Dai J, Tan Y, Yang W, Wen L, Shen X.

[Investigation of wind resource characteristics in mountain wind farm using multiple-unit SCADA data in Chenzhou: A case study.](#)

Energy Conversion and Management 2017, 148, 378-393

Copyright:

© 2017. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<https://doi.org/10.1016/j.enconman.2017.05.075>

Date deposited:

28/06/2017

Embargo release date:

13 June 2018

This work is licensed under a

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Investigation of wind resource characteristics in mountain wind farm using multiple-unit SCADA data in Chenzhou: a case study

Juchuan Dai^{a,b*}, Yayi Tan^a, Wenxian Yang^b, Li Wen^a, Xiangbin Shen^c

^a. School of Electromechanical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

^b. School of Marine Science and Technology, Newcastle University, Newcastle upon Tyne NE1 7RU, UK

^c. XEMC Windpower Co., Ltd., Xiangtan 411000, China

Abstract: Wind resource characteristics in a mountain wind farm is more complicated and changeable than that in a flat wind farm due to the influence of complex mountain terrain-topography on air flow. Although there have been a large number of investigations on wind resource characteristics, the local wind resource characteristics in a mountain wind farm are still a lack. To fill this knowledge gap, high-resolution wind data are extracted for investigating from multiple-unit SCADA data, that is, wind data are collected from four selected wind turbines with SCADA system, which means wind data are from four different locations. For each location, more than thirty million sets of wind data are extracted for investigation in 2015. Then, several kinds of probability density functions (PDFs) are compared, and one-dimensional and multidimensional kernel density estimation method is selected for the investigation of the frequency distribution. Data preprocessing methods of both wind speed and wind direction are also presented. Finally, the results of this investigation reveal the specific change characteristics of wind speed and wind direction in the investigated mountain wind farm. Yearly mean wind speeds for four different locations are 3.8m/s, 5.1m/s, 5.4m/s and 5.6m/s, respectively.

Keywords: wind resource assessment; mountain wind farm; SCADA data; frequency distribution

1. Introduction

Wind power has the advantage of rich resources, little environmental pollution, and strong economic competitiveness is being recognized as clean energy sources to withstand environmental damage and avoid future crises[1]. Speeding up the wind power development has become a common understanding to support the Human Development[2, 3]. In recent years, wind power industry continues to develop rapidly. The growth rates of the total wind capacity of the world from 2013 to 2015 for three consecutive years are more than 10%. The main characteristic of wind power is that it is greatly influenced by environmental characteristics, that is, wind resource characteristics. In other words, wind resource characteristics are related to the assessment of energy output on a wind turbine and the whole wind farm, which also have important significance for the location of the wind farm and the layout of the wind turbine. Thus, the accurate and detailed knowledge of wind resource is essential for efficient wind energy conversion and utilization[4].

In recent years, the investigations of wind resource characteristics are mainly focused on the following aspects. (1) Investigations of wind power potential, which is one of the basic indicators of wind resource, have attracted extensive attention. For example, an investigation of wind energy in Turkey showed the total theoretically potential was estimated to be about 88,000 MW[5, 6]; investigation on a provincial scale wind resource in China suggested that an annual wind output of 2000 TWh to 3500 TWh could be reached[7]; in Kingdom of Bahrain, the mean of annual wind power density was found to be 817 W/m² at 60 m height[8]; calculating of offshore wind power in U.S. Atlantic Coast suggested that it could supply an annual mean of 14,087 MW[9]. (2) Investigations of the regional difference of wind resources, which are helpful to grasp the characteristics macroscopically, are also carried out for many regions. For instance, local wind resource characteristics in Turkey showed that the yearly power density mean in the eastern Mediterranean region could reach about 44W/m² [10], but in Keban-Elazig (east region of Turkey), the yearly mean

* Corresponding author.
E-mail address: daijuchuan@163.com

wind power density might be only 15.6 W/m²[11]. (3) Investigations focusing on the rational layout of wind turbines according to the characteristics of wind resources are another hot. For example, an investigation of the wind potential in Greece showed that a 4.5 m diameter wind turbine is sufficient for a pilot solar-wind desalination unit requirement[12] in Mikra-Thessaloniki region; in the coastal region of South Africa, the wind turbine with a hub height of 70 m and a rated power of 1600 kW seems to be a perfect choice [13].

In addition, to obtain more reliable and more abundant information of wind resources, many statistical methods are used in this field. For example, Weibull function is an earlier, popular, and classic function for the frequency distribution analysis of wind speed[14-17]. Considering a positive value of Weibull location parameter causes the unrealistic condition of zero probability, inverse Gaussian distribution was suggested as a possible alternative function[18]. Besides Weibull and inverse Gaussian distributions, there are many other functions, such as Rayleigh[19-22], the maximum entropy principle (MEP) [23, 24], the minimum cross-entropy (MinxEnt) principle[25], hidden Markov models (HMMs) [26], Kappa, Wakeby[27, 28], gamma, lognormal[29], Bayesian [30], and kernel density method[31, 32]. A comprehensive evaluation including gamma, lognormal, inverse Gaussian, MEP, Weibull and Rayleigh distributions were discussed[29]. Parameter calculation method of Weibull function was also investigated, such as the maximum likelihood method[33], the moment, empirical, graphical, and energy pattern factor method[34, 35].

However, almost all the above research doesn't concern the local wind resource characteristic in the same wind farm, especially in a mountain farm. In some countries, such as China, mountain wind farm has been an important source of the wind power output. Thus, it is quite meaningful to fill this knowledge gap through the investigation of wind resource characteristics in a mountain wind farm. On the other hand, the local wind characteristics in a mountain wind farm, are determined by not only the overall weather system, but also the local terrain-topography such as hills, cliffs, and ridges. Thus, wind resource investigation methods applicable to flat wind farms are not applicable. In other words, this requires more wind towers to get local wind resource characteristics. At the same time, it is unrealistic to install wind towers near each wind turbine. So, wind data are extracted from the actual supervisory control and data acquisition (SCADA) system in this paper, that is, wind data are obtained from the anemometer fixed on the nacelle of the turbine. Moreover, different from the traditional wind resource investigation usually with hourly data, high-resolution ratio wind resource data with a one second sampling period will be used. It can provide a more microscopic comparison such as wind speed fluctuations. The remaining part is organized as follows. In Section 2, both environmental condition of the investigated wind farm and data source are introduced, and then five wind resource assessment criteria and the corresponding steps are presented. In Section 3, investigation of wind speed characteristics is carried out, including frequency distribution, mean wind speed, and wind speed fluctuation. Several kinds of probability density functions (PDFs) have been introduced and compared in this section. Also, the specific data preprocessing measures are proposed based on wind speed data characteristic. In Section 4, wind direction characteristics are investigated, including frequency distribution of wind direction and wind direction fluctuation. Both the preprocessing algorithm of wind direction data and fluctuation coefficient definition of wind direction are put forward in this section. Then, a joint investigation of wind speed and wind direction based on multi-dimensional kernel density estimation is presented in Section 5. Finally, Section 6 ends the paper by summarizing the main achievements of the paper and open problems and future issues for further investigation.

