Environmental Research 160 (2018) 365–371

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

A new method for wind speed forecasting based on copula theory

Yuankun Wang[a], Huiqun Ma[b], Dong Wang[a,*], Guizuo Wang[c], Jichun Wu[a], Jinyu Bian[d], Jiufu Liu[d]

[a] Key Laboratory of Surficial Geochemistry, Ministry of Education, Department of Hydrosciences, School of Earth Sciences and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing 210023, China

[b] Shandong Electric Power Engineering Consulting Institute Corp, Ltd., Jinan 250013, China

[c] Development Research Center of the Ministry of Water Resources, Beijing 100038, China

[d] Nanjing Hydraulic Research Institute, Nanjing 210029, China

ARTICLE INFO

Keywords:
Wind farm
Wind speed
Copula method
Conditional distribution

ABSTRACT

How to determine representative wind speed is crucial in wind resource assessment. Accurate wind resource assessments are important to wind farms development. Linear regressions are usually used to obtain the representative wind speed. However, terrain flexibility of wind farm and long distance between wind speed sites often lead to low correlation. In this study, copula method is used to determine the representative year's wind speed in wind farm by interpreting the interaction of the local wind farm and the meteorological station. The result shows that the method proposed here can not only determine the relationship between the local anemometric tower and nearby meteorological station through Kendall's tau, but also determine the joint distribution without assuming the variables to be independent. Moreover, the representative wind data can be obtained by the conditional distribution much more reasonably. We hope this study could provide scientific reference for accurate wind resource assessments.

1. Introduction

Wind energy was introduced as an energy alternative source in 1970s and grew faster than any other energy sources in 1990s. In 2003, more than 68,000 wind turbines with a capacity of 40, 300 MW had been installed all over the world (Zarma, 2005; Razavieh et al., 2014). China will adjust its energy structure, promote industrial upgrading and develop non-fossil energy technologies (Duan, 2017), such as wind power and solar power. In the recent years, the government of China has encouraged the construction of wind farm while limiting the coal-fire power plants construction. More and more wind farms have been installed in the last ten years (Xia and Song, 2009). In feasible research of wind farm construction, wind resources assessment is an important process. The grade of wind resources is the crucial qualification in the construction. It determines whether the wind farm is profitable or not. The best method to determine the wind energy potential s is to measure the wind on site for several years. However, it is time-consuming. One would like to calculate the wind power with atmospheric models. Although it is not as exact as direct measurements, it is expected that favorable and unfavorable regions can be distinguished using numerical simulations.

Generally, one or more anemometric tower will be built in the wind farm. After more than an entire year's observation, wind resources assessment can be carried through by the short-term data and 30 year's data nearby meteorological station (Du and Feng, 2010). The observed data can only represent the wind resources status of the given year not for a long-term time series. Because wind is highly variable every year, short-term onsite measurements can result in highly inaccurate energy estimates (Yang et al., 2015). Therefore, wind speed data from nearby long-term meteorological stations are used to adjust the onsite data. Traditional methods for determining the representative wind speed is using the liner correlation between the observed data and the same term's data of the nearby meteorological station. However, the correlation is not always good as a result of terrain and distance between the wind farm and meteorological station. How to determine the representative wind speed with poor correlation is still unresolved.

Multivariate analyses, such as factor analysis and multivariate regression analysis, have been proved useful for understanding the relationships among meteorological factors in highly complex systems. In fact, the interaction of these variables is highly complex. For example, determination of the joint distribution of variables and conditional distribution are important for wind speed analysis (Feng et al., 2015), but they cannot be evaluated by the previous methods effectively. Correlation analysis, to some extent, reflects the relationship between different variables. However, it cannot fully represent the inherent relationship of the variables in a complex system. Thus, a more effective

* Corresponding author.
E-mail address: wangdong@nju.edu.cn (D. Wang).

http://dx.doi.org/10.1016/j.envres.2017.09.034
Received 4 August 2017; Received in revised form 14 September 2017; Accepted 29 September 2017

tool is necessary for better understanding the interaction of different climatic variables. To accomplish this goal, copulas are employed in this study to characterize the inherent relationship of the wind variables.

Copulas proposed by Sklar (1959) offer an effective method for the dependence and frequency analysis of random variables. Couplas methods have been widely used in many different fields (Patton, 2006; Zhang and Singh, 2007; Lzel and Friederichs, 2008; Haghi et al., 2010; Wang et al., 2012; Ganguli and Reddy, 2014; Wu, 2014; Tiwari and Singh, 2015). Many studies also have been done in wind power with copulas theory (Sun et al., 2016; Jia and Zheng, 2017; Pircalabu et al., 2017). Haghi et al. (2010) used Archimedean copulas to investigate combination of photovoltaics and wind turbines distributed in a distribution network based on the stochastic modeling. Sun et al. (2016) proposed a fuzzy copula model to express the possibilistic uncertainty of wind speed correlation. Bessa et al. (2014) developed a novel time-adaptive quantile-copula estimator for kernel density forecast and a discussion of how to select the adequate kernels for modeling the different variables of the problem. In order to achieve the increased sample size for the subsequent Monte Carlo simulation, Hagspiel et al. (2012) applied copula theory to generate a synthetic set of data from scarce wind speed reanalysis data. Veeramachaneni et al. (2015) proposed multivariate copulas for modeling multiple joint distributions of wind speeds at a wind farm site and neighboring wind source, and results show that a n-dimensional Gaussian copula and multiple copula graphical models enhance the quality of the prediction. If there is no neighboring wind source available, then how to determine representative wind speed, the previous studies did not answer this question.

