Wind energy potential assessment based on wind speed, its direction and power data

Zhiming Wang (wangzhiming301@sohu.com)
Lanzhou University of Technology
Weimin Liu
Lanzhou University of Technology

Research Article

Keywords: Finite Mixture Statistical Distributions, Angular-linear Modeling, Weibull Mixture Distribution, Wide Wind Turbine Applications

Posted Date: July 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-594128/v1

Version of Record: A version of this preprint was published at Scientific Reports on August 19th, 2021. See the published version at https://doi.org/10.1038/s41598-021-96376-7.

Wind energy potential assessment based on wind speed, its direction and power data

Zhiming Wang[1,*] and Weimin Liu[1,2]

[1]School of Mechanical and Electronic Engineering, Lanzhou University of Technology, Lanzhou 730050, China.
[2]Gansu Province Special Equipment Inspection and Testing Institute, Lanzhou 730050, China
*corresponding. wangzhiming301@sohu.com

ABSTRACT

Based on wind speed, direction and power data, an assessment method of wind energy potential using finite mixture statistical distributions is proposed. Considering the correlation existing and the effect between wind speed and direction, the angular-linear modeling approach is adopted to construct the joint probability density function of wind speed and direction. For modeling the distribution of wind power density and estimating model parameters, based on expectation-maximization algorithm, a two-component three-parameter Weibull mixture distribution is chosen as wind speed model, and a von Mises mixture distribution with nine components and six components are selected as wind direction and relation circular variable models, respectively. A comprehensive technique of model selection, which includes Akaike information criterion, Bayesian information criterion, the coefficient of determination R^2 and root mean squared error, is used to select the optimal model in all candidate models. The proposed method is applied to averaged 10-minute field monitoring wind data and compared with the other estimation methods and judged by the values of R^2 and root mean squared error, histogram plot and wind rose diagram. The results show that the proposed method is effective and the area under study is not suitable for wide wind turbine applications, and the estimated wind energy potential would be inaccuracy without considering the influence of wind direction.

Introduction

Energy consumption increases dramatically with the rapid development of society and economy. Wind energy has attracted more and more attention because of its advantages such as abundance, renewability, natural cleanness, low cost and little negative impact on the environment, and has been used as an alternative to fossil fuels[1,2]. Therefore, the application of wind energy has already been selected as an important measure for the sustainable development of resources and environment all over the world. Wind energy also plays an important role in national economic growth which creates more employment opportunities[3,4]. Before developing wind power in a certain site, including the design, arrangement and condition monitoring of wind turbine systems, it is necessary to assess the wind energy potential and wind characteristics[5].

As wind energy is proportional to the cube of wind speed, this means that even a small increase in wind speed results in a large increase in wind energy, therefore, the most important factor affecting wind energy is wind speed. But wind speed is not constant, it always fluctuates with the varying of air temperature over a period of time in different geographic locations and seasons. In this case, we can take wind speed as a random variable and describe it by a probability density function (pdf). Therefore, the pdf of wind speed becomes an important basis for evaluating wind energy potential and wind stochastic characteristics[6]. If the frequency distribution of wind speed is comprehensively expressed by an estimated pdf, the wind power density and wind energy output of

wind turbines can be evaluated, which can help us make a reasonable decision whether to build a wind farm in the observed area or not, and reduce the uncertainties and the errors of wind power output estimation[7]. At the same time, the accurate estimated pdf also helps to select an optimal wind energy conversion system and to evaluate the reliability of generation system. Therefore, accurate evaluation of wind speed pdf is conducive to the prediction of wind energy potential and the selection of an optimal wind energy conversion system[8].

There are usually two kinds of models describing wind characteristics, one is parametric model, and the other is non-parametric model. The parametric model is also divided into single and mixture distribution models. At present, the most widely used single distribution models include Gamma, Raleigh, Inverse Gaussian distribution, lognormal and Weibull distribution models[2-14], etc. Among them, the two-parameter Weibull distribution model is often recognized as an effective model and is widely used in the field of wind industry to estimate wind energy potential mainly due to its simplicity. In some cases, the probability of calms (null wind speeds) or the wind speed below 2 m/s is significant, and the two-parameter Weibull distribution performs poorly for a high percentage of null wind speeds. While it can be observed that the three-parameter Weibull distribution gives a better result of energy calculation when the frequency of null wind speeds is higher[4,13]. A wind turbine can generate electricity only when wind speed exceeds the cut-in speed. So Deep et al.[14] pointed out that a three-parameter Weibull distribution must be used to model wind speed data between the cut-in and cut-out wind speeds, and the location parameter can be equivalent to the cut-in wind speed. On the other hand, when a distribution of wind speed is bimodal or multimodal, a single distribution model cannot perform well. In this case, some mixture distribution models[5,8,15-24], which consist of several single distribution models (called components), are used, such as the Weibull-Weibull mixture, the Gamma-Weibull mixture, the truncated Normal-Weibull mixture, etc.

Wind direction, which strongly influences wind speed, is also an important aspect affecting the wind energy when evaluating wind characteristics in a certain area. Gugliani et al.[25] argued that it is futile to study wind power at a particular site when wind direction was not analyzed. Han and Chu[26] also considered that the available wind resources change with the wind direction, especially in the low-speed and complex terrain areas. Therefore, the mixture model of von Mises (voM) distribution is commonly used for modelling wind direction data[27-32]. Besides the voM distribution, the other circular statistical distributions including the uniform distribution, wrapped-normal distribution and wrapped-Cauchy distribution, etc. are also suitable for modelling analysis of wind direction. The research results showed that a mixture of voM distributions provided a flexible model for studies of wind direction that have several modes[27,28]. However, compared to wind speed model, the statistical modeling of wind direction is more difficult and complex. On the other hand, wind speed and wind direction are dependent random variables, wind speed has a directional characteristic and wind direction can complement information about wind speed in analyses of wind energy potential[28,29]. Therefore, in the past fewer decades, several bivariate distribution models for simultaneously describing wind speed and direction, which named joint distribution models of wind speed and direction, have been proposed by different authors[33-40]. Carta et al.[33] have presented a joint distribution from two marginal distributions, a single truncated from below Normal-Weibull mixture distribution for wind speed and a finite mixture of voM distributions for wind direction. Erdem and Shi[35] given a comparison of bivariate distribution models for analyzing wind speed and direction data of multiple sites in North Dakota, USA.

The methods commonly used to estimate model parameters of wind speed or wind direction are

the graphical method or least square method (LSM), maximum likelihood estimate (MLE), modified maximum likelihood estimate (MMLE), empirical method (EmM), moment method (MoM), power density method (PDM), energy pattern factor method (EPFM), equivalent energy method (EEM) and copula-based approach, etc. Non-parametric model methods were also proposed in literature. These methods include the minimum cross entropy (MCE) method[41], maximum entropy principle method (MEPM)[42], kernel density estimation[43-45] and root-transformed local linear regression method[46], etc.

In this study, to evaluate wind energy potential, the single and mixture of two-parameter and three-parameter Weibull distributions were used as candidate models for wind speed data, and a finite mixture of voM distributions was used for wind direction data. Based on MLE, the expectation-maximization (EM)[8,29,45,47] optimization algorithm is applied to estimate the model parameters of mixture distributions. As Carta et al.[27] pointed out that although the mixture distributions enrich the modelling and have high degrees of fits, the model complexity increases with the increasing of more number of model parameters. Therefore, we use Akaike information criterion (AIC) and Bayesian information criterion (BIC) to select the optimal model, and adopt the coefficient of determination R^2 and the root mean squared error (RMSE) to evaluate the goodness-of-fit of model. The number of components in mixture model does not need to be known in advance.

The paper is organized as follows. In "Methodology" we give some details on the modelling for wind data using the Weibull and voM distributions, including parameter estimation, model selection and validation. The assessment of wind energy potential is described in "Wind power estimation". While in "Case study" presents some information about the observed field and the statistical description for wind speed, direction and power. Results and comparison with the observation data are presented with details in "Results and discussion". Conclusions are drawn in the final and concluding section.