2. Data collection and assessment procedure

2.1 Environment descriptions and data collection

This mountain wind farm lies in Chenzhou, which is in the south of China, a subtropical climate zone. The whole

terrain-topography of Chenzhou tilts from the southeast to the northwest, the east is higher than the west, and the south is higher than the north. Several main mountains range from the northeast to the southwest in this region, which blocks and uplifts the cold air from the north, is a barrier to the warm airflow from the southwest. Therefore, apart from the main characteristics of sub-tropical humid climate in this region, there are obvious characteristics of local small climate. The observational data in recent years show that the coldest month is January, with an average temperature of 3 degrees Celsius and the lowest temperature of -8 degrees Celsius. The hottest month is July, with an average temperature of 28 degrees Celsius and the highest temperature of 37 degrees Celsius. As spring comes, the temperature rises rapidly in March and April. After the summer, the temperature drops; from September to December, the monthly reduction of the temperature is more than 5 degrees Celsius. This change in temperature reflects the characteristics of a continental climate. According to the open meteorological data, from 1 January 2011 to 1 September 2016, the south wind blows 836 days, the north wind 834 days, no persistent wind direction 175 days, the east wind 1 day, the west wind 9 days, and other wind direction 157 days. Obviously, the south wind and the north wind prevail, accounting for 83%. The wind less than Force 3 (3.4m/s-5.4m/s) blows 1471 days; the wind of Force 3 blows 418 days; the wind between Force 3 and Force 4 (5.5m/s-7.9m/s) blows 72 days; the wind greater than Force 4 and less than Force 6 blows 6 days. However, it should be pointed out that these data are the general meteorological observation results, not the actual wind force acting on the wind turbine, which is generally installed on the top of the mountain. There are some differences due to the large local airflow difference in mountain wind farm.

Fig.1 shows the specific terrain-topography of the wind farm where wind data are collected from four selected wind turbines. The distance between wind turbine 1 (WT1) and wind turbine 2 (WT2) is about 5900m; the distance between wind turbine 2 (WT2) and wind turbine 3 (WT3) is about 6100m; the distance between wind turbine 3 (WT3) and wind turbine 4 (WT4) is about 3300m. The altitude of WT1 is about 450 meters, WT2 440 meters, WT3 400 meters, and WT4 500 meters. All the selected wind turbines are the same type with exactly the same structure. In this mountain wind farm, supervisory control and data acquisition (SCADA) system has been installed in each turbine. So, the real-time running parameters can be monitored including the wind signals from the anemometer fixed on the nacelle with 80-meter height from the ground.

Fig.1 Terrain-topography of mountain wind farm

Fig.2 shows the wind speed and generator power data from SCADA system in 2.5 hours for a turbine. From the figure, it can be seen that generator power change correspondingly with the change of wind speed. Specifically, the trend of the two is basically the same. It indicates that though there is a difference between the measured wind speed and the actual wind speed, the measured wind speed has the potential to reflect the actual wind speed. Since four wind turbines distribute in different locations, the wind signal can fully reflect the local characteristics of wind resource.

Fig.2 Wind speed, rotor speed and generator power from SCADA system (WT4)

2.2 Assessment criteria and steps

To obtain reliable and complete investigation of wind resource characteristic for mountain wind farm, reasonable assessment criteria are important apart from the use of multiple-unit SCADA data. Furthermore, the selection of assessment criteria should consider the final objectives of the investigation. For wind energy utilization by wind turbines, the following five criteria are employed.

- Frequency distribution of wind speed
- Mean wind speed
- Wind speed fluctuation
- Frequency distribution of wind direction
- Wind direction fluctuation

Frequency distribution of wind speed (probability density function (PDF) of wind speed) is an important also the most common criterion for assessing wind resource. The PDF of a random variable is a model that describes the probability for this variable occurring at a certain point in each observation interval[36]. Thus, it can reflect the consistency of the wind speed at the measured location. If the frequency distribution of wind speed is narrow, it means the consistency is relatively good. To the contrary, it means the wind speed varies greatly. Mean wind speed is another important criterion for assessing wind resource. Combined with the standard deviation, it can reflect the magnitude of wind speed and the amount of variation. Usually, both yearly analysis and monthly analysis are needed, which ask for a long sample data. Apart from the frequency distribution and mean wind speed, a short-time wind speed fluctuation is also significant, which focuses on the size of wind disturbance component. To a turbine, large wind speed fluctuation means a heavy fatigue load. Compared with wind speed distribution, it is a micro assessment.

In terms of wind direction, as long as it is stable no obvious impact acts on wind turbines. Therefore, compared with a long-time frequency distribution of wind direction, the short-time wind direction fluctuation is more important because it will influence the energy capture and the corresponding yaw strategy. For modern large-scale wind turbines, yaw error influences not only the power capture but also the fatigue life of yaw mechanism, because when the yaw angle is larger than a set critical value, yaw mechanism will have to work to adjust wind rotor axis to the wind direction[37, 38].

Fig.3 shows the assessment criteria and steps of wind resource investigation. After the above-mentioned physical

144 meaning analysis for assessment criteria, it is necessary to understand the basic characteristics of wind data. Wind
 145 data (wind speed and wind direction) collected by SCADA system is measured by an anemometer and wind vane
 146 fixed on the nacelle; the sampling frequency is 1Hz. Due to sensing system error, there are some null values in
 147 SCADA data set, also some abnormal values such as the value being greater than or less than 2 times adjacent values.
 148 Then, for different assessment criteria, the needed SCADA data sample length may be different. For instance, one-
 149 year SCADA data is needed for the yearly frequency distribution investigation of wind speed, but one-hour SCADA
 150 data could be enough for wind speed fluctuation investigation. Usually, the extracted SCADA data cannot be directly
 151 used for data calculation. Null values and abnormal values should be removed effectively. Furthermore, since the
 152 sampling frequency is 1Hz, which causes the data amount for the yearly frequency distribution of wind speed is too
 153 large to analyze, 10-minutes mean value would be better. Therefore, data preprocessing is important before data
 154 calculation. For the specific assessment criterion, a reasonable data process algorithm can ensure reliable, accurate
 155 and identifiable results, which asks for a careful consideration on the specific algorithm. With these preparations,
 156 data calculation and result analysis are reliable and accurate, to grasp the characteristics of wind resource.

Fig.3 Assessment criteria and steps

159 3. Investigation of wind speed characteristics

160 3.1 Frequency distribution of wind speed

161 3.1.1 Mathematical methods

162 To date, many kinds of PDFs have the potentiality to be the underlying methods of the frequency distribution of wind
 163 speed. In [Tab.1](#), probability density functions are given for nine kinds of PDFs, namely Weibull distribution (W.pdf),
 164 Beta distribution (B.pdf), Normal distribution (N.pdf), Gamma distribution (G.pdf), Erlang distribution (E.pdf),
 165 Inverse Gaussian distribution (IG.pdf), Rayleigh distribution (R.pdf), the maximum entropy principle distribution
 166 (MEP.pdf), Kernel density estimation (KDE.pdf). Some of them have already been practiced for wind speed
 167 distribution in many sites. To select a reliable and accurate investigating method, the basic introductions on these
 168 PDFs are given in the following[\[29, 39\]](#).

