

A Multivariate and Multimodal Wind Distribution model

Jie Zhang^a, Souma Chowdhury^a, Achille Messac^{b,*}, Luciano Castillo^c

^a Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

^b Department of Mechanical and Aerospace Engineering, Syracuse University, Syracuse, NY 13244, USA

^c Mechanical Engineering Department, Texas Tech University, Lubbock, TX 79409, USA

ARTICLE INFO

Article history:

Received 28 October 2011

Accepted 22 September 2012

Available online 6 November 2012

Keywords:

Energy
Kernel density estimation
Multimodal
Offshore
Wind distribution
Wind power density

ABSTRACT

This paper presents a new methodology to accurately characterize and predict the annual variation of wind conditions. The estimate of the distribution of wind conditions is necessary to quantify the available energy (power density) at a site, and to design optimal wind farm configurations. A smooth multivariate wind distribution model is developed to capture the coupled variation of wind speed, wind direction, and air density. The wind distribution model developed in this paper avoids the limiting assumption of unimodality of the distribution. This method, which we call the *Multivariate and Multimodal Wind Distribution* (MMWD) model, is an evolution from existing wind distribution modeling techniques. *Multivariate kernel density estimation*, a standard non-parametric approach to estimate the probability density function of random variables, is adopted for this purpose. The MMWD technique is successfully applied to model (i) the distribution of wind speed (univariate); (ii) the joint distribution of wind speed and wind direction (bivariate); and (iii) the joint distribution of wind speed, wind direction, and air density (multivariate). The latter is a novel contribution of this paper, while the former offers opportunities for validation. Both onshore and offshore wind distributions are estimated using the MMWD model. Recorded wind data, obtained from the *North Dakota Agricultural Weather Network* (NDAWN) and the *National Data Buoy Center* (NDBC), is used in this paper. The coupled distribution was found to be multimodal. A strong correlation among the wind condition parameters was also observed.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past five years, the worldwide installed wind capacity has been growing at an average rate of 27.8% per year [1]. The available energy from a wind resource varies appreciably over one year. The uncertainty in wind resource potential is a constraining factor in the growth of wind energy market share. Accurate determination of energy variability at a site serves these important objectives: (i) accurate assessment of the potential effectiveness of a wind farm site, (ii) effective design of the wind farm layout, and (iii) appropriate selection of turbine types for the site.

Wind Power Density (WPD) is a useful way to evaluate the wind resource available at a potential site. The WPD, measured in *watts per square meter*, indicates how much energy is available at the site. WPD (W/m^2) is a nonlinear function of the *probability density function* (*pdf*) of wind velocity and air density, which is expressed as

$$\text{WPD} = \int_{0^\circ}^{360^\circ} \int_0^{U_{\max}} \int_{\rho_{\min}}^{\rho_{\max}} \frac{1}{2} \rho U^3 f(U, \theta, \rho) d\rho dU d\theta \quad (1)$$

where U and θ represent the wind speed and wind direction, respectively; U_{\max} is the maximum possible wind speed at that location; ρ represents the air density; ρ_{\min} and ρ_{\max} are the maximum and minimum air density in that location, respectively; and $f(U, \theta, \rho)$ is the *pdf* of the wind condition (speed, direction and air density).

The most widely used distribution for the characterization of wind speed is the 2-parameter Weibull distribution [2–7]. Other distributions used to characterize wind speed include 1-parameter Rayleigh distribution, 3-parameter generalized Gamma distribution, 2-parameter Lognormal distribution, 3-parameter Beta distribution, 2-parameter inverse Gaussian distribution, singly truncated normal Weibull mixture distribution, and the maximum entropy probability density function [4,7].

1.1. Research objectives and motivation

Wind energy sources appear in the form of wind farms that consist of multiple wind turbines located in an arrangement over

* Corresponding author. Tel.: +1 315 443 2341.

E-mail address: messac@syrr.edu (A. Messac).

Nomenclature

U	wind speed (m/s)
θ	wind direction ($^{\circ}$)
U_{\max}	maximum possible wind speed (m/s)
ρ	air density (kg/m^3)
ρ_{\max}	maximum air density (kg/m^3)
ρ_{\min}	minimum air density (kg/m^3)
$K(\cdot)$	Kernel function
h	the bandwidth of kernel density estimation
H	the bandwidth matrix of multivariate kernel density estimation
$MISE$	Mean Integrated Squared Error
U_m	measured wind speed (m/s)
z_0	the average roughness length in the farm region (m)
p	the absolute pressure (Pa)
h_a	the altitude above sea level (m)
R^2	coefficient of determination

a substantial stretch of land (onshore), or water body (offshore). Sorensen and Nielsen [8] showed that the total power extracted by a wind farm is significantly less than the product of the power extracted by a stand-alone turbine and the number of wind turbines in the farm. This difference is attributed to the loss in the availability of energy due to wake effects – the mutual shading effect of wind turbines [9]. An optimal layout of turbines that maximizes farm efficiency is of utmost importance in conceiving a wind farm project, and increasing the wind energy market in general.