Methodology

Wind speed model with Weibull distribution. When the frequency of low wind speed, especially of null winds is significant, a three-parameter Weibull distribution can be used to model this wind speed data well and a more appropriate results can be obtained. The pdf of wind speed using the three-parameter Weibull distribution is given by[4,13]:

$$f\left(v;\eta,\beta,\gamma\right)=\left(\frac{\beta}{\eta}\right)\left(\frac{v-\gamma}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{v-\gamma}{\eta}\right)^{\beta}\right] \tag{1}$$

where v is wind speed, η is the scale parameter (m/s), $\eta > 0$, β represents the shape parameter, $\beta > 0$, and γ is the position parameter, $\gamma \le 0$. When $\gamma = 0$, three-parameter Weibull distribution reduces to two-parameter Weibull distribution.

Mixture distributions are defined as linear combinations of two or several distributions. Therefore, there are more parameters needed to estimate for mixture distribution model than that of single distribution model. And parameter estimation of mixture distribution model is more complex and difficult. The MLE is one of the efficient methods to estimate model parameters.

Mixture model is a weighted sum with several single models[16-19,22-24], therefore, a pdf of mixture distribution model for m-component three-parameter Weibull distributions can be given by

$$f_V\left(v;w_i,\eta_i,\beta_i,\gamma_i\right)=\sum_{i=1}^{m}w_i f\left(v;\eta_i,\beta_i,\gamma_i\right)=\sum_{i=1}^{m}w_i\left(\frac{\beta_i}{\eta_i}\right)\left(\frac{v-\gamma_i}{\eta_i}\right)^{\beta_i-1}\exp\left[-\left(\frac{v-\gamma_i}{\eta_i}\right)^{\beta_i}\right] \tag{2}$$

where w_i is the weight coefficient of the ith component, and must satisfy the following conditions:

$$0 \le w_i \le 1 \ \left(i = 1, ..., m\right) \text{ and } \sum_{i=1}^{m} w_i = 1 \tag{3}$$

Given the n observed wind speed data $\mathbf{V} = [v_1, v_2, \ldots, v_n]$, the likelihood function on $\mathbf{V}$ can be obtained by

$$L\left(\mathbf{V}; \mathbf{\Lambda}\right) = \prod_{j=1}^{n} \sum_{i=1}^{m} \left\{ w_i \left(\frac{\beta_i}{\eta_i}\right) \left(\frac{v_j - \gamma_i}{\eta_i}\right)^{\beta_i - 1} \exp\left[-\left(\frac{v_j - \gamma_i}{\eta_i}\right)^{\beta_i} \right] \right\} \tag{4}$$

where $\mathbf{\Lambda} = [\boldsymbol{w}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}]$ are unknown model parameters of wind speed.

Then the log-likelihood function can be given as follows:

$$\ln L\left(\mathbf{V}; \mathbf{\Lambda}\right) = \sum_{j=1}^{n} \ln \sum_{i=1}^{m} \left\{ w_i \left(\frac{\beta_i}{\eta_i}\right) \left(\frac{v_j - \gamma_i}{\eta_i}\right)^{\beta_i - 1} \exp\left[-\left(\frac{v_j - \gamma_i}{\eta_i}\right)^{\beta_i} \right] \right\} \tag{5}$$

Due to the complexity of the log-likelihood function, the model parameters cannot be got by taking the partial derivatives of log-likelihood function with respect to each parameter and setting them equal to zero. Therefore, the log-likelihood function is maximized directly to estimate the model parameters. Unfortunately, there is no closed-form expression for computing them, it only can be numerically estimated. Therefore, a numerical method, such as EM algorithm is needed to find the maximum likelihood estimates of the parameters.

EM algorithm is an iterative method for finding maximum likelihood or maximum a posteriori estimates of model parameters for statistical distribution from a given data set. It proceeds iteratively in two steps, the expectation (E) step and maximization (M) step. In the E-step, a function for the expectation of the log-likelihood is created, and the hidden variables or missing data are estimated given the observed data and current estimator of model parameters. In the M-step, the likelihood function defined by the previous E-step is maximized to obtain new parameter estimations under the assumption that the hidden variables or missing data are known. It should be noted that the E-step and M-step in EM algorithm are performed iteratively until the algorithm converges. Initial values are required for the iterative procedure. In this study, the population is divided into m components, and the estimated parameters are assumed to have the same values with a single Weibull distribution for each component. These estimates are considered as initial values for the iterative procedure.

For an m-component mixture model, Eq (6) is used to find the mean c_1, variance c_2, the coefficients of skewness c_3 and kurtosis c_4 of wind speed, respectively.

$$E\left(v^d\right) = \sum_{i=1}^{m} w_i \int_0^{\infty} v^d f_V\left(v; w_i, \eta_i, \beta_i, \gamma_i\right) \mathrm{d}v, \ d = 1, 2, 3, 4 \tag{6}$$

When the model parameters of mixture three-parameter Weibull distributions are known, based on Eq (6), the values of c_1, c_2, c_3 and c_4 can be obtained as follows, respectively:

$$c_1 = \sum_{i=1}^{m} w_i \left[\gamma_i + \eta_i \Gamma\left(1 + \frac{1}{\beta_i}\right) \right] \tag{7}$$

$$c_2 = \sum_{i=1}^{m} w_i \left[\eta_i^2 \left(\Gamma\left(1 + \frac{2}{\beta_i}\right) - \Gamma^2\left(1 + \frac{1}{\beta_i}\right) \right) \right] \tag{8}$$

$$c_3=\sum_{i=1}^{m} w_i\left[\frac{\Gamma\left(1+\frac{3}{\beta_i}\right)\text{-}3\Gamma\left(1+\frac{1}{\beta_i}\right)\Gamma\left(1+\frac{2}{\beta_i}\right)+2\Gamma^3\left(1+\frac{1}{\beta_i}\right)}{\left[\Gamma\left(1+\frac{2}{\beta_i}\right)\text{-}\Gamma^2\left(1+\frac{1}{\beta_i}\right)\right]^{3/2}}\right] \tag{9}$$

$$c_4=\left(\sum_{i=1}^{m} w_i\left[\frac{\Gamma\left(1+\frac{4}{\beta_i}\right)\text{-}4\Gamma\left(1+\frac{1}{\beta_i}\right)\Gamma\left(1+\frac{3}{\beta_i}\right)+6\Gamma\left(1+\frac{2}{\beta_i}\right)\Gamma^2\left(1+\frac{1}{\beta_i}\right)\text{-}3\Gamma^4\left(1+\frac{1}{\beta_i}\right)}{\left[\Gamma\left(1+\frac{2}{\beta_i}\right)\text{-}\Gamma^2\left(1+\frac{1}{\beta_i}\right)\right]^{2}}\right]\right)\text{-}3 \tag{10}$$

Wind direction model with von Mises distribution. For the assessment of wind direction, the voM distribution is used. Consider a random variable θ following the voM distribution, the corresponding pdf is[27,28]

$$f(\theta;\mu,\alpha)=\frac{\exp\left[\alpha\cos(\theta-\mu)\right]}{2\pi I_0(\alpha)};\ 0\le\theta\le 2\pi \tag{11}$$

where θ is wind direction in radians units, μ denotes location parameter or mean direction on the circle, $0 \le \mu \le 2\pi$, α represents concentration parameter, $\alpha \ge 0$, and $I_0(\alpha)$ is the modified Bessel function of the first kind of order zero, given by[27,28]

$$I_0(\alpha)=\frac{1}{2\pi}\int_0^{2\pi}\exp(\alpha\cos\theta)\mathrm{d}\theta=\sum_{r=0}^{\infty}\frac{1}{(r!)^2}\left(\frac{\alpha}{2}\right)^{2r} \tag{12}$$