169 W.pdf, named after Waloddi Weibull who described it in 1951, is a continuous probability distribution and has
 170 been employed to describe wind speed distribution in many cases[\[40\]](#). Its distribution function is a stretched
 171 exponential function. Apart from the typical two-parameter W.pdf used to describe unimodal frequency distribution

of wind speed, a two-component mixture W.pdf is also proposed to describe bimodal frequency distribution[41]. B.pdf parametrized by two shape parameters is used to describe the distribution of random variables for intervals of finite length. Using an approximation function for B.pdf parameters, it was used to estimate wind power forecast error[42]. For probability forecasting of wind power generation, it was found to be more suitable for summer than other seasons[43]. N.pdf is a well-known continuous probability distribution and useful due to the central limit theorem. In many cases, random variables obey normal distribution when its number is sufficiently large. Some novel N.pdf-based models have also been proposed and applied in wind power industry, such as square-root normal distribution[39], truncated distribution[44], lognormal distribution[45], Normal Weibull mixture distribution[46]. G.pdf, as a two-parameter continuous probability distribution, is a common type of distribution that is a general case of the two parameters W.pdf [39], can also be transformed into other PDFs such as B.pdf, E.pdf with some conditions[29]. In general, wind regimes are different in different regions. G.pdf has been found to be well adapted to wind speed distribution in southern Sabah[47]. E.pdf is a special case for G.pdf, can also be simplified to the exponential distribution when shape parameter α is equal to one. A case in [48] shows E.pdf could be suitable for long-term wind speed distribution. IG.pdf, also known as Wald distribution, is a two-parameter continuous probability distribution, whose cumulated generating function is the inverse of that of a Gaussian random variable. It was suggested by Bardsley in 1980 to be an alternative in some situations for indicating the use of three-parameter W.pdf with the positive location parameter[18]. R.pdf is a special and simplified case of two-parameter W.pdf in the situation when its shape parameter is assumed to equal to two, and scale parameter is $1/\sqrt{2}$ times of that of W.pdf. It is also widely used in investigating wind speed distribution. Its probability density and cumulative distribution function may be obtained based on the mean value of wind speed, which is a distinct advantage. However, as a simplified case of W.pdf, R.pdf is sometimes poorer than W.pdf in fitting wind resource distribution[20, 21]. MEP.pdf is based on the principle of maximum entropy (if the distribution of discrete random variable X is $P_r(X = x_k) = p_k$, the ‘entropy’ of X is defined as $H(X) = -\sum_{k=1} \log p_k$), suitable for the situation that nothing is known about a distribution except belonging to a certain class. In other words, by changing the parameter number and value (i.e. $\lambda_0, \lambda_1, \dots, \lambda_n$), it has the potential to fit various data distributions[29]. Application of MEP.pdf in wind speed distribution has shown that it agrees better with a variety of wind speed than the conventional W.pdf[23].

KDE.pdf, also known as Parzen method, is a non-parametric method to estimate the PDF of random variables. It can effectively solve the problems of density functions from histograms: (1) the resulting estimate is unsmooth; (2) performance depends on the start and end of the bins. On the other hand, comparing with the conventional parametric method, an attractive advantage of KDE.pdf is the direct use of sample data without the need to estimate characteristic parameters needed in a theoretical distribution. Therefore, it has the potential to overcome the assumption error existing in parametric methods because there is an ineluctable deviation between estimated parameters and actual situations of wind speed in conventional parametric methods[31]. In fact, the probability distribution of wind speed in different sites is very different; it is hard to give a precise estimate. Therefore, KDE.pdf would be a better option to estimate frequency distribution of wind speed. In the following section, KDE.pdf is employed to estimate frequency distribution of wind speed.

Assuming that $X = \{x_i\}$ is the wind speed sample, the PDF $f(x)$ of wind speed can be estimated by the following Kernel density function[49]:

$$\hat{f}(x) = n^{-1} \sum_{i=1}^n h^{-1} K(h^{-1}(x - x_i)) \quad (1)$$

where, $\hat{f}(x)$ is the estimator of the true density function $f(x)$; h is the smoothing parameter called bandwidth for controlling the size of the neighborhood around x ; $K(*)$ is called the kernel function; n is the sample size.

The kernel function $K(*)$ should be symmetric about zero. Apart from this, another three properties in Eq. (2) is

214 also needed.

$$215 \quad \begin{cases} K(u) \geq 0 \\ \int K(u) du = 1 \\ \int u^2 K(u) du > 0 \end{cases} \quad (2)$$

216 Though there are many different kinds of kernel functions available, uniform, Epanechnikov, and Gaussian kernels
217 are the most used kernel functions. Substituting Gaussian kernel $\left(K(u) = \exp(-0.5u^2) / \sqrt{2\pi}\right)$ into Eq. (1), one has

$$218 \quad \hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \exp \left(-\left(\frac{x-x_i}{\sqrt{2}h} \right)^2 \right) \right) \quad (3)$$

219 In Eq. (3), the bandwidth h is a free parameter, which has a strong influence on the resulting estimate. If the
220 bandwidth is too large or too small, it is possible to bring about a large deviation between the estimated density and
221 the true density. A commonly used criterion to select this parameter is based on the mean integrated squared error
222 (MISE), that is

$$223 \quad \min \text{MISE}(h) = E \left\| \hat{f}(x) - f(x) \right\|_2^2 \quad (4)$$

224 where, E indicates the expected value of sample.

225 Since the true density function is not known, Eq. (4) cannot be used directly to calculate the bandwidth parameter.
226 To solve this problem, some bandwidth selection methods have been developed, such as NR bandwidth selection, PI
227 bandwidth selection, SEQ bandwidth selection, CV bandwidth selection, and BT bandwidth selection[50].
228 Corresponding to Eq. (4) deduced from Gaussian kernel, a widely-used bandwidth calculation method is

$$229 \quad h = \left(\frac{4\hat{\sigma}^5}{3n} \right)^{0.2} \quad (5)$$

230 where, $\hat{\sigma}$ is the standard deviation of the wind speed sample.

231 Tab.1 Probability density function for wind speed distribution

Distribution	Probability density function	Parameter explanation
Weibull distribution (W.pdf)	$\frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha-1} \exp \left(-\left(\frac{x}{\beta} \right)^{\alpha} \right)$	α , shape parameter; β , scale parameter
Beta distribution (B.pdf)	$\frac{1}{B(\alpha_1, \alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1}$	(α_1, α_2) , shape parameter
Normal distribution (N.pdf)	$\frac{1}{\sqrt{2\sigma^2\pi}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$	μ , mean; σ , standard deviation
Gamma distribution (G.pdf)	$\frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp(-x/\beta)$	α , shape parameter; β , scale parameter
Erlang distribution (E.pdf)	$\frac{\lambda^{\alpha} x^{\alpha-1}}{(\alpha-1)!} \exp(-\lambda x)$	α , shape parameter; λ , rate parameter
Inverse Gaussian distribution (IG.pdf)	$\left(\frac{\alpha}{2\pi x^3} \right)^{0.5} \exp \left(-\frac{\alpha(x-\mu)^2}{2\mu^2 x} \right)$	α , shape parameter; μ , mean
Rayleigh distribution (R.pdf)	$\frac{x}{\beta^2} \exp \left(-\frac{x^2}{2\beta^2} \right)$	β , scale parameter
MEP distribution (MEP.pdf)	$\exp \left(\lambda_0 + \sum_{i=1}^n (\lambda_i x^i) \right)$	λ_i , Lagrangian multipliers
Kernel density estimation (KDE.pdf)	$\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x-x_i}{h} \right)$	h , bandwidth

232 3.1.2 Comparison of calculation result

233 As explained above, by investigating yearly and seasonal frequency distribution of wind speed, the basic wind
234 speed distribution characteristic in the wind farm could be obtained. However, the data amount based on one-second

sample frequency is rather difficult to process since it is more than thirty million sets for four wind turbines in one year. It would be better to use wind speed mean in a period, for example, a 10-minute wind speed mean. There are two main reasons for choosing a mean of 10 minutes. (1) There is a data packet every 10 minutes for the investigated SCADA data, thus it is relatively convenient to process 10-minute data. (2) In the actual wind turbine control, 10 minutes is also a common control observation unit, such as yaw control. But, due to sensing system error, some null and abnormal data may exist in wind speed sample. As introduced in Section 2.2, data preprocessing is necessary to improve the data reliability. Also, abnormal data should include the date being greater than or less than 2 times adjacent data.

