OMS, reporting of electric quantity             |
| Power prediction       | 15 min           | OMS, power prediction                           |
| AGC data               | 1 min            | EMS, AGC                                       |

Table 2. Turbine dimension required data

| Data item                                      | Data granularity | Data source                                      |
|-----------------------------------------------|------------------|-------------------------------------------------|
| Name and serial number of standalone device   |                  | OMS, standalone information                      |
| Standalone wind velocity                      | 1 min            | Power plant                                     |
| Standalone status                             | 1 min            | Power plant                                     |
| Standalone active power                       | 1 min            | Power plant                                     |

The data are different in time granularity (resolution). For instance, the power prediction data is 15 min, electric energy production data of power plant is day, and standalone data is 1 min. The storage lists of the data are mutually independent, without associated fields. For the sake of unified data analysis of a single power plant, the data need to be associated through the unified power plant ID, and then the required parameter value, intermediate calculated value, and calculated result value are respectively stored according to the time granularity. Meanwhile, the corresponding data can be saved by regions (or sections), power plants, and standalone dimensions. For the data of the same dimension, different saved values should be divided through the measurement type, e.g. standalone data, and the measurement types include standalone wind velocity (0), standalone active power (1), and standalone status (2).

After being associated, the data should be verified according to different rules. For example, the miss rate, invalid rate, and out-of-limit rate of standalone real-time data (active power, wind velocity, etc.) should be verified; the integral electric quantity of active power, and the electric quantity of sample machine reported by power plant should be cross-validated; the electric energy loss of the power plant should be cross-validated with the electric energy loss reported by the power plant through the standalone wind velocity method; the integrity and accuracy tests are required for the power prediction of wind power plants; the integrity rate, out-of-limit rate, tendency check and
correlation check are needed for the anemometer tower data of the power plant. After the data fusion and verification, the applications can be developed through computing and invoking, e.g., generation of theoretical power curve for the draught fan via the nacelle anemometry [4]. After the data like standalone wind velocity, temperature, humidity, atmospheric pressure, and wind direction are fused, the three factors most highly correlated with the power output of draught fan are selected from five factors—wind velocity, wind direction, atmospheric pressure, humidity, and temperature—through the covariance, respectively being wind velocity, atmospheric pressure, and temperature. The maximum thresholds of the three factors are set, and the historical data are successively iterated and screened according to the correlation values (from high to low) of the factors to obtain the corresponding result set. If the result set is empty after a factor is iterated and screened, the domain length of this factor is increased, and the iteration is continued until the result set is nonempty. After the three factors are iterated and screened from high to low, the mean value of the power values ranking in the first 40% in the result set is taken as the standalone theoretical power value within the threshold [5]. The calculation method of standalone theoretical power will be applied to the deviation analysis of utilization hours.

2) Data feature extraction

In this paper, the feature extraction mainly focuses on the selection of text representation model, and feature word selection algorithm.

For the standalone status reported by the new energy power plant, it was found that many data were inaccurate. During the verification process, the standalone status was judged by extracting the key information in dispatching log through the data labels. The information extracted from the dispatching log (Figure 1) are listed in Table 3. The extracted information can be applied to the correction of standalone status, and the data after the governance can be applied to the deviation analysis of utilization hours.

2017-12-05 12:10:43 (TIME) Reported by Chengde ground dispatcher (Zhang Yong): The commands expressing abnormality (FAULT) of two sets of protective channels (NSR and PCS), and device abnormality were sent from two sides of the Muhei Line. Application was submitted for transforming the equipment and 220 kV (VOLT) Muhei line in Datang Daheishan Wind Power Plant (STATION) to hot standby. The dispatcher in Northern Hebei Province already permitted. At 12:40, the equipment (EQUIPMENT) and 220 kV Muhei line (LINE) in Datang Daheishan Wind Power Plant were transformed into hot standby.