For a given farm layout, the direction of the wind has a strong influence on the wakes created, and subsequently on the overall flow pattern in the wind farm. A bivariate distribution of the wind speed and wind direction would be helpful for the wind farm layout optimization. Lackner and Elkinton [10] characterized the wind speed data by direction sector and fitted a Weibull distribution for each direction sector. Vega and Letchford [11] used Weibull distribution to estimate the wind speed probability, and modeled the shape parameter and the scale parameter as functions of wind direction. Carta et al. [12] presented a joint probability density function of wind speed and wind direction for wind energy analysis. Erdem and Shi [13] compared three differing bivariate joint distributions (angular-linear, Farlie-Gumbel-Morgenstern, and anisotropic lognormal approaches) to represent wind speed and wind direction data.

These existing wind distribution modeling approaches can be broadly classified into: (i) univariate and unimodal distributions of wind speed (such as Weibull, Rayleigh, and Gamma distributions), and (ii) bivariate and unimodal distributions of wind speed and wind direction [11–13]. *These wind distribution models make limiting assumptions regarding the correlativity and the modality of the distribution of the wind.* These assumptions lead to approximations that deviate significantly from the actual scenario. In addition, it is seen from Eq. (1) that the WPD is directly proportional to the air density. For the real life case study (a site in North Dakota [14]) in this paper, the annual variation in air density is estimated to be 30%. Neglecting such an appreciable variation (in air density), by assuming a constant air density value, can lead to significant errors in the predicted power available at a wind site. *Therefore, we believe that a robust multivariate probability distribution of wind speed, wind direction and air density can address the above limiting assumptions.* To the best of the authors' knowledge, such a wind distribution model is yet to be found in the literature.

In addition, such a wind distribution model is crucial for wind resource assessment and wind farm layout optimization [15–17].

In this paper, a new methodology to represent the multivariate (and possibly multimodal) distribution of wind conditions is developed and explored. *Multivariate kernel density estimation* [18] has been adopted to develop the distribution. The remainder of the paper is organized as follows. The *Multivariate and Multimodal Wind Distribution* (MMWD) model is developed in Section 2. The ten-year wind data used for testing the new MMWD distribution is provided in Section 3. Section 4 presents the results and discussion for the three scenarios studied. Concluding remarks and future work are given in the last Section.

2. Multivariate and Multimodal Wind Distribution (MMWD) model

The *Kernel Density Estimation* (KDE) method [18] is adopted to represent the distribution of the wind conditions in this paper. The following assumptions are made in the development of the KDE-based MMWD model.

1. This model does not account for the frequency of calms. At a particular wind site, the frequency of calms may vary 15–20% over a year [19].
2. Wind direction is a periodic variable. The probability of the wind coming from the 0° direction should be equal to that coming from the 360° direction. This attribute is not captured in the KDE-based model.

The MMWD model presented in this paper makes unique contributions as follows.

1. This model is capable of representing multimodal wind data, which is rare in the wind distribution literature.
2. This model can capture the joint variations of wind speed, wind direction and air density.

2.1. Kernel density estimation (KDE)

KDE, also known as the Parzen-Rosenblatt window method [20,21], is a non-parametric approach to estimate the *pdf* of a random variable. For an independent and identically distributed sample, x_1, x_2, \dots, x_n , drawn from some distribution with an unknown density f , the KDE is defined to be [22]

$$\hat{f}(x; h) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2)$$

In the equation, $K(\cdot) = (1/h)K(\cdot/h)$ for a *kernel function* K (often taken to be a symmetric probability density) and a *bandwidth* h (the smoothing parameter).