Eq. (6) shows a specific data preprocessing strategy: (1) smoothing abnormal data, (2) eliminating null data, (3) data mean calculating.

$$\begin{cases} x_i = 0, & \text{if } x_i \text{ is null} \\ x_i = \frac{x_{i-1} + x_{i+1}}{2}, & \text{for } x_i \geq 2x_{i-1} \ \& \ x_i \geq 2x_{i+1} \\ x_i = x_{i+1}, n = n-1, & \text{for } x_i \leq 0; \ i = 1, 2, 3, \dots, n \\ \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \end{cases} \quad (6)$$

where, $\bar{x}$ is the mean of the wind speed sample.

The theoretical item number n for a 10-minute wind speed sample is equal to 600, but after processing by Eq. (6), it may be smaller.

Fig.4 shows the yearly frequency distribution of wind speeds of the four wind turbines. Both the KDE curves and frequency histograms are given. The differences between KDE curves for different wind turbines are obvious, wind speeds corresponding to the maximum probability density for WT1, WT2, WT3 and WT4 are 2.2m/s, 1.9m/s, 4.3m/s and 6.6m/s, respectively. WT1 has the maximum probability density of 0.224 which means it has a more concentrated frequency distribution of wind speed than others. Though WT2 has the minimum probability density peak, it has two probability density peaks, which means it has a more uniform density distribution than others. Relatively, the difference between WT3 and WT4 is small. For the aim of wind power generation, frequency distributions of wind speed in WT3 and WT4 are better than WT1 and WT2.

Fig.4 Yearly frequency distribution of wind speed

Seasonal frequency distribution of wind speed is illustrated in Fig.5. Density distribution curves in Figs.5(a), (b), (c) and (d) are corresponding to WT1, WT2, WT3, and WT4, respectively. Data for spring is from February, March, and April; summer from May, June, and July; autumn from August, September, and October; winter from November, December and January. Probability density curves in different seasons are also different and vary in different wind turbines. For WT1 and WT2, the maximum wind speed corresponding to probability density peak appears in winter, but it appears in spring for WT3 and in summer for WT4. For WT1, summer has a more concentrated frequency distribution of wind speed, its wind speed corresponding to the probability density peak is 2.3m/s, which is 5m/s in winter. Similar probability density distributions are in spring and autumn. For WT2, wind speed corresponding to probability density peak are basically consistent in the vicinity of 2m/s in spring, summer, and autumn; wind speed corresponding to the first probability density peak in winter is also about 2m/s. Furthermore, probability density curves are very similar for spring, summer, and autumn. For WT3, except autumn, probability density distributions are close and probability density peaks are also relatively flat. In terms of wind power generation, the frequency distribution of wind speed in autumn seems to be not as good as that in the other seasons because low wind speed has a greater probability in the distribution curve. For WT4, the difference of probability density distributions in four seasons is relatively small. The minimum wind speed corresponding to probability density peak is about 4.8m/s in autumn and the maximum wind speed is about 7.4m/s in summer and the difference is 2.6m/s, which is 3m/s for WT1, 5.8m/s for WT2, and 2.6m/s for WT3.

Fig.5 Seasonal frequency distribution of wind speed
(a) WT1, (b) WT2, (c) WT3, (d) WT4

3.2 Mean wind speed

In this section, the mean wind speed including yearly, seasonal, monthly, diurnal and hourly results is given. To quantify the amount of variation of wind speed, the corresponding sample standard deviations are also given, which is expressed as

$$s = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}} \quad (7)$$

286 where, s denotes the standard deviation of the sample $\{x_i\}$.

287 Yearly, seasonal and monthly mean wind speed and the corresponding standard deviation are given in Tab.2 based
 288 on SCADA data in 2015. The yearly mean wind speed for WT1 is obviously smaller than that of the other turbines.
 289 Meanwhile, its standard deviation is also smaller. The most significant difference of mean wind speed between
 290 different seasons is in WT2 which is 1.9m/s, the least significant difference is in WT1 which is 0.5m/s. For WT1 and
 291 WT2, spring and winter have larger standard deviation than summer and autumn. For WT3 and WT4, spring and
 292 summer have larger standard deviation than autumn and winter. The maximum mean wind speed appears in
 293 December for WT1 and WT1, April for WT3, and June for WT4. The maximum differences of mean wind speed
 294 between different months are 2.1m/s, 3m/s, 2.9m/s and 1.8m/s for WT1, WT2, WT3, and WT4, respectively. The
 295 maximum standard deviation appears in November for WT1, February for WT2, April for WT3 and WT4.

296 Tab.2 Yearly, seasonal and monthly mean wind speed

Mean (m/s)																	
Year	Season				Months												
	Spr	Sum	Aut	Win	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
WT1	3.8	3.9	3.2	3.6	4.4	4.6	4.0	3.9	3.8	3.1	2.8	3.8	3.2	3.5	4.1	3.8	4.9
WT2	5.1	5.2	4.2	4.7	6.1	6.3	5.5	5.3	5.0	3.9	4.0	4.8	4.0	4.4	5.6	5.1	6.9
WT3	5.4	6.1	5.4	4.5	5.6	5.8	5.4	5.8	7.2	5.5	5.4	5.2	4.3	4.4	4.9	5.0	6.0
WT4	5.6	6.0	6.0	5.0	5.3	5.7	5.6	6.1	6.2	5.5	6.6	5.9	4.9	4.8	5.1	4.8	5.6

Standard Deviation (m/s)																	
Year	Season				Months												
	Spr	Sum	Aut	Win	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
WT1	1.9	2.0	1.7	1.8	2.0	2.0	2.1	1.9	1.9	1.6	1.1	2.0	1.6	1.8	1.9	2.2	1.9
WT2	2.9	3.0	2.5	2.8	3.0	2.9	3.3	3.1	2.9	2.5	2.2	2.7	2.5	2.7	2.9	3.1	2.7
WT3	2.3	2.6	2.3	2.0	2.2	2.0	2.4	2.1	2.9	2.8	2.0	2.1	1.8	2.0	2.0	2.3	2.1
WT4	2.4	2.7	2.4	2.0	2.1	2.0	2.7	2.5	3.0	2.7	2.3	2.0	2.0	2.1	1.8	2.2	1.8

297 To investigate diurnal and hourly mean wind speed of four turbines, only data in the selected month is used. The
 298 screening conditions include two: (1) the most similar mean wind, (2) the most similar standard deviation. In other
 299 words, the selected month should have the minimum mean difference and standard deviation difference between four
 300 turbines. By comparing, both September and November are suitable for investigating, in which the mean difference
 301 is 1.3m/s and the standard deviation difference is 0.9m/s. In the following, data in September is used. Diurnal wind
 302 speed mean of four turbines in September is shown in Fig.6(a), and the corresponding standard deviations are in
 303 Fig.6(b). Hourly mean wind speed from 1 September to 3 September is shown in Fig.6(c), and the corresponding
 304 standard deviations are in Fig.6(d). Though diurnal mean wind speeds are different in Fig.6(a), trends of wind speed
 305 of four turbines have similarity. Generally, diurnal mean wind speed for WT1 is smaller than that for other turbines
 306 and its standard deviation is also smaller. From Tab.2, monthly mean wind speed for WT2 is equal to that for WT3
 307

308 but its diurnal variance is larger than that for WT3. The maximum standard deviation appears on 25 September for
 309 WT2. According to hourly mean wind speed and standard deviation, WT1 and WT2 are relatively close, and then
 310 WT3 and WT4 are relatively close except a sudden change in WT3. If this sudden change is regarded as an abnormal
 311 phenomenon and neglected, WT4 has the maximum mean wind speed and standard deviation in the figures.