When wind direction has several modes or prevailing wind directions, the distribution of wind direction comprises a finite mixture of voM distributions. Thus, based on Eq.(11), the corresponding pdf of mixture distribution model can be given by

$$f_\Theta(\theta;p_i,\mu_i,\alpha_i)=\sum_{i=1}^{k}p_i f(\theta;\mu_i,\alpha_i)=\sum_{i=1}^{k}\frac{p_i}{2\pi I_0(\alpha_i)}\exp\left[\alpha_i\cos(\theta-\mu_i)\right] \tag{13}$$

where k is the number of components and p_i is the weight coefficient of the ith component that sum to one, given by

$$0\le p_i\le 1\ (i=1,...,k)\ \text{and}\ \sum_{i=1}^{k}p_i=1 \tag{14}$$

The mixture of voM distributions corresponds to the weighted sum of several voM distributions. It is thus suitable for the statistical description of multimodal datasets. Given the n observed wind direction data $\mathbf{\Theta} = [\theta_1, \theta_2, \ldots, \theta_n]$, the likelihood function on $\mathbf{\Theta}$ is given by

$$L(\mathbf{\Theta};\mathbf{\Delta})=\prod_{j=1}^{n}\sum_{i=1}^{k}\left\{\frac{p_i\exp\left[\alpha_i\cos(\theta_j-\mu_i)\right]}{2\pi I_0(\alpha_i)}\right\} \tag{15}$$

where $\mathbf{\Delta} = [\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\alpha}]$ are unknown model parameters of wind direction. Then the log-likelihood function is computed by the following expression:

$$\ln L(\mathbf{\Theta};\mathbf{\Delta})=\sum_{j=1}^{n}\ln\sum_{i=1}^{k}\left\{\frac{p_i\exp\left[\alpha_i\cos(\theta_j-\mu_i)\right]}{2\pi I_0(\alpha_i)}\right\} \tag{16}$$

The model parameters can also be estimated by the EM algorithm for maximum likelihood estimation[29,47].

If the model parameters of the voM mixture distributions are known, the mean b_1 and variance b_2 can also be got as follows, respectively:

$$b_1 = \sum_{i=1}^{k} p_i \mu_i \tag{17}$$

$$b_2 = \sum_{i=1}^{k} p_i \left[1 - \frac{I_1(\alpha_i)}{I_0(\alpha_i)} \right] \tag{18}$$

Joint distribution model of wind speed and direction. Based on Eq. (2), the corresponding cumulative distribution function (CDF) for wind speed is given as follows:

$$F_V(v;\eta_i,\beta_i,\gamma_i) = \int_0^v f_V(v')\mathrm{d}v' = \sum_{i=1}^{m} w_i \left\{ 1 - \exp\left[-\left(\frac{v-\gamma_i}{\eta_i} \right)^{\beta_i} \right] \right\} \tag{19}$$

Eq. (13) can also be expressed as a series of Bessel functions and given by

$$f_\Theta(\theta;p_i,\mu_i,\alpha_i) = \sum_{i=1}^{k} \left[\frac{p_i}{2\pi I_0(\alpha_i)} \left(I_0(\alpha_i) + 2\sum_{q=1}^{\infty} I_q(\alpha_i)\cos\left[q(\theta-\mu_i) \right] \right) \right] \tag{20}$$

where $I_q(\alpha)$ is the modified Bessel function of the first kind of order q, whose expression is given by

$$I_q(\alpha) = \sum_{r=0}^{\infty} \frac{1}{r!(q+r)!} \left(\frac{\alpha}{2} \right)^{q+2r}, \quad q = 1,2,\ldots \tag{21}$$

Therefore, using Eq. (20), the CDF for wind direction can be obtained as follows:

$$F_\Theta(\theta;p_i,\mu_i,\alpha_i) = \int_0^\theta f_\Theta(\theta')\mathrm{d}\theta' = \sum_{i=1}^{k} \frac{p_i}{2\pi I_0(\alpha_i)} \left\{ \theta I_0(\alpha_i) + 2\sum_{q=1}^{\infty} \frac{I_q(\alpha_i)\sin\left[q(\theta-\mu_i) \right]}{q} \right\} \tag{22}$$

The joint angular-linear pdf of wind speed and direction is then given as[33]

$$f_{V,\Theta}(v,\theta) = 2\pi g(\zeta) f_V(v) f_\Theta(\theta); 0 \le \theta \le 2\pi; -\infty \le v \le \infty \tag{23}$$

where $g(\zeta)$ is the correlation pdf of circular variable ζ between wind speed v and direction θ, and ζ given by[33]

$$\zeta = 2\pi\left[F_V(v) - F_\Theta(\theta) \right] \tag{24}$$

Using the above definitions, the values of ζ_j can be obtained for each pair of values of wind speed v_j and direction θ_j from a sample of size n given as

$$\zeta_j = 2\pi\left[F_V(v_j) - F_\Theta(\theta_j) \right] \quad j = 1,\ldots,n \tag{25}$$

with the following condition[33]:

$$\zeta_j = 2\pi + \zeta_j \quad \text{if } \zeta_j < 0; \; j = 1,\ldots,n \tag{26}$$

Therefore, based on Eqs. (2), (13), (19), (22) and (24), the pdf $g(\zeta)$ of the variable ζ can also be described by a mixture of voM distributions.

Selection of optimal model. For mixture distribution model, the selection of the number of components is important. The log-likelihood cannot be directly used for selecting the number of

components, since the log-likelihood value increases with the number of components. The best models of wind speed and direction are selected by two information criteria, which are named the Akaike information criterion (AIC)[12,25,39,47] and Bayesian information criterion (BIC)[37]. The values of AIC and BIC are defined as

$$\mathrm{AIC} = -2\max\ln L + 2l,\ \mathrm{BIC} = -2\max\ln L + l\ln n \tag{27}$$

where l is the number of estimated parameters, n is the number of all observations, and maxln L is the maximized log-likelihood. The unknown parameters $l = 4m$-1 for m components three-parameter Weibull mixture models, and $l = 3m$-1 for m components voM mixture models. AIC and BIC contain the penalization terms that take into account the number of model parameters and all observed values to counterbalance the maximized log-likelihood. To avoid overfitting, AIC penalizes model for its complexity only with model parameters, while BIC imposes a greater penalty for additional parameters than AIC. So, AIC and BIC give a comprehensive balance in order to find a good tradeoff between the goodness-of-fit and the complexity of the model, and avoid the risk of choosing a complex model with a poor generalization. The smaller the values of AIC and BIC are, the higher the fit accuracy of the model is. Therefore, the number of components does not need to be known in advance.

Validation of model. The coefficient of determination (R^2) and the root mean squared error (RMSE) are used to judge the goodness-of-fit of different mixture models to wind speed and direction data, because it quantifies the correlation between the observed and the estimated probability density according to a particular distribution.

The coefficient of determination is the square of the correlation between the estimated values and observed values, and is defined by[13,16]

$$R^2 = 1 - \frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{\sum_{j=1}^{n}\left(y_j - y_m\right)^2} \tag{28}$$

where y_j is the jth observed value, y_m is the mean value of all observations, and $\hat{y}_j$ is the jth estimated value, respectively.

A large value of R^2 indicates the proposed distribution fits the wind speed data set well in all candidate models. Unlike the value of R^2, a high RMSE value indicates a poor fit. The smaller the values of RMSE are, the better the proposed distribution function approximates the observed data. It can be given by[13,16]

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{n}} \tag{29}$$

Wind power estimation

The density of air changes slightly with air temperature and with altitude at a potential site, supposed that the air density is constant, based on a pdf $f(v)$ of wind speed v at a height for a wind turbine, in theory, the available wind power $P(v)$ in W can be computed by[10,12]

$$P(v) = \frac{1}{2}\rho A\int_0^{\infty} v^3 f(v)\,\mathrm{d}v \tag{30}$$

where ρ is the air density (kg/m^3), A is the sweep area of a wind turbine rotor (m^2).

Using the real wind speed values of time series wind data, the mean wind power density or

effective wind energy density, denoted as $p_e(v)$ in W/m^2 is estimated as follows[12,22,42]:

$$p_e(v) = \frac{1}{2}\rho\sum_{j=1}^{n_e}\frac{v_j^3}{n_e} \tag{31}$$

where n_e is the number of effective wind speed, which lies between the cut-in wind speed v_{in} and cut-out wind speed v_{out} of wind turbine.