The objective of this study is to apply the copula theory to develop a method to determine the representative year's wind speed in wind farm by analyzing the relationship between the wind speeds of anemometer tower in wind farm and that of meteorological station. A case study is presented for a wind farm in Laixi City, Shandong Province, China.

2. Theory and methods

2.1. Pearson correlation coefficient

Suppose that a random sample (u_1, v_2), …, (u_n, v_n) is given from some pair (u, v) of continuous variables. Then, the well-known non-parametric measures of dependence, Pearson correlation coefficient can be got:

$$\rho = \frac{\sum_{i=1}^{n}(u_i-\bar{u})(v_i-\bar{v})}{\sqrt{\sum_{i=1}^{n}(u_i-\bar{u})^2\sum_{i=1}^{n}(v_i-\bar{v})^2}} \in [-1, 1] \tag{1}$$

where

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i \tag{2}$$

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i \tag{3}$$

This parameter shows the linear relationship between two variables, namely, only if U and V are linear functions of one another, $\rho = \pm 1$.

2.2. Kendall's Tau

Kendall's tau is another well-known measure of dependence. It is based on ranks, whose empirical expression is given by

$$\tau_N = \binom{N}{2}^{-1}(P_n - Q_n) \tag{4}$$

where P_n means number of concordant pairs, $P_n = P_n + 1$ if $(u_i - u_j)(v_i - v_j) \geq 0$; In contrast, Q_n means number of discordant pairs, $Q_n = Q_n + 1$ if $(u_i - u_j)(v_i - v_j) < 0$. N is the number of observations; τ_N is the estimate of τ from observations (Wang et al., 2012).

2.3. Copula theory

Restricting attention to the bi-variable in the interest of simplification, the copula approach is rooted in a representation theorem due to Sklar (1959). The joint cumulative distribution function (CDF), $H(u, v)$ of any pair U, V of continuous random variables may be written in the following form

$$H(u, v) = C(F(u), G(v)) \quad u, v \in R \tag{5}$$

where $F(u)$ and $G(v)$ are marginal distributions; C is copula.

In this study, a parametric family C_θ of copulas is being considered as a model for the dependence between two random variables U, V. Archimedean copulas are used because of the explicit relation between the parameter θ and the population value τ of Kendall's tau.

In classical statistics, maximum likelihood estimation is a well-known method that is usually more efficient. The method of maximum pseudo-likelihood, which requires that C_θ be absolutely continuous with density c_θ. The estimate of θ is obtained through the maximization of a function of the form

$$l(\theta) = \sum_{i=1}^{n} \log[c_\theta(\hat{F}(u), \hat{G}(v))] \tag{6}$$

Letting $\dot{c}_\theta(u, v) = \partial c_\theta(u, v)/\partial\theta$, under mild regularity conditions that the root $\hat{\theta}_n$ of the equation can be got

$$\dot{l}(\theta) = \frac{\partial}{\partial\theta}l(\theta) = \sum_{i=1}^{n}\frac{\dot{c}_\theta(u_i, v_i)}{c_\theta(u_i, v_i)} = 0 \tag{7}$$

2.4. Conditional copula

The conditional distribution using the copula method can be expressed. Let X and Y be random variables with $u = F_X(x)$, $v = F_Y(y)$. As an example, the conditional distribution function of X given $Y = y$ can be expressed by the copula method as

$$\begin{aligned} F(X \leq x|Y = y) &= C_\theta(u|V = v) \\ &= \lim_{\Delta v \to 0}\frac{C_\theta(u, v + \Delta v) - C_\theta(u, v)}{\Delta v} \\ &= \frac{\partial}{\partial v}C_\theta(u, v)|V = v \end{aligned} \tag{8}$$

Similarly, an equivalent formula for the conditional distribution function for Y *given* $X = x$ can be obtained. Furthermore, the conditional distribution function of X given $Y \leq y$ can be expressed by the copula method as

$$F(X \leq x|Y \leq y) = C_\theta(u|V \leq v) = \frac{C_\theta(u, v)}{v} \tag{9}$$

Likewise, an equivalent formula for the conditional distribution function for Y given $X \leq x$ can be obtained (Zhang and Singh, 2007).

3. Applications

Shandong is part of the East China region (Fig. 1), one of the developed provinces of China, is the biggest industrial producer and one of the top manufacturing provinces in China. To protect air environment quality, Shandong government has been devoted to reducing the proportion of thermal power in energy consumption and enlarging the clean energy, such as wind power. Clean energy utilization increased from 127 million kilowatts in 2009 to 1317 million kilowatts in 2015 (SPBS, 2016). Wind energy resources of inland areas and offshore are very rich. At the same time, it was rarely affected by typhoon and