Fig.6 Diurnal and hourly mean wind speed

3.3 Wind speed fluctuation

316 Wind speed fluctuation, which describes the real-time wind speed variation characteristic with a higher sampling
 317 frequency (for example, 1Hz), is also an important criterion for wind resource assessment because wind speed
 318 fluctuation has a significant effect on the turbine, for example, it will influence the aerodynamic load and further
 319 impact the fatigue life. From another point of view, wind speed fluctuation can be regarded as a microscopic
 320 assessment of wind resource characteristic since SCADA data based on the one-second sample will be employed. To
 321 compare the synchronous difference of wind speed fluctuation of four turbines, a wind speed fluctuation coefficient
 322 presented in [51] is introduced which is expressed as

$$w_s(i) = \begin{cases} \frac{\sum_{k=l+i}^{n+i} |x_k - x_{k+1}|}{\sqrt{\sum_{k=l+i}^{n+i} |x_k^2 - x_{k+1}^2|}}, & \sum_{k=l+i}^{n+i} |x_k^2 - x_{k+1}^2| \neq 0 \\ 0, & \sum_{k=l+i}^{n+i} |x_k^2 - x_{k+1}^2| = 0 \end{cases} \quad (8)$$

324 where, w_s is wind speed fluctuation coefficient; x_k is wind speed data (sampling once a second); $i=1,2,3,\dots$

325 The main advantage of using Eq. (8) is that wind speed fluctuation coefficient could be limited to the range of
 326 $[0,1]$, thus is convenient to evaluate the relative fluctuation of wind speed. According to the equation characteristic,
 327 time continuous data is needed. In other words, to maintain the continuity of data in the time field, zero value should
 328 not be eliminated when investigating wind speed fluctuation. Thus, zero value is preprocessed according to Eq. (9).
 329 Moreover, it should assure that no continuous two or more zero values are included in the selected data set.

330

$$\begin{cases} x_i = 0, & \text{if } x_i \text{ is null} \\ x_i = \frac{x_{i-1} + x_{i+1}}{2}, & \text{for } \begin{cases} x_i \geq 2x_{i-1} \ \& \ x_i \geq 2x_{i+1} \\ \text{or} \\ x_i = 0 \end{cases} \end{cases} \quad (9)$$

331

332

333

334

335

336

337

338

339

340

341

342

343

SCADA data in the 39th hour in Fig.6(c) is employed to analyze wind speed fluctuation in which WT1 and WT2 have the same mean wind speed of 4.8m/s, WT3 and WT4 have the close mean wind speed about 8m/s. Fig.7(a) shows the wind speed for WT1 and WT2, and Fig.7(b) shows the corresponding fluctuation coefficients. Fig.7(c) shows the wind speed for WT3 and WT4, and the corresponding fluctuation coefficient is in Fig.7(d). Both Fig.7 (b) and Fig.7 (d) are under the condition of $n=10$ which used in Eq. (8). This parameter selection is mainly based on the mechanical time constant of wind rotor installed in the farm. The standard deviation of wind speed for WT2 is larger than that for WT1, which is illustrated in Fig.6(d), and wind speed fluctuation coefficient for WT2 is also larger than that for WT1. The maximum wind speed fluctuation coefficient for WT1 arrives at 0.95, and 0.55 for WT2. Both mean wind speed and its standard deviation for WT3 and WT4 are close, but wind speed fluctuation coefficient for WT4 is much larger than that for WT3. The variance ranges of fluctuation coefficient for WT1, WT2, WT3, and WT4 are 0.36, 0.74, 0.45 and 0.84. The means of fluctuation coefficient for WT1, WT2, WT3, and WT4 are 0.36, 0.49, 0.30 and 0.60 in Fig.7. Obviously, except for frequency distribution and mean wind speed, wind speed fluctuation also has significant differences at different sites.

344

345

346

Fig.7 Wind speed fluctuation

347

4. Investigation of wind direction characteristics

348

4.1 Frequency distribution of wind direction

349

350

351

352

In this section, the yearly, seasonal and monthly frequency distributions of wind direction are investigated using 10-minutes mean of wind direction data. For obtaining reliable 10-minutes mean, it is necessary to preprocess the SCADA data of 1Hz sampling frequency. Except for null data, if a datum is 90 degrees larger than its neighbor data, it is also considered an abnormal data, which may be caused by sensing signal error. Moreover, a wind direction of 0

degrees is actually equivalent with that of 360 degrees. Then, assuming $X = \{x_i\}$ is wind direction data sample, a preprocessing method is presented as

$$\text{if } x_i \text{ is null, } x_i = 0$$

$$\text{if } \begin{cases} |x_i - x_{i-1}| > 90^\circ \\ |x_i - x_{i-1} - 360^\circ| > 90^\circ \\ |x_i - x_{i+1}| > 90^\circ \\ |x_i - x_{i+1} - 360^\circ| > 90^\circ \end{cases}, \text{ then } \begin{cases} x_i = \text{mod}(\frac{x_{i-1} + x_{i+1}}{2} + 180^\circ, 360^\circ), & |x_{i-1} - x_{i+1}| > 180^\circ \\ x_i = \frac{x_{i-1} + x_{i+1}}{2}, & |x_{i-1} - x_{i+1}| \leq 180^\circ \end{cases} \quad (10)$$

where, the mod is a remainder operation, $\text{mod}((x_{i-1} + x_{i+1})/2 + 180^\circ, 360^\circ)$ means giving the remainder after the division of $((x_{i-1} + x_{i+1})/2 + 180^\circ)$ by 360° .

In Eq. (10), to maintain the continuity of sampling data, all null values are set to be zero. However, null values should not be involved in the calculation of the average value. Thus, after preprocessing wind direction data using Eq. (10), all null values should be eliminated. If a wind direction datum and the corresponding wind speed datum are equal to zero at the same time, it is judged to be abnormal data and is eliminated. Then, wind direction data of yearly 10-minutes mean could be obtained according to the following equation, that is

$$\begin{cases} X_a = X(0^\circ < x_i < 180^\circ) \\ X_b = X(180^\circ \leq x_i \leq 360^\circ) \\ \bar{x}_a = \frac{1}{n} \sum_{i=1}^n x_i, & x_i \in X_a, n > 0 \\ \bar{x}_b = \frac{1}{m} \sum_{i=1}^m x_i, & x_i \in X_b, m > 0 \\ \bar{x} = \begin{cases} \frac{m(\bar{x}_b - 360^\circ)}{m+n} + \frac{n \cdot \bar{x}_a}{m+n}, & |\bar{x}_a - \bar{x}_b| > 180^\circ \text{ \& } \bar{x} \geq 0 \\ \frac{m(\bar{x}_b - 360^\circ)}{m+n} + \frac{n \cdot \bar{x}_a}{m+n} + 360^\circ, & |\bar{x}_a - \bar{x}_b| > 180^\circ \text{ \& } \bar{x} < 0 \\ \frac{\bar{x}_a + \bar{x}_b}{2}, & |\bar{x}_a - \bar{x}_b| \leq 180^\circ \\ \bar{x} = \bar{x}_b, & n = 0 \\ \bar{x} = \bar{x}_a, & m = 0 \end{cases} \end{cases} \quad (11)$$

where, n is the sample size of X_a , m is the sample size of X_b .