If the wind direction influence is ignored, the wind power density $p(v)$ for a specified wind turbine can be obtained numerically by

$$p(v) = \frac{1}{2}\rho\int_{v_{\text{in}}}^{v_{\text{out}}} v^3 f(v)\,\mathrm{d}v \tag{32}$$

When considering the effect of wind direction, similar to Eq. (32), the wind power density $p(v,\theta)$ at different wind speed and direction can be computed numerically by[37]

$$p(v,\theta) = \frac{1}{2}\rho\int_0^{2\pi}\int_{v_{\text{in}}}^{v_{\text{out}}} v^3 f_{V,\Theta}(v,\theta)\,\mathrm{d}v\mathrm{d}\theta \tag{33}$$

The power curve gives a relation between the output power $P_w(v)$ and wind speed v, and this relation can be expressed by a polynomial function of degree u as follows[48-50]:

$$P_w(v) = \begin{cases} 0 & v < v_{\text{in}}, v \geq v_{\text{out}} \\ a_0 + \sum_{i=1}^{u} a_i v^i & v_{\text{in}} \leq v < v_r \\ P_r & v_r \leq v < v_{\text{out}} \end{cases} \tag{34}$$

where $P_w(v)$ is the output power, P_r denotes the rated power of wind turbine, v_r represent the rated speed, a_0 and a_i are the regression constants which can be obtained using a polynomial regression method.

Based on the power curve of a specified wind tribune, the wind energy output $E(v,\theta)$ with a joint pdf within a period of time t can be calculated numerically by[33]

$$E(v,\theta) = t\int_0^{2\pi}\left[\int_{v_{\text{in}}}^{v_{\text{out}}} P_w(v) f_{V,\Theta}(v,\theta)\,\mathrm{d}v\right]\mathrm{d}\theta \tag{35}$$

Case study

Wind speeds are continuously acquired for a significant time, usually not less than one year. Therefore, the data used in this study were collected in a period of one year (January 1, 2019 to December 31, 2019) and measured at a height of 30 m above the ground level from the Maling Mountain wind farm (34°31.4′ N and 118°45.7′ E) located in Jiangsu Province, China. The data records corresponding to the periods of averaged 10 min each and containing the wind speed, direction and power information available from supervisory control and data acquisition (SCADA) systems. The Maling Mountain wind farm is selected due to the fact that the histogram of wind speed indicates that the frequency of 0-2 m/s wind speed range is 8.24% for this station. The percentage of null wind speed or calms (0-0.2 m/s) at this wind farm is 0.50%. Wind direction data are circular because they are recorded in terms of degrees, from 0° clockwise to 360°. However, for modeling convenient, the data were converted into radian units. After removing some abnormal and unreasonable data such as the missing data by sensor fault, measurement error data and low temporal resolution data, a total of 47084 wind data are collected. The statistical description of wind speed, its direction and power data are shown in Table 1.

	Wind speed (m/s)	Wind direction (rad)	Power (kW)
Mean	4.3902	2.7238	367.6699
Variance	3.3240	2.7547	172406.7517
Skewness	0.8119	0.3562	2.1340
Kurtosis	2.4538	-0.9237	1.5765
Min	0	0.0003	-4.0816
Max	18.9878	6.2830	1842.7400
Number	47084	47084	47084

Table 1. Statistical description of wind speed, its direction and power data

Results and discussion

The estimated parameters with different methods for the Weibull distribution wind speed model are shown in Table 2.

Model	Method	Parameter				
		i	w_i	η_i	β_i	γ_i
Single model	2-p LSE	1	1	5.8556	1.3907	0
	2-p MLE	1	1	4.9040	2.4252	0
	2-p MoM	1	1	4.9436	2.5850	0
	3-p EM	1	1	5.2625	3.2919	-0.2326
Mixture model	3-p EM (m = 2)	**1**	**0.6525**	**4.8655**	**3.6591**	**0**
		2	**0.3475**	**6.0169**	**2.3126**	**-0.1415**
	3-p EM (m = 3)	1	0.0010	2.0090	1.1026	-0.2019
		2	0.3241	5.0169	1.4774	-0.1440
		3	0.6749	6.0114	1.5496	-0.0010

Table 2. Parameter estimation results of wind speed model with 2-parameter and 3-parameter Weibull distributions using different methods

A comparison results with different methods for the Weibull distribution wind speed model are also given in Table 3. Based on the information criteria of AIC and BIC, we can see that the fit accuracy of mixture model is higher than that of single model, and the accuracy of LSE is the lowest. For single model, three-parameter model has a higher fit accuracy than that of two-parameter model. The reason is that the former considers the null wind speed. However, for mixture model, as the number of components increases, the accuracy of the model decreases. Therefore, we select two-component three-parameter Weibull mixture model as the optimal model for wind speed. This result is also confirmed by the values of RMSE and R^2. Because a lower value of RMSE and a higher value of R^2 indicate a better fit of the model to the data, and two-component three-parameter Weibull mixture model has the lowest value of RMSE and highest value of R^2 in all candidate models. It is worth noting that in the case of multi-modal data, the fitting, modeling and analysis for a statistical distribution are more accurate than an ordinary regression analysis, a value of R^2 only more than 0.7 is not sufficient[28]. In this study, this value is high, it is 0.9944.

Model	Method	ln L	AIC	BIC	RMSE	R^2
	2-p LSE	-48145.4007	96294.8014	96312.3208	0.0520	0.8505

Single model	2-p MLE	-21961.6692	43927.3384	43944.8578	0.0211	0.9257
	2-p MoM	-18957.1679	37918.3358	37935.8552	0.0169	0.9526
	3-p EM	-11746.7293	23499.4586	23525.7377	0.0080	0.9894
Mixture model	3-p EM (m = 2)	**-9309.0261**	**18632.0522**	**18693.3700**	**0.0058**	**0.9944**
	3-p EM (m = 3)	-9450.2055	18922.4110	19018.7676	0.0484	0.8098

Table 3. Comparison results with different methods for Weibull wind speed model

The fit results of different models are given in Fig. 1 where SM denotes single model, MM means mixture model and two-parameter is abbreviated as 2-p. It also shows that two-component three-parameter Weibull mixture model adequately fits the frequency histogram of wind speed well than other models. The fit accuracy of three-component three-parameter Weibull mixture model and two-parameter Weibull single model with LSE method is the lowest in all models. To fit the sample histogram, the wind speed and direction intervals must be given. In this case, the bin size of wind speed and direction interval was selected as 0.5 m/s and 10° (0.1745 rad)[7,21,27,29,35], which is often considered to be reasonable in wind energy analyses. The speed interval of 0.5 m/s is also close to the value of 2 km/h (approximates 0.56 m/s) used by Deep et al.[14] and Gugliani et al[25]. They concluded that the class width of 2 km/h gives a minimum error for modelling wind speed data.

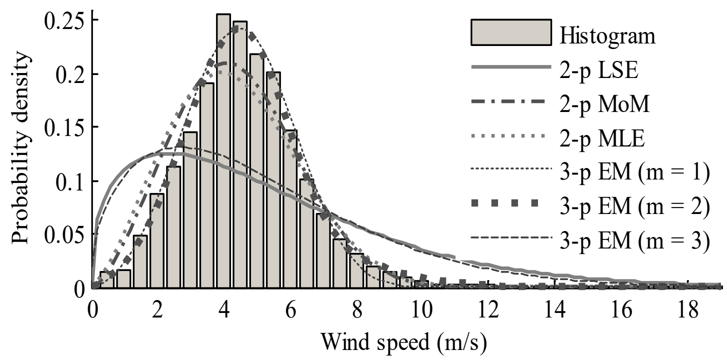

Figure 1. Histogram of wind speeds

Using Eqs. (7)-(10), we can obtain the estimated values of the mean, variance, the coefficients of skewness and kurtosis for wind speed, respectively, as follows: $c_1 = 4.6666$, $c_2 = 3.2406$, $c_3 = 0.1470$ and $c_4 = 2.8026$. Compared to the real statistical values in Table 1, the relative errors of mean and variance are small, they are only 6.30% and 2.51%, respectively. The estimated values of $c_3 > 0$ and $c_4 < 3$ all indicate that the probability distribution of wind speed has a long and light right tail than a normal distribution. This result agrees well with Fig. 1.