![Fig 1. Dispatching log text paragraph](image)

| Time          | Name of power plant | Unit No. | Event     | Property |
|--------------|---------------------|----------|-----------|----------|
| 2017/12/5 12:40:00 | Daheishan             | All      | Power cut | Fault    |

4. Deviation Analysis of Utilization Hours

The influence factors of utilization hours in wind power plants are preliminarily judged, including resource factors, equipment availability, AGC command balance degree, AGC command following performance, and power generation capacity. The eigenvalues of the 5 influence factors will be
hereby calculated. Based on the mean value of the sections, the calculation method for the contribution degree of each factor will be discussed.

1) Resource factors
As an index used to measure the wind power plant resources, the average wind velocity can be acquired by multiple means, including anemometer tower measured wind velocity, numerical weather prediction, meteorological station data, nacelle anemometry, etc. The average wind velocity of each power plant is calculated through the nacelle anemometry. The nacelle wind velocity of standalone draught fan is minute data, assume that there are m draught fans in a power plant, and then the daily average wind velocity of standalone device is as follow:

$$v_{dat} = \frac{v_1 + v_2 + \cdots + v_{1440}}{1440}$$

The daily average wind velocity of the power plant is:

$$v_{das} = \frac{v_{dat1} + v_{dat2} + \cdots + v_{datm}}{m}$$

Assume that there are n power plants within one section, and then the daily average wind velocity in this section is expressed as below:

$$v_{daa} = \frac{v_{das1} + v_{das2} + \cdots + v_{dasn}}{n}$$

The quantities $Q_{das}$ and $Q_{daa}$ of electric energy generated under $v_{das}$ and $v_{daa}$ in each power plant are respectively calculated. If the capacity of a power plant is C, the corresponding contribution to utilization hours is:

$$h_{dv} = \frac{Q_{das} - Q_{daa}}{C}$$

2) Equipment availability
The measurement index—equipment availability—is used to calculate the average availability of a power plant based on the standalone availability \([4]\).

According to standalone status reported by the power plant, the standalone status is minute data, where 0 (wind standby), 1 (power generation), and 2 (derating and power rationing by dispatcher) all belong to operating statuses of draught fan. The minutes of above everyday statuses are accumulated to obtain the standalone operating time, and thus the daily availability of standalone device is:

$$A_{dt} = \frac{\text{Daily operating time of standalone device}}{1440}$$

Daily availability of a power plant is:

$$A_{ds} = \frac{A_{dt1} + A_{dt2} + \cdots + A_{dtm}}{m}$$

Daily availability of a section is:

$$A_{da} = \frac{A_{das1} + A_{das2} + \cdots + A_{dasn}}{n}$$

The corresponding contribution to the utilization hours is:

$$h_{da} = (A_{ds} - A_{da}) \times \frac{Q_{das}}{C}$$

3) AGC command balance degree
It is the fairness of command issuing that is considered in the measurement of AGC command balance degree. The installed capacity, power generation plant reported by power plant, and its actual power output can be selected according to the reference standard, and the index ratios of the power plant are designed as follows:

$$r_1 = \frac{\text{Command}}{\text{Installed capacity}}$$

$$r_2 = \frac{\text{Command}}{\text{Reported power generation plan}}$$

$$r_3 = \frac{\text{Command}}{\text{Actual power output}}$$
If the AGC command issued by a provincial dispatching department is minute data, then the one-day average ratio of one power plant is:

\[ r_{1da} = \frac{\sum_{1}^{1440} r_1}{1440} \]

The one-day average ratio of this section is as below:

\[ R_{1da} = \frac{\sum_{1}^{n} r_{1dan}}{n} \]

By reference to the one-day average ratio of this section, the contribution degree of AGC command balance degree to the utilization hours can be measured via \((r_{1da} - R_{1da})\), and the corresponding daily contribution to the utilization hours is expressed as below:

\[ h_{dr} = (r_{1da} - R_{1da}) \times 24 \]