Its frequency distributions are shown in Figs.8 (a), (b), (c), and (d) corresponding to WT1, WT2, WT3, and WT4, respectively. From the figures, the apparent differences of wind direction distribution are observed for different turbines. For WT1, a dominant wind direction is observed which is near 0/360 degrees. Wind direction of 6 degrees has the maximum probability distribution, and wind direction data between 0 degrees and 10 degrees accounted for the total data of 20%. For WT2, there are three distribution peaks of wind direction near 310 degrees, 290 degrees, and 140degrees. Especially, near 310 degrees, the wind direction distribution dominates. Wind data between 305 degrees and 315 degrees accounted for the total data of 19%. For WT3, two dominant wind directions are obvious which are approximately in two opposite directions. One is near the direction of 40 degrees and the other is near the direction of 210 degrees. Wind direction data between 205 degrees and 215 degrees accounted for the total data of 22%. For WT4, wind direction also dominates in two directions, namely near 15 degrees and 175 degrees. Wind direction of 175 degrees has the maximum probability distribution. Wind direction data between 170 degrees and 180 degrees accounted for the total data of 19%. By comparison of the yearly mean of wind direction data, the common point for four turbines is that there are clearly dominant wind directions for four turbines; the different point is the

dominant wind directions are different for different turbines. It shows that the local wind direction characteristics can be significantly changed in mountain wind farm.

Fig.8 Yearly frequency distribution of wind direction

Seasonal frequency distributions of wind direction for four turbines are shown in Fig.9. For the convenience of analysis, wind direction data in spring and summer are arranged in a figure and that in autumn and winter in a figure. From the figures, obvious seasonal differences of the frequency distribution of wind speed can be observed. In general, there is a close frequency distribution in spring and summer, and then another close frequency distribution in autumn and winter. Also, some local differences in different seasons or turbines appear. Specifically, two dominant wind directions near 150 degrees and 0 degrees appear in spring and summer, and a predominant wind direction in near 0 degrees in autumn and winter for WT1. The maximum frequency distribution appears near 0 degrees in four seasons. Obviously, autumn and winter have more stable wind direction than spring and summer. For WT2, three predominant wind directions are observed in four seasons, namely near 140 degrees, 290 degrees and 310 degrees. Wind direction near 310 degrees has the maximum frequency distribution for spring, autumn, and winter, and wind direction near 140 degrees has the maximum frequency distribution for summer. For WT3, two predominant wind directions in two approximately opposite directions appear for all seasons. One is near 40 degrees and the other is near 210 degrees. Wind direction near 210 degrees dominates in spring, autumn, and winter and wind direction near 40 degrees dominates in summer. Especially in winter, wind direction distribution has a rather prominent peak near 210 degrees, and data between 205 degrees and 215 degrees accounted for the total data of 34%. For WT4, like the law in WT3, two predominant wind directions near 15 degrees and 175 degrees appear for all seasons in two approximately opposite directions. In spring, wind directions near 175 degrees and 15 degrees have a close frequency distribution. In summer, wind direction near 15 degrees has a prominent frequency distribution peak, and data between 10 degrees and 20 degrees accounted for the total data of 31%. In autumn and winter, wind direction near 175 degrees has the maximum frequency distribution. Generally, the law of seasonal wind direction distribution between WT3 and WT4 is relatively close. The predominant wind direction may appear an obvious change in different seasons or different turbines.

Fig.9 Seasonal frequency distribution of wind direction

After processing monthly 10-minutes mean wind direction data using KDE method illustrated in Eq. (3), KDE curves of wind direction distribution for four turbines could be obtained and are shown in Fig.10. For WT1, WT3 and WT4, two probability distribution peaks appear even in all months, but for WT2 three probability distribution peaks appear. For WT1, the highest distribution peak appears in October, in December for WT2, in September for WT3 and in June for WT4. In other words, wind distribution is relatively concentrated in these months. The lowest distribution peak appears in Feb for WT1, July for WT2, WT3, and WT4. In addition, wind direction distribution in

June is obviously different with other months for all turbines despite the variation of wind direction distribution in different turbines. For the same turbine, wind direction distribution law between different months has some similarity but the significant difference is also observed. For example, the number of distribution peaks for different months is roughly the same, but the difference of distribution density is very big.

Fig.10 Monthly frequency distribution of wind direction
(a) WT1, (b) WT2, (c) WT3, (d) WT4

4.2 Wind direction fluctuation

Wind direction fluctuation is presented to describe the real-time wind direction variation characteristic with a higher sampling frequency (for example, 1Hz). A fluctuation coefficient of wind direction is defined to investigate the fluctuation characteristic in the following.

Firstly, the wind direction difference between two sampling is written as

$$\Delta x_k = |x_k - x_{k+1}| \quad (12)$$

where, x_k means wind direction data (sampling once a second).

Since wind direction of 0 degrees and wind direction of 360 degrees are essentially equivalent, Eq. (12) should be revised to be

$$\Delta x_e = \min \left\{ \begin{array}{l} \Delta x_k \\ |\Delta x_k - 360^\circ| \end{array} \right. \quad (13)$$

To construct the fluctuation coefficient of wind direction, a base of wind direction difference is needed, which is defined as

$$\Delta x_b = \max \left(\frac{1}{n} \sum_{k=1+i}^{n+i} \Delta x_e \right) \quad (14)$$

where, $i=1,2,3,\dots; n$ is the number of data point for one-time calculation.

Finally, fluctuation coefficient of wind direction w_d could be defined as

$$w_d(i) = \begin{cases} \frac{1}{n} \sum_{k=1+i}^{n+i} \Delta x_e, & \Delta x_b \neq 0 \\ 0, & \Delta x_b = 0 \end{cases} \quad (15)$$

Like the use of wind speed fluctuation coefficient, time continuous data is needed. Data should be preprocessed according to Eq. (9). Using wind direction data corresponding to wind speed data in Fig.7, wind direction fluctuation curves are obtained (Fig.11). Fig.11(a) shows the wind direction time-varying curve for WT1 and WT2, and Fig.11(b) shows the corresponding fluctuation coefficients. Fig.11(c) shows the wind direction time-varying curve for WT3 and WT4, and the corresponding fluctuation coefficient is in Fig.11(d). Both Fig.11(b) and Fig.11 (d) are under the condition of $n=10$ which used in Eq. (15). For the convenience of analysis, wind direction data greater than 300 degrees is represented by a negative data in Fig.11 (c). From the figure, it can be seen that WT1 and WT2 have a close wind direction distribution, WT3 and WT4 have another close wind direction distribution. Fluctuation coefficient of wind direction has the characteristic of jumping for all turbines. Moreover, fluctuation coefficients of wind direction for WT1 and WT2 are much larger than that for WT3 and WT4. Especially, WT3 has a very low fluctuation coefficient of wind direction. In general terms, wind direction fluctuation is significantly different for different turbines, which also shows the characteristics of mountain wind farm.

Fig.11 Wind direction fluctuation

5. Joint investigation of wind speed and wind direction

To investigate the joint distribution of wind speed and wind direction, Eq. (1) is rewritten as Eq. (16) which is a multi-dimensional kernel density estimation[52].