In Table 4, the estimated parameters with different components for voM wind direction model are given.

k	***p***	***α***	***μ***	***k***	***p***	***α***	***μ***	***k***	***p***	***α***	***μ***
1	1.0000	0.4329	2.0577	7	0.1952	2.7826	2.9787	9	**0.1729**	**5.8521**	**3.0405**
2	0.4725	0.0010	3.6480		0.0635	25.8134	0.5254		**0.1302**	**15.4270**	**0.5544**
	0.5275	0.8540	2.0127		0.4711	1.7217	1.6196		**0.0021**	**2.5792**	**2.2854**
3	0.7561	0.8822	2.0134		0.0010	4.3601	0.0010		**0.0010**	**4.3479**	**0.0010**
	0.0447	39.3747	0.5511		0.2218	2.1591	4.5968		**0.2297**	**2.4925**	**4.3874**
	0.1992	1.5674	4.7000		0.0464	12.1652	5.6580		**0.0969**	**6.0480**	**5.6471**

4	0.2908	1.1954	4.9078		0.0010	4.9832	1.4314		**0.1922**	**8.0584**	**2.1458**
	0.0542	38.0446	0.5274	8	0.1369	2.5621	3.1525		**0.1740**	**9.2402**	**1.3416**
	0.1466	4.0033	1.2829		0.2117	3.3017	0.7875		**0.0010**	**4.9909**	**1.5989**
	0.5084	1.2699	2.4422		0.3744	2.0779	2.0594	10	0.1709	5.8695	3.0507
5	0.4350	1.4041	2.6909		0.0010	4.3421	0.0010		0.1276	15.4501	0.5504
	0.0010	0.1277	0.0010		0.1568	2.1431	4.2825		0.0010	2.5804	2.2857
	0.2030	3.6102	1.5138		0.1172	3.4021	5.2958		0.0010	4.3455	0.0010
	0.0941	13.5594	0.5751		0.0010	4.9803	1.5391		0.2289	2.4971	4.3906
	0.2669	1.4701	4.9751		0.0010	4.9875	1.4903		0.0965	6.0483	5.6461
6	0.4295	1.5940	2.7449						0.1761	8.0646	2.1734
	0.1015	14.8627	0.5334						0.1424	9.2373	1.3224
	0.2346	3.4898	1.4698						0.0546	4.9903	1.5884
	0.0010	1.3469	0.9287						0.0010	4.9907	1.5990
	0.1467	3.6726	4.5639								
	0.0867	7.7691	5.6304								

Table 4. Parameter estimation results for mixture voM wind direction model

Table 5 is a comparison results with the different mixture models for wind direction. It can be observed that increasing the number of voM components from one to ten in the mixture distributions will increase the value of the R^2 coefficient, which indicates a better fit to the data. However, when the number of components increases to ten, the value of the R^2 coefficient does not increase any more, it is the same as the nine-component mixture model has. At the same time, a nine-component mixture model has the same value of RMSE as a ten-component mixture model. In this situation, how to select the best wind direction model? However, it is noticed that nine-component mixture model has lower values of AIC and BIC, so based on the information criteria of AIC and BIC, and the values of RMSE and R^2, the best wind direction model would be selected using the comprehensive criteria of information and goodness-of-fit. It can be concluded that the most suitable model for wind direction at this station is a voM mixture model with nine components distribution. Generally speaking, a voM mixture distribution with six components for the modelling of wind direction is enough, increasing the number of components of mixture distributions, the variations in value of R^2 are not significant[27]. On the contrary, it would yield a complex model. Therefore, combining with the comprehensive criteria of model selection, an appropriate number of mixture distributions can be selected, which not only reduces the computational burden but also improves the model accuracy, and the model has a higher predictive ability.

k	ln L	AIC	BIC	RMSE	R^2
1	-84740.7450	169485.4899	169503.0093	0.0308	0.5798
2	-84706.8360	169423.6719	169467.4703	0.0304	0.5854
3	-84394.3660	168804.7320	168874.8095	0.0199	0.8522
4	-84387.5764	168797.1529	168893.5094	0.0169	0.8990
5	-84405.4189	168838.8377	168961.4734	0.0184	0.8774
6	-84410.4464	168854.8928	169003.8075	0.0161	0.9105
7	-84394.6663	168829.3325	169004.5263	0.0155	0.9164
8	-84471.9027	168989.8054	169191.2782	0.0227	0.7936

9	**-84377.1065**	**168806.2129**	**169033.9648**	**0.0134**	**0.9394**
10	-84374.6762	168807.3524	169061.3834	0.0134	0.9394

Table 5. Comparison results with different mixture model for wind direction

The fit results of different models are given in Figs. 2(a) and 2(b), it also shows that nine-component mixture model fits the frequency histogram of wind direction well than other models.

In Table 4, from the estimated model parameters of nine-component mixture model of wind direction, we can see that the parameters of μ, which correspond to the nine main wind directions, are 0.0010, 0.5544, 1.3416, 1.5989, 2.1458, 2.2854, 3.0405, 4.3874 and 5.6471 rad (or 0.06°, 31.77°, 76.87°, 91.61°, 122.94°, 130.94°, 174.21°, 251.38° and 323.55°), respectively. As shown in Fig. 2, the prevailing wind directions covering from 0° to 180°.

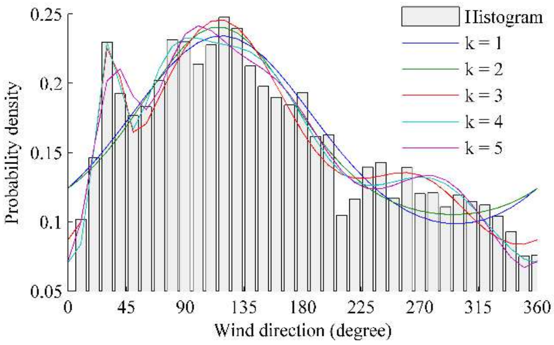

Figure 2. (a) Histogram of wind directions

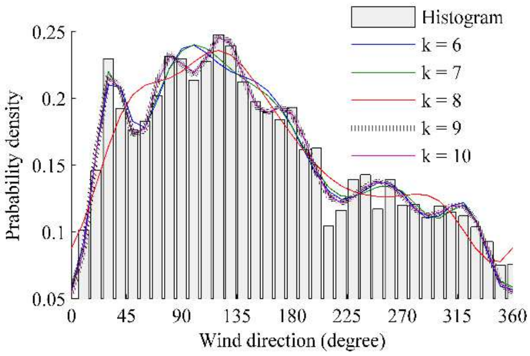

Figure 2. (b) Histogram of wind directions

This result is also confirmed by the wind rose diagram shown in Fig. 3 where the wind direction is generally divided into 16 different sections. According to the frequency of occurrence for wind direction, the nine main wind directions are classified into three kinds in descending order: the first four, the middle three and the last two main wind directions. We can see that the first four prevailing wind directions are the ESE, ENE, E and NNE directions with 9.39%, 8.92%, 8.72% and 7.95% of frequency of occurrence, respectively. In Table 4, the four estimated parameters μ of nine-component mixture model are 2.1458, 1.3416, 1.5989 and 0.5544 rad, which correspond to the wind directions are 122.94°, 76.87°, 91.61° and 31.77°, respectively. This result is close to the result given by Fig. 2 with the ESE, ENE, E and NNE directions (112.5°, 67.5°, 90° and 22.5°). The middle three predominant directions are the SE, SSE and N directions (135°, 157.5° and 0°) with 7.74%, 7.42% and 7.40% of occurrence. This result also agrees well with the result given by the estimated parameters μ = 2.2854, 3.0405 and 0.0010 rad (130.94°, 174.21° and 0.06°) in nine-component mixture model of wind direction. The last two main wind directions are the WSW and NW directions (247.5° and 315°) with 5.11% and 4.20% frequency. In the range from the directions N clockwise to S, including seven main wind directions, with 57.54% frequency totally. The other directions show an approximately uniform dispersion. Therefore, the yaw system of a wind turbine can be arranged to the ESE direction, since most of the wind blows from this direction, which will enable the wind turbine to be positioned in such a way as to maximize the captured energy.