4) AGC command following performance

In order to measure the following performance of AGC command, namely the ability of actual generated power to follow the command power, the minute following performance index of a power plant is defined as below:

\[ T_{\text{min}} = \frac{\text{Actual power output}}{\text{Command power of the last minute}} \]

Daily following performance of the power plant is:

\[ T_{ds} = \frac{\sum_{1}^{1439} T_{\text{min}}}{1439} \]

Daily following performance of the section is:

\[ T_{da} = \frac{\sum_{1}^{n} T_{dsn}}{n} \]

Based on the daily following performance of the section, the contribution of AGC command following performance to the utilization hours is as follow:

\[ h_{dT} = (T_{ds} - T_{da}) \times \frac{Q_{das}}{C} \]

5) Power generation capacity

To measure the pure power generation performance of a power plant, the influences of other factors should be excluded, e.g. downtime overhaul status. Therefore, the actual daily electric energy production and theoretical daily electric energy production of a power plant are calculated especially under the complete standalone availability (status 1) status, and then the daily power generation capacity of this power plant is:

\[ G_{ds} = \frac{\text{Actual daily electric energy production of the power plant}}{\text{Theoretical daily electric energy production of the power plant}} \]

The daily electric power generation of the section is:

\[ G_{da} = \frac{\sum_{1}^{n} G_{dsn}}{n} \]

By reference to the daily power generation capacity of the section, the contribution of power generation capacity of a power plant to the utilization hours is:

\[ h_{dG} = (G_{ds} - G_{da}) \times \frac{Q_{das}}{C} \]

5. Example Illustration

The difference analysis of utilization hours was implemented with the data acquired by a wind power plant in the control center of a provincial company. The acquired data are shown in Table 4:
Table 4. Some provincial dispatching and control center wind farm utilization hour analysis collecting data

| Data item                          | Data granularity | Data source                              |
|-----------------------------------|------------------|------------------------------------------|
| Name of power plant               | Wind power plant management table |
| Electric energy production of power plant | Day              | Electric quantity table reported by power plant |
| Installed capacity                | Wind power plant management table |
| Standalone wind velocity          | min              | Information table of standalone device |
| Standalone status                 | min              | Information table of standalone device |
| Standalone active power           | min              | Information table of standalone device |
| Command output power              | min              | AGC command table                        |
| Actual power output               | min              | AGC command table                        |
| Power prediction                  | 15 min           | Power prediction table of power plants   |
| Abbreviation of wind power plant  |                  | Comparison table of power plant names    |

During the data fusion process, it was found that the power plant names from different data sources could not completely correspond to each other. Therefore, the data fusion was implemented by extracting the keywords, namely the power plants named presented in different tables were correlated by using the field “abbreviation of wind power plant” in the “comparison table of power plant names”.

Afterwards, the data of a wind power plant within a section from January to June were calculated, including monthly average wind velocity, average availability, AGC command balance degree, AGC command following performance index, and power generation capacity percentage. The monthly mean values within the section were calculated as the reference indexes. The contribution degrees of the influence factors to the accumulative utilization hours were calculated as seen in Table 4-Table 8. As the access to theoretical power of standalone device was not gained, the contribution degrees of some factors to the utilization hours could not be calculated temporarily, and the data in the example illustration was replaced by anomaly percentage. The calculated number of utilization hours in a target power plant was partially lower than the regional mean value.