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \left\{ \prod_{j=1}^d \frac{1}{h_j} K\left(\frac{x_j - x_{ij}}{h_j}\right) \right\} \quad (16)$$

where, $X = \{x_{ij}\}$ is the wind speed and wind direction sample; n is the sample size; K is the multivariate kernel function with d arguments ($d=2$ in this paper).

Fig.12 shows the yearly joint distribution of wind speed and wind direction. From the figures, it can be seen

that the joint distribution surface of wind speed and wind direction has prominent distribution peaks, which means a data point taking into account the maximum probability of wind speed and wind direction is obtained. Moreover, wind direction distribution for WT3 and WT4 is more concentrated than that for WT1 and WT2. Especially for WT1, the wind direction distribution is obviously very scattered. Specifically, three distribution peaks appear for WT1 and two peaks appear for WT2, WT3, and WT4. When wind speed is 2.6m/s and wind direction is 170 degrees, the maximum joint distribution peak for WT1 appears. When wind speed is 7.5m/s and wind direction is 314 degrees, the maximum joint distribution peak for WT2 appears. When wind speed is 6.3m/s and wind direction is 214 degrees, the maximum joint distribution peak for WT3 appears. When wind speed is 4.8m/s and wind direction is 173 degrees, the maximum joint distribution peak for WT4 appears. Relatively speaking, the joint distribution of wind speed and wind direction for WT2 seems to be more conducive to wind power because it has the maximum wind speed corresponding the maximum joint distribution peak.

Fig.12 Yearly joint distribution of wind speed and wind direction

Seasonal joint distribution of wind speed and wind direction for four turbines is shown in Fig.13. The seasonal joint distribution for WT1 is in Fig.13 (a) to Fig.13 (d), for WT2 in Fig.13 (e) to Fig.13 (h), for WT3 in Fig.13 (i) to Fig.13 (l), for WT4 in Fig.13 (m) to Fig.13 (p). Except for WT1, there are two obvious joint distribution peaks even in each season for WT2, WT3, and WT4. In other words, for WT1, the joint distribution of wind speed and wind direction is relatively messy; more regular distributions appear in WT2, WT3, and WT4. Also, the joint distribution is obviously different in different seasons. Compared with spring, summer, and winter, the joint distribution of wind speed and wind direction were more concentrated in autumn for WT1. When wind speed is 4.2m/s and wind direction is 14 degrees, the maximum joint distribution peak appears in autumn. For WT2, the most concentrated joint distribution of wind speed and wind direction appears in winter. Corresponding to the maximum peak, wind speed is 8.3m/s, the wind direction is 309 degrees. Winter seems to have the best wind energy utilization potential because it not only has a centralized joint distribution of wind speed and wind direction, but also the wind speed corresponding to the maximum peak is the largest than other

485 seasons. For WT3, the most concentrated joint distribution appears in winter like that in WT2. Corresponding
 486 to the maximum peak, wind speed is 6.8m/s, the wind direction is 213 degrees. For WT4, in both summer and
 487 winter, the relatively concentrated joint distribution appears. Corresponding to the maximum peak, wind speed
 488 is 7.6m/s, the wind direction is 13 degrees in summer; the wind speed is 6.4m/s, the wind direction is 173
 489 degrees in winter. Generally, the joint distribution of wind speed and wind direction will change with location
 490 change in the investigated mountain wind farm.

491

492

493

494

495

Fig.13 Seasonal joint distribution of wind speed and wind direction

6. Conclusion

For a mountain wind farm, the local wind characteristics are determined by not only the overall weather system, but also the local terrain-topography such as hills, cliffs, and ridges. Thus, Wind resource characteristics in the mountain wind farm are inevitably more complicated and changeable than that in a flat wind farm. Through the separate investigation of wind speed and wind direction characteristics, and the joint investigation of wind speed and wind direction, the specific characteristics of wind resource in the investigated mountain wind farm are revealed. Significant location difference, seasonal difference, and monthly difference of wind characteristics are observed. Yearly mean wind speeds

for the four locations are 3.8m/s, 5.1m/s, 5.4m/s and 5.6m/s, respectively. This gives an important information that in the mountain wind farm, the layout of wind turbines should take into account local geographical differences. For example, the length of wind turbine blades at different locations may vary.

To solve the issue of big data processing and data error, data preprocessing methods are proposed; to solve the lack of prior knowledge on wind distribution, several kinds of PDFs are compared, and kernel density estimation method is employed including one-dimensional and multidimensional estimation. The methods provided are universal and can also be used for the processing and analysis of other wind farm data.

In the future, by combining wind resource characteristics and energy output characteristics of wind turbines, more in-depth investigation conclusions will be obtained. Moreover, high-resolution wind data from other mountain wind farms will be extracted to obtain a more general conclusion. All these new investigation achievements will be reported in separate papers.

Acknowledgements

This work is supported by the National Scholarship Foundation, National Natural Science Foundation of People's Republic of China (grant number 51675175, 51475160), State Scholarship Fund (grant number 201408430267), Hunan Provincial Natural Science Foundation (grant number 2016JJ5024) and Scientific Research Fund of Hunan Provincial Education Department (grant number 15B084).

References:

- [1] NEMATOLLAHI O, HOGHOOGHI H, RASTI M, et al. Energy demands and renewable energy resources in the Middle East[J]. Renewable and Sustainable Energy Reviews, 2016, 54: 1172-1181
- [2] BECERRA M, MOR N J, JEREZ A, et al. Wind energy potential in Chile: Assessment of a small scale wind farm for residential clients[J]. Energy Conversion and Management, 2017(140): 71–90
- [3] VALVERDE L, PINO F J, GUERRA J, et al. Definition, analysis and experimental investigation of operation modes in hydrogen-renewable-based power plants incorporating hybrid energy storage[J]. Energy Conversion and Management, 2016(113): 290-311
- [4] NOOROLLAHI Y, JOKAR M A, KALHOR A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran[J]. Energy Conversion and Management, 2016, 115: 17-25
- [5] HEPBASLI A, OZGENER O. A review on the development of wind energy in Turkey[J]. Renewable and Sustainable Energy Reviews, 2004, 8(3): 257-276
- [6] KOSE R, OZGUR M A, ERBAS O, et al. The analysis of wind data and wind energy potential in Kutahya, Turkey[J]. Renewable and Sustainable Energy Reviews, 2004, 8(3): 277-288
- [7] HE G, KAMMEN D M. Where, when and how much wind is available? A provincial-scale wind resource assessment for China[J]. Energy Policy, 2014, 74: 116-122
- [8] JOWDER F A. Wind power analysis and site matching of wind turbine generators in Kingdom of Bahrain[J]. Applied Energy, 2009, 86(4): 538-545
- [9] SHERIDAN B, BAKER S D, PEARRE N S, et al. Calculating the offshore wind power resource: Robust assessment methods applied to the US Atlantic Coast[J]. Renewable Energy, 2012, 43: 224-233
- [10] SAHIN B, BILGILI M, AKILLI H. The wind power potential of the Eastern Mediterranean region of Turkey[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2005, 93(2): 171-183
- [11] AKPINAR E K, AKPINAR S. A statistical analysis of wind speed data used in installation of wind energy conversion systems[J]. Energy conversion and management, 2005, 46(4): 515-532
- [12] VOGIATZIS N, KOTTI K, SPANOMITSIOS S, et al. Analysis of wind potential and characteristics in North Aegean, Greece[J]. Renewable energy, 2004, 29(7): 1193-1208
- [13] AYODELE T, JIMOH A, MUNDA J, et al. Wind distribution and capacity factor estimation for wind turbines in the coastal region of South Africa[J]. Energy Conversion and Management, 2012, 64: 614-625