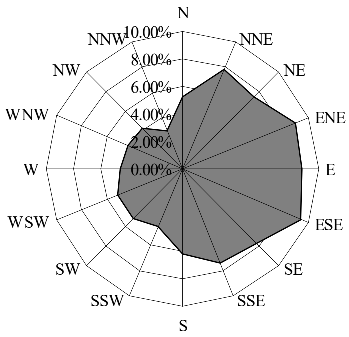

Figure 3. Wind rose diagram of wind directions

Using Eqs. (17) and (18), the estimated mean and variance of wind direction can be given as follows: b_1 = 2.8051 rad (160.72°) and b_2 = 0.1052, respectively. The relative error of b_1 is very small, it is only 2.99%.

At last, the parameter estimation results for the pdf $g(\xi)$ of circular variable ζ between the wind speed and direction using a voM mixture distribution with six components are given in Table 6.

The statistical association between the wind speed and wind direction is measured by the linear–circular correlation coefficient, r^2, given by[33,37,40]

$$r^2 = \frac{r_{vc}^2 + r_{vs}^2 - 2r_{vc}r_{vs}}{1 - r_{cs}^2} \tag{36}$$

where r_{vc} = corr(v, cosθ), r_{vs} = corr(v, sinθ) and r_{cs} = corr(cosθ, sinθ). The correlation coefficient between wind speed and direction should satisfy the requirement of $|r| < 1/3$. In this study, the value of r is small and equals to 0.1101, therefore, it can be seen that there exists a weak correlation between wind speed and direction. The absolute value of correlation coefficient is within 1/3 satisfying the condition of use for voM wind direction model. The estimated parameters of different components for circular variable between the wind speed and direction are given in Table 6.

k	***p***	***α***	***μ***	***k***	***p***	***α***	***μ***	***k***	***p***	***α***	***μ***
1	1.0000	0.1283	1.7772	5	0.4786	0.6301	4.6067	7	0.0271	25.9206	5.9223
2	0.0641	4.8654	1.1485		0.0350	18.8312	1.1307		0.0235	17.4631	1.4179
	0.9359	0.0777	2.8944		0.0295	17.4734	5.8814		0.0461	16.0716	0.9755
3	0.2179	1.2932	2.0530		0.0380	19.9723	0.5709		0.0384	19.1911	0.4917
	0.0673	6.6117	0.9532		0.4189	0.9275	1.9459		0.5399	0.5717	4.6824
	0.7148	0.2678	4.8013	6	**0.0271**	**25.9203**	**5.9236**		0.2986	1.1985	2.1294
4	0.3865	0.8576	4.6962		**0.0228**	**17.4639**	**1.4361**		0.0264	1.2281	2.1992
	0.4695	0.8295	2.1480		**0.0476**	**16.0592**	**0.9895**				
	0.0324	16.4754	5.8558		**0.0395**	**19.1996**	**0.4968**				
	0.1116	4.5147	0.8659		**0.5507**	**0.5568**	**4.6757**				
					0.3123	**1.2454**	**2.1411**				

Table 6. Parameter estimation results for circular variable between the wind speed and direction using different voM mixture distributions

Table 7 is a comparison results with different components for circular variable.

k	$\ln L$	AIC	BIC	RMSE	R^2
1	-86400.2443	172804.4887	172822.0081	0.0280	0.9693
2	-86396.3473	172802.6946	172846.4930	0.0258	0.9739
3	-86419.6016	172855.2032	172925.2808	0.0248	0.9758
4	-86428.6593	172879.3187	172975.6753	0.0238	0.9777
5	-86433.5052	172895.0104	173017.6461	0.0236	0.9781
6	**-86446.5538**	**172927.1075**	**173076.0222**	**0.0233**	**0.9785**
7	-86445.9581	172931.9162	173107.1100	0.0233	0.9785

Table 7. Comparison results with different components for circular variable

From Table 7, it can be found that increasing the number of voM components from one to six in the mixture distributions will increase the value of the R^2 coefficient. However, when the number of voM components increase to seven, the value of the R^2 coefficient does not increase, it is the same as the six-component mixture model has. At the same time, a seven-component mixture model has the same value of RMSE as a six-component mixture model. It is also noticed that six-component mixture model has a lower value of AIC and BIC, so based on comprehensive criteria of model selection, the most suitable model for the wind direction at this station is six-component VoM mixture model.

When considering the correlation between wind speed and direction, the number of parameters of μ, which correspond to the main wind directions, are decreased from 9 to 6. They are 0.4968, 0.9895, 1.4361, 2.1411, 4.6757 and 5.9236 rad (28.46°, 56.69°, 82.28°, 122.68°, 267.90° and 339.40°), respectively. Compared with the nine main wind directions (0.06°, 31.77°, 76.87°, 91.61°, 122.94°, 130.94°, 174.21°, 251.38° and 323.55°) without considering the correlation, we find an interesting phenomenon: the deleted three main wind directions (0.06°, 130.94° and 174.21°) are happened to be the middle three main wind directions analyzed in Table 4. A possible explanation is that the N and S directions, which correspond 0.06° and 174.21°, fall in the edge of the section of main wind directions ranging from the N direction clockwise to S direction, so the wind energy potential of these two directions is not significant than other parts of the section. On the other hand, 130.94° is approximately close to 122.94°, therefore, it is also deleted. At last, only six directions (31.77°, 76.87°, 91.61°, 122.94°, 251.38° and 323.55°) are left, and they are close to these six directions: 28.46°, 56.69°, 82.28°, 122.68°, 267.90° and 339.40°, as shown in Fig. 4. The maximum error does not exceed the value of one section (22.5°).

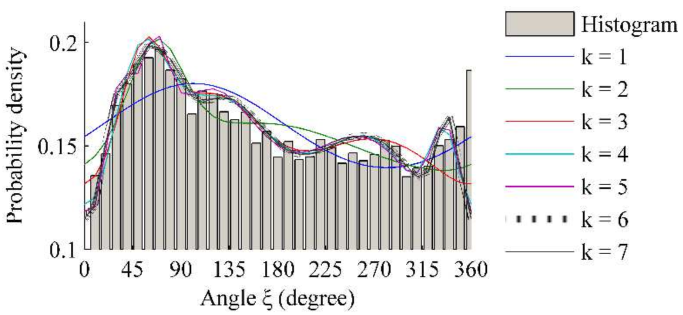

Figure 4. Histogram of circular variable

The joint pdf of wind speed and direction and wind power probability density are shown in Figs. 5 and 6.

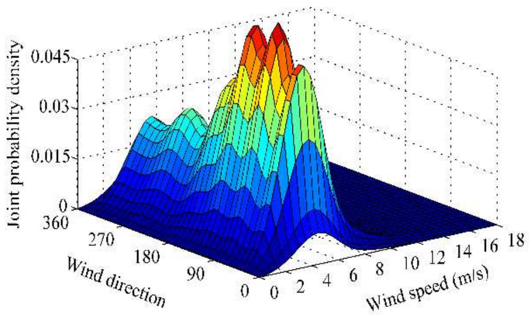

Figure 5. Joint probability density of wind speed and direction

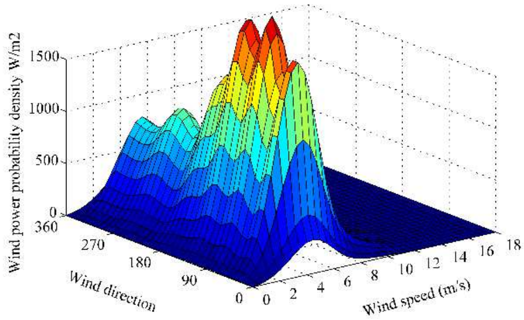

Figure 6. Wind power probability density

To estimate the wind power output of a wind turbine, it is necessary to know wind speed and the number of hours of the year, in which the wind blows at velocity v.

The scatter plot of wind power output versus wind speed for 1.8 MW wind turbine is shown in Fig. 7, after pre-processing with a data-cleaning method, some abnormal and unreasonable data are discarded. Using a polynomial regression method to fit the data of wind power, a mathematical expression of wind power output with speed can be obtained. Fig. 8 is the power curve of wind turbine.