Table 5. Each wind farm wind resource factor contribution to utilization hour in some section

| Section | Wind power plant | Monthly average wind velocity (m/s) | Anomaly percentage |
|---------|------------------|-------------------------------------|--------------------|
| Mulan   | Jifeng           | 6.52                                | -11.17%            |
| Mulan   | Daheishan        | 7.45                                | 1.15%              |
| Mulan   | Xunfeng          | 6.37                                | -13.23%            |
| Mulan   | Songshan         | 8.18                                | 11.4%              |
| Mulan   | Jufeng           | 7.44                                | 1.37%              |
Table 6. Each wind farm equipment availability factor contribution to utilization hour in some section

| Section | Wind power plant | Monthly average equipment availability | Anomaly percentage |
|---------|------------------|-----------------------------------------|--------------------|
| Mulan   | Jifeng           | 0.9997                                  | 7.94%              |
| Mulan   | Daheishan        | 0.5306                                  | -42.71%            |
| Mulan   | Xunfeng          | 0.9991                                  | 7.88%              |
| Mulan   | Songshan         | 0.8580                                  | -7.38%             |
| Mulan   | Jufen            | 0.9990                                  | 7.87%              |
| Sectional average | | 0.9261 |  |

Table 7. Each wind farm AGC balance degree factor contribution to utilization hour in some section

| Section | Wind power plant | Monthly average AGC ratio | Anomaly percentage |
|---------|------------------|---------------------------|--------------------|
| Mulan   | Jifeng           | 0.6483                    | -2.42%             |
| Mulan   | Daheishan        | 0.6533                    | -1.67%             |
| Mulan   | Xunfeng          | 0.6433                    | -3.18%             |
| Mulan   | Songshan         | 0.6533                    | -1.67%             |
| Mulan   | Jufen            | 0.6583                    | -0.92%             |
| Sectional average | | 0.6644 |  |

According to the calculation results, the main cause for partially low utilization hours in both Jifeng Wind Power Plant and Xunfeng Wind Power Plant was resource factor; the partially low utilization hours in Daheishan Wind Power Plant and Songshan Wind Power Plant were mainly ascribed to equipment availability; AGC command balance degree and other factors decided the partially low utilization hours in Jufeng Wind Power Plant. In addition, the calculated AGC command balance degrees in all target power plants tended to be low.

6. Conclusion

Based on the data fusion and feature extraction technologies, the contributions of five dimensions—resources, equipment availability, AGC command balance degree, AGC command following performance, and power generation capacity—to the utilization hours of wind power plants were quantitatively analyzed, and the calculation formulas for the measurement indexes and contributions of the factors were derived. According to the wind power plant data of the control center in a provincial company, the wind field data of a section from January to June in a year were fused for trial calculation. The results showed that the reasons for partially low utilization hours in Jifeng Wind Power Plant and Xunfeng Wind Power Plant were resource factors, that for Daheishan Wind Power Plant and Songshan Wind Power Plant was equipment availability, and that for Jufeng Wind Power Plant was possibly the equipment following performance or equipment power generation capacity. Furthermore, the AGC command balance degrees of all target power plants were partially low. However, the quantification of contributions of the influence factors to the utilization hours is just a preliminary exploration, and more accurate calculation methods should be sought for in the practical comparisons.
References

[1] Guo L J, Gu Y G. Research on relationship between installed generating capacity and electric energy production. Science Technology and Industry. 2012, 12 (12): 71-72.

[2] Wang M. Research on subsidy gaps and high-range electric energy loss problems in China’s wind power and photovoltaic power generation. International Economic Review [J]. 2018, 4: 71-74.

[3] He H T, Zang N N, Du X W, Ma Y W. Review on utilization hours of electric energy generated under new normal of China’s economy. China Power Enterprise Management [J]. 2018, 02: 39-40.

[4] Tang N, Dai Z S. Application of Electronic Technique. Supplementary issue in 2018 [J]: 44-45.

[5] Yang Z. Model driven research on data cleaning modules. Yunnan Electric Power [J]. 2005:73-74.

[6] Jiang W L. Study on energy loss calculation during wind power curtailment based on wind speed measured by turbine nacelle anemometers. Power System Technology [J]: 2014, 03 (38) (3): 648-650.

[7] Zhou J. Research on performance evaluation of wind turbine based on operation data [D]. 2015: 16-21.

[8] Fan L. Research on wind power efficiency evaluation and its ways of promotion on wind power project [D]. 2017: 21-22.