- [14] JUSTUS C, HARGRAVES W, MIKHAIL A, et al. Methods for estimating wind speed frequency distributions[J]. Journal of applied meteorology, 1978, 17(3): 350-353
- [15] TULLER S E, BRETT A C. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis[J]. Journal of Climate and Applied Meteorology, 1984, 23(1): 124-134
- [16] WEISSER D. A wind energy analysis of Grenada: An estimation using the 'Weibull' density function[J]. Renewable energy, 2003, 28(11): 1803-1812
- [17] AKDAĞ S, BAGIORGAS H, MIHALAKAKOU G. Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean[J]. Applied Energy, 2010, 87(8): 2566-2573
- [18] BARDSLEY W. Note on the use of the Inverse Gaussian distribution for wind energy applications[J]. Journal of Applied Meteorology, 1980, 19(9): 1126-1130
- [19] ŞAHİN A Z, AKSAKAL A. Wind power energy potential at the Northeastern region of Saudi Arabia[J]. Renewable Energy, 1998, 14(1): 435-440
- [20] CELİK A N. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the Southern region of Turkey[J]. Renewable energy, 2004, 29(4): 593-604
- [21] SAFARI B, GASORE J. A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda[J]. Renewable Energy, 2010, 35(12): 2874-2880
- [22] PISHGAR-KOMLEH S, KEYHANI A, SEFEEDPARI P. Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooch county of Iran)[J]. Renewable and Sustainable Energy Reviews, 2015, 42: 313-322
- [23] LI M, LI X. Mep-type distribution function: A better alternative to Weibull function for wind speed distributions[J]. Renewable energy, 2005, 30(8): 1221-1240
- [24] SHAMILOV A, KANTAR Y M, USTA I. Use of minmaxent distributions defined on basis of maxent method in wind power study[J]. Energy Conversion and Management, 2008, 49(4): 660-677
- [25] KANTAR Y M, USTA I. Analysis of wind speed distributions: Wind distribution function derived from minimum cross entropy principles as better alternative to Weibull function[J]. Energy Conversion and Management, 2008, 49(5): 962-973
- [26] HOCAOĞLU F O, GEREK Ö N, KURBAN M. A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2010, 98(8): 472-481
- [27] MORGAN E C, LACKNER M, VOGEL R M, et al. Probability distributions for offshore wind speeds[J]. Energy Conversion and Management, 2011, 52(1): 15-26
- [28] SOUKISSIAN T. Use of multi-parameter distributions for offshore wind speed modeling: The Johnson SB distribution[J]. Applied energy, 2013, 111: 982-1000
- [29] ZHOU J, ERDEM E, LI G, et al. Comprehensive evaluation of wind speed distribution models: A case study for North Dakota Sites[J]. Energy Conversion and Management, 2010, 51(7): 1449-1458
- [30] LI G, SHI J. Application of Bayesian model averaging in modeling long-term wind speed distributions[J]. Renewable Energy, 2010, 35(6): 1192-1202
- [31] QIN Z, LI W, XIONG X. Estimating wind speed probability distribution using kernel density method[J]. Electric Power Systems Research, 2011, 81(12): 2139-2146
- [32] ZHANG J, CHOWDHURY S, MESSAC A, et al. A multivariate and multimodal wind distribution model[J]. Renewable Energy, 2013, 51: 436-447
- [33] SEGURO J, LAMBERT T. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis[J]. Journal of Wind Engineering and Industrial Aerodynamics, 2000, 85(1): 75-84
- [34] CHANG T P. Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application[J]. Applied Energy, 2011, 88(1): 272-282
- [35] ROCHA P A C, DE SOUSA R C, DE ANDRADE C F, et al. Comparison of seven numerical methods for determining

598 Weibull parameters for wind energy generation in the Northeast region of Brazil[J]. *Applied Energy*, 2012, 89(1):
599 395-400

600 [36] ALAVI O, MOHAMMADI K, MOSTAFAEIPOUR A. Evaluating the suitability of wind speed probability
601 distribution models: A case of study of East and Southeast parts of Iran[J]. *Energy Conversion and Management*
602 2016(119): 101-108

603 [37] GEBRAAD P M O, TEEUWISSE F W, WINGERDEN J W V, et al. Wind plant power optimization through yaw
604 control using a parametric model for wake effects—a CFD simulation study[J]. *Wind Energy*, 2016, 19(1): 95-114

605 [38] FLEMING P A, NING A, GEBRAAD P M O, et al. Wind plant system engineering through optimization of layout
606 and yaw control[J]. *Wind Energy*, 2016, 19(2): 329-344

607 [39] CARTA J A, RAMIREZ P, VELAZQUEZ S. A review of wind speed probability distributions used in wind energy
608 analysis: Case studies in the Canary Islands[J]. *Renewable and Sustainable Energy Reviews*, 2009, 13(5): 933-955

609 [40] WAIS P. Two and three-parameter Weibull distribution in available wind power analysis[J]. *Renewable Energy*, 2017,
610 103: 15–29

611 [41] CARTA J, RAMIREZ P. Analysis of two-component mixture Weibull statistics for estimation of wind speed
612 distributions[J]. *Renewable Energy*, 2007, 32(3): 518-531

613 [42] BLUDSZUWEIT H, DOM NGUEZ-NAVARRO J A, LLOMBART A. Statistical analysis of wind power forecast
614 error[J]. *IEEE Transactions on Power Systems*, 2008, 23(3): 983-991

615 [43] WAN C, XU Z, PINSON P, et al. Probabilistic forecasting of wind power generation using extreme learning
616 machine[J]. *IEEE Transactions on Power Systems*, 2014, 29(3): 1033-1044

617 [44] BARAN S. Probabilistic wind speed forecasting using bayesian model averaging with Truncated normal
618 components[J]. *Computational Statistics & Data Analysis*, 2014, 75: 227-238

619 [45] BARAN S, LERCH S. Log-normal distribution based ensemble model output statistics models for probabilistic wind-
620 speed forecasting[J]. *Quarterly Journal of the Royal Meteorological Society*, 2015, 141(691): 2289-2299

621 [46] BASILE S, BURLON R, MORALES F. Joint probability distributions for wind speed and direction. A case study in
622 Sicily[C].2015 International Conference on Renewable Energy Research and Applications (ICRERA), 2015: IEEE:
623 1591-1596.

624 [47] MASSERAN N, RAZALI A M, IBRAHIM K, et al. The probability distribution model of wind speed over East
625 Malaysia[J]. *Research Journal of Applied Sciences, Engineering and Technology*, 2013, 6(10): 1774-1779

626 [48] KAHROBAEE S, ASGARPOOR S. Short and long-term reliability assessment of wind farms[C].North American
627 Power Symposium (NAPS), 2010, 2010: IEEE: 1-6.

628 [49] SHEATHER S J, JONES M C. A reliable data-based bandwidth selection method for kernel density estimation[J].
629 *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991: 683-690

630 [50] DELAIGLE A, GIJBELS I. Practical bandwidth selection in deconvolution kernel density estimation[J].
631 *Computational statistics & data analysis*, 2004, 45(2): 249-267

632 [51] DAI J, CAO J, LIU D, et al. Power fluctuation evaluation of large scale wind turbines based on SCADA data[J]. *IET*
633 *Renewable Power Generation*, 2016

634 [52] DUONG T. Ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R[J]. *Journal of*
635 *Statistical Software*, 2007, 21(7): 1-16

636