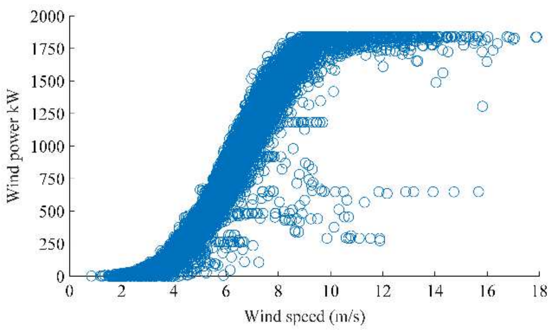

Figure 7. Scatter plot of wind power

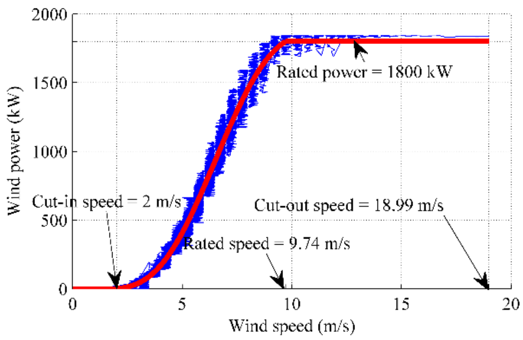

Figure 8. Wind power curve

The different fitting results are shown in Table 8, based on the correlation coefficient R^2, we select the eight-degree polynomial as the best model for wind power output.

	i **= 3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
a_0	190.9752	300.1130	50.3861	-71.7432	-29.7084	**8.2388**	0.9578	-9.6752
a_1	-269.9311	-382.1442	-22.8256	233.5924	103.3182	**-73.2569**	-23.2012	83.4538
a_2	83.2476	117.1197	-36.9182	-191.5619	-85.8043	**100.0191**	34.3568	-136.7091
a_3	-3.8966	-7.6637	19.2781	58.9500	22.0650	**-61.4883**	-24.9013	91.1207
a_4		0.1333	-1.8654	-6.6945	-0.2396	**19.1791**	8.4253	-33.6795
a_5			0.0517	0.3245	-0.2651	**-2.7679**	-0.9466	8.0854
a_6				-0.0057	0.0210	**0.2006**	0.0175	-1.1788
a_7					-0.0005	**-0.0072**	0.0036	0.1022
a_8						**0.0001**	-0.0002	-0.0051
a_9							0.0000	0.0001
a_{10}								0.0000
R^2	0.9851	0.9864	0.9915	0.9928	0.9930	0.9932	0.9933	0.9933

Table 8. The fitting results for wind power output with different degree i

The power curve of the studied wind turbine as shown in Fig. 8. The expression of wind power in W with wind speed in m/s is given as follows:

$$P(v) = 8238.7825 - 73256.9343v + 100019.0726v^2 - 61488.3128v^3 + 19179.0859v^4 - 2767.8827v^5 + 200.6144v^6 - 7.1645v^7 + 0.1006v^8 \text{ W} \tag{37}$$

When the wind speed arrives at the cut-in speed, the wind turbine starts to generate useable power. That is, if the wind speed is below the cut-in wind speed, then the wind turbine cannot generate electricity; At the rated output speed, the maximum output power is generated; while at the cut-out speed, the wind turbine was shut down to prevent damage caused by a strong wind. In this study, the cut-out wind speed (maximum allowable speed) and rated output power are known, they are 18.99 m/s and 1800 kW, respectively. The rated wind speed can be found by examining the power curve, i.e., the lowest wind speed, where the wind turbine first reaches its rated power, is the rated wind speed. Therefore, set Eq. (37) equals to 1.8×10^6 and 0, we can get the rated output speed, $v_r = 9.74$ m/s, and the cut-in speed, $v_{\text{in}} = 2$ m/s, respectively, as shown in Fig. 8.

In this study, the period t of one year is used, i.e., approximates 8760 h and the air density is 1.216 kg/m^3. Subsequently, the estimated values of wind power density at this wind farm were calculated using Eq. (33) is 93.27 W/m^2. This value is close to the reference value of mean wind power density of 88.14 W/m^2 which is obtained using Eq. (31). It indicates that the region under study stands in class 1[10], which belongs to the low-wind speed wind power development area, and is generally not suitable for wide wind turbine establishment and wind farm investment. However, it would be possible to exploit the wind power applications for small scale wind turbines at this area. Based on Eq. (32), a comparison result without taking into account of wind direction is 125.78 W/m^2, is also presented. It is clearly shown that the wind energy potential is overestimated without considering the effect of wind direction. Using Eq. (35), the annual energy output can also be obtained, it is 2.21 GWh.

Conclusions

In this paper, the wind energy potential of the Maling Mountain in China is studied by using the measured wind data for a period of one year at a height of 30 m. Based on EM algorithm, an assessment method of wind energy potential using finite mixture statistical distribution model is proposed, the probability density function of wind power density and the annual energy output are given for use in wind energy analyses. The suitability of the model is judged using a comprehensive technique of model selection including AIC, BIC information criteria, coefficient of determination R^2 and RMSE. Field monitoring wind data are used to verify the effectiveness and validity of the proposed method by comparing it with other estimation methods in accordance with the value of R^2 and RMSE, the histogram plot and wind rose diagram. The main conclusions are drawn from the study as follows:

1. The proposed method takes into account the effect of wind speed and direction simultaneously, the correlation existing between both variables, as well as the bimodal or multimodal distributions of them. The mixture distribution model provides a better fitting result for wind data than single distribution model, and it can therefore be used in the assessment of the wind energy at a potential site and the assessment result is more accuracy and close to the reality. On the other hand, three-parameter Weibull distribution considers the frequency of calm winds, it shows a good fit to wind speed data with a significant null wind speed or high percentage of low wind speed and is particularly suitable for a skewed data with a long tail in histogram plot.

2. The best mixture model with the lowest AIC and BIC values is selected as the optimal model

from all candidate models with a finite component number, therefore it is not necessary to know the number of components in mixture model in advance. Increasing the number of components of the mixture distributions, the variations in value of R^2 are not significant. On the contrary, it might yield complex. Therefore, combining with the comprehensive technique of model selection, an appropriate number of mixture distributions can be selected, which not only reduces the computational burden but also improves the model accuracy, and the model has a higher predictive ability.

3. Compared with the real mean wind power density of time series wind speed data, it also shown that when there exists a correlation between wind speed and its direction, the estimated results of wind energy potential is more close to the real situation when considering the influence of wind direction.

References

1. Ogulata, R. T. Energy sector and wind energy potential in Turkey. *Renew. Sustain. Energy Rev*. **7**, 469–484 (2003).
2. Eskin, N., Artar, H. & Tolun, S. Wind energy potential of Gokceada Island in Turkey. *Renew. Sustain. Energy Rev*. **12**, 839–851 (2008).
3. Philippopoulos, K., Deligiorgi, D. & Karvounis, G. Wind speed distribution modeling in the greater area of Chania, Greece. *Int. J. Green Energy* **9**, 174–193 (2012).
4. Wais, P. A review of Weibull functions in wind sector. *Renew. Sustain. Energy Rev*. **70**, 1099–1107 (2017).
5. Kiss, P. & Janosi, I. M. Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe. *Energy Convers. Manag*. **49**, 2142–2151 (2008).
6. Aslam, M. Testing average wind speed using sampling plan for Weibull distribution under indeterminacy. *Sci. Rep.* **11**, 7532-1-9 (2021).
7. Chen, H., Birkelund, Y., Anfinsen, S. N., Staupe-Delgado, R. & Yuan, F. Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic. *Sci. Rep.* **11**, 7613-1-11 (2021).
8. Hu, Q., Wang, Y., Xie, Z., Zhu, P. & Yu, D. On estimating uncertainty of wind energy with mixture of distributions. *Energy* **112**, 935–962 (2016).
9. Aries, N., Boudia, S. M. & Ounis, H. Deep assessment of wind speed distribution models: A case study of four sites in Algeria. *Energy Convers. Manag*. **155**, 78–90 (2018).
10. Pishgar-Komleh, S. H., Keyhani, A. & Sefeedpari, P. Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooh county of Iran). *Renew. Sustain. Energy Rev*. **42**, 313–322 (2015).
11. Akdag, S. A. & Dinler, A. A new method to estimate Weibull parameters for wind energy applications. *Energy Convers. Manag*. **50**, 1761–1766 (2009).
12. Kantar, Y. M. & Usta, I. Analysis of the upper-truncated Weibull distribution for wind speed. *Energy Convers. Manag*. **96**, 81–88 (2015).
13. Wais, P. Two and three-parameter Weibull distribution in available wind power analysis. *Renew. Energy* **103**, 15–29 (2017).
14. Deep, S., Sarkar, A., Ghawat, M. & Rajak, M. K. Estimation of the wind energy potential for coastal locations in India using the Weibull model. *Renew. Energy* **161**, 319–339 (2020).
15. Carta, J. A. & Mentado, D. A continuous bivariate model for wind power density and wind turbine energy output estimations. *Energy Convers. Manag*. **48**, 420–432 (2007).

16. Carta, J. A. & Ramirez, P. Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. *Energy Convers. Manag*. **48**, 281–291 (2007).
17. Carta, J. A. & Ramirez, P. Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. *Renew. Energy* **32**, 518–531 (2007).
18. Akpinar, S. & Akpinar, E. K. Estimation of wind energy potential using finite mixture distribution models. *Energy Convers. Manag*. **50**, 877–884 (2009).
19. Akdag, S. A., Bagiorgas, H. S. & Mihalakakou, G. Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. *Appl. Energy* **87**, 2566–2573 (2010).
20. Zhang, J., Chowdhury, S., Messac, A. & Castillo, L. A multivariate and multimodal wind distribution model. *Renew. Energy* **51**, 436–447 (2013).
21. Ouarda, T. B. M. J. *et al*. Probability distributions of wind speed in the UAE. *Energy Convers. Manag*. **93**, 414–434 (2015).
22. Mahbudi, S., Jamalizadeh, A. & Farnoosh, R. Use of finite mixture models with skew-t-normal Birnbaum-Saunders components in the analysis of wind speed: Case studies in Ontario, Canada. *Renew. Energy* **162**, 196–211 (2020).
23. Mazzeo, D., Oliveti, G. & Labonia, E. Estimation of wind speed probability density function using a mixture of two truncated normal distributions. *Renew. Energy* **115**, 1260–1280 (2018).
24. Ouarda, T. B. M. J. & Charron, C. On the mixture of wind speed distribution in a Nordic region. *Energy Convers. Manag*. **174**, 33–44 (2018).
25. Gugliani, G. K., Sarkar, A., Ley, C. & Mandal, S. New methods to assess wind resources in terms of wind speed, load, power and direction. *Renew. Energy* **129**, 168–182 (2018).
26. Han, Q. & Chu, F. Directional wind energy assessment of China based on nonparametric copula models. *Renew. Energy* **164**, 1334–1349 (2021).
27. Carta, J. A., Bueno, C. & Ramirez, P. Statistical modelling of directional wind speeds using mixtures of von Mises distributions: Case study. *Energy Convers. Manag*. **49**, 897–907 (2008).
28. Masseran, N., Razali, A. M., Ibrahim, K. & Latif, M. T. Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia. *Energy Convers. Manag*. **72**, 94–102 (2013).
29. Zou, M. *et al*. Evaluation of wind turbine power outputs with and without uncertainties in input wind speed and wind direction data. *IET Renew. Power Gener*. **14**, 2801–2809 (2020).
30. Soukissian, T. H. Probabilistic modeling of directional and linear characteristics of wind and sea states. *Ocean Eng.* **91**, 91–110 (2014).
31. Horn, J., Gregersen, E. B., Krokstad, J. R., Leira, B. J. & Amdahl, J. A new combination of conditional environmental distributions. *Appl. Ocean Res*. **73**, 17–26 (2018).
32. Vanem, E., Hafver, A. & Nalvarte, G. Environmental contours for circular-linear variables based on the direct sampling method. *Wind Energy* **23**, 563–574 (2020).
33. Carta, J. A., Ramirez, P. & Bueno, C. A joint probability density function of wind speed and direction for wind energy analysis. *Energy Convers. Manag*. **49**, 1309–1320 (2008).
34. Carta, J. A., Ramirez, P. & Velazquez, S. A review of wind speed probability distributions used in wind energy analysis Case studies in the Canary Islands. *Renew. Sustain. Energy Rev*. **13**, 933–955 (2009).
35. Erdem, E. & Shi, J. Comparison of bivariate distribution construction approaches for analysing wind speed and direction data. *Wind Energy* **14**, 27–41 (2011).
36. Ovgor, B., Lee, S. K. & Lee, S. A method of micrositing of wind turbine on building roof-top

by using joint distribution of wind speed and direction, and computational fluid dynamics. *J. Mech. Sci. Technol*. **26**, 3981–3988 (2012).

37. Soukissian, T. H. & Karathanasi, F. E. On the selection of bivariate parametric models for wind data. *Appl. Energy* **188**, 280–304 (2017).
38. Han, Q., Hao, Z., Hu, T. & Chu, F. Non-parametric models for joint probabilistic distributions of wind speed and direction data. *Renew. Energy* **126**, 1032–1042 (2018).
39. Ye, X. W., Xi, P. S. & Nagode, M. Extension of REBMIX algorithm to von Mises parametric family for modeling joint distribution of wind speed and direction. *Eng. Struct*. **183**, 1134–1145 (2019).
40. Li, H. N., Zheng, X. W. & Li, C. Copula-based joint distribution analysis of wind speed and direction. *J. Eng. Mech*. **145**, 04019024–1–12 (2019).
41. 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. *Energy Convers. Manag*. **49**, 962–973 (2008).
42. Zhang, H., Yu, Y. J. & Liu, Z. Y. Study on the Maximum Entropy Principle applied to the annual wind speed probability distribution: A case study for observations of intertidal zone anemometer towers of Rudong in East China Sea. *Appl. Energy* **114**, 931–938 (2014).
43. Miao, S. *et al*. A mixture kernel density model for wind speed probability distribution estimation. *Energy Convers. Manag*. **126**, 1066–1083 (2016).
44. Han, Q., Ma, S., Wang, T. & Chu, F. Kernel density estimation model for wind speed probability distribution with applicability to wind energy assessment in China. *Renew. Sustain. Energy Rev*. **115**, 109387–1–14 (2019).
45. Guan, J. S., Lin, J., Guan, J. J. & Mokaramian, E. A novel probabilistic short-term wind energy forecasting model based on an improved kernel density estimation. *Int. J. Hydrog. Energy* **45**, 23791–23808 (2020).
46. Wahbah, M., Feng, S. F., EL-Fouly, T. H. M. & Zahawi, B. Wind speed probability density estimation using root-transformed local linear regression. *Energy Convers. Manag*. **199**, 111889–1–12 (2019).
47. Ye, X. W., Ding, Y. & Wan, H. P. Statistical evaluation of wind properties based on long-term monitoring data. *J. Civ. Struct. Health* **10**, 987–1000 (2020).
48. Gungor, A., Gokcek, M., Ucar, H., Arabaci, E. & Akyüz, A. Analysis of wind energy potential and Weibull parameter estimation methods: a case study from turkey. *Int. J. Environ. Sci. Te.* **17**, 1011–1020 (2020).
49. Yan, J., Zhang, H., Liu, Y., Han, S. & Li, L. Uncertainty estimation for wind energy conversion by probabilistic wind turbine power curve modelling. *Appl. Energy* **239,** 1356–1370 (2019).
50. Wang, L., Liu, J. & Qian, F. Wind speed frequency distribution modeling and wind energy resource assessment based on polynomial regression model. *Int. J. Electr. Power Energy Syst.* **130**, 106964–1–12 (2021).

Acknowledgements

We would like to thank the staff of the Gansu Province Special Equipment Inspection and Testing Institute for providing wind data used in this study.

Author contributions statement

W.Z. is responsible for developing methods, drawing charts, analyzing results, and writing the

original draft. L.W. is responsible for data curation and methodology improvements.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W. Z.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Winddata.xls