2.2. Multivariate kernel density estimation

For a d -variate random sample X_1, X_2, \dots, X_n drawn from a density f , the multivariate KDE is defined to be

$$\hat{f}(x; H) = n^{-1} \sum_{i=1}^n K_H(x - X_i) \quad (3)$$

where $x = (x_1, x_2, \dots, x_d)^T$ and $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$, $i = 1, 2, \dots, n$. Here, $K(x)$ is the kernel that is a symmetric probability density function, H is the bandwidth matrix which is symmetric and positive-definite,

and $K_H(x) = |H|^{-1/2}K(H^{-1/2}x)$. The choice of K is not crucial to the accuracy of kernel density estimators [23]. In this paper, $K(x) = (2\pi)^{-d/2}\exp(-\frac{1}{2}x^T x)$ is considered throughout. In contrast, the choice of H is crucial in determining the performance of \hat{f} [24].

2.3. Optimal bandwidth matrix selection

The most commonly used optimality criterion for selecting a bandwidth matrix is the Mean Integrated Squared Error (MISE) [24], which is expressed as

$$MISE(H) = E \int [\hat{f}(x; H) - f(x)]^2 dx \tag{4}$$

It is usual to employ an asymptotic approximation, known as the AMISE (Asymptotic MISE) [24], which is expressed as

$$AMISE(H) = n^{-1}(4\pi)^{-d/2}|H|^{-1/2} + \frac{1}{4}(vec^T H)\psi_4(vecH) \tag{5}$$

where vec is the vector half operator, given by

$$vecH = vec \begin{bmatrix} h_1^2 & h_{12} \\ h_{12} & h_2^2 \end{bmatrix} = \begin{bmatrix} h_1^2 \\ h_{12} \\ h_2^2 \end{bmatrix}$$

The general expression of ψ_4 can be found in the paper by Wand and Jones [25]. An ideal optimal bandwidth is estimated to be

$$H_{AMISE} = arg \min_H AMISE(H) \tag{6}$$

In this paper, the multivariate plug-in selector (PI(H)) developed by Wand and Jones [26] is used, which is given by

$$PI(H) = n^{-1}(4\pi)^{-d/2}|H|^{-1/2} + \frac{1}{4}(vec^T H)\hat{\psi}_4(vecH) \tag{7}$$

The plug-in estimate of the AMISE can be numerically minimized to give the plug-in bandwidth matrix, \hat{H}_{PI} .

2.4. Illustrating the multimodality of the wind distribution

As briefly discussed in the previous section, the state of the art in wind distribution estimation makes limiting assumptions regarding the correlativity and the modality of the distribution of the wind. Two sample sites are used to explore this issue and illustrate the multimodality of the wind distribution. These sites are: (1) an onshore site at the Baker wind station in North Dakota,

and (2) an offshore site at the wind Station 44013-BOSTON, 16 NM East of Boston, Massachusetts. The details of the wind data for the two sites are presented in Section 3.

The histograms of the wind speed and wind direction at the two sites are illustrated in Fig. 1(a) and (b). It is observed from Fig. 1(a) that the distribution of wind speed and direction at the onshore site is multimodal. Therefore, the MMWD model proposed in this paper can prove to be uniquely helpful in capturing this multimodal characteristic of the wind distribution and provide more credibility to the available energy estimates. Fig. 1(b) illustrates that the distribution of wind speed and direction at the offshore site is practically unimodal.

2.5. Applicability of kernel density estimation to wind distribution

The literature shows that the choice of the bandwidth (or bandwidth matrix) is much more important for the behavior of $\hat{f}(x; h)$ (or $\hat{f}(x; H)$) than the choice of K . Small values of h (or H) will introduce spurious oscillation into the estimation; while too large a value of h (or H) will lead to too smooth an estimate, which may not reveal the actual features, such as multimodality [27].

As illustrated in the previous subsection, the distribution of wind condition could be multimodal or unimodal for different wind farm sites. Therefore, the selection of the value of h (or H) would affect the accuracy of estimated wind distribution. The histograms of wind conditions (speed, direction, air density) could roughly show the trend of distribution (Fig. 1), which is useful to assess the accuracy of the developed MMWD model.

3. Wind condition data

Both onshore and offshore wind distributions are estimated using the MMWD model later in Section 4. The onshore wind data used in this paper is obtained from the North Dakota Agricultural Weather Network (NDAWN) [14]. The daily averaged data for wind speed, wind direction, and air temperature measured at the Baker station (Fig. 2(a)) between the year 2000 and 2009 is used. The offshore wind data is obtained from the National Data Buoy Center (NDBC) [28]. The daily averaged data is measured at the Station 44013-BOSTON 16 NM East of Boston (Fig. 2(b)) between the year 2000 and 2009. Table 1 shows the geographical information of the two stations. The measurement information is listed as follows.

1. Wind speed is measured at 3 m above the soil surface at the onshore station (Baker), and at 5 m above the sea level at the offshore station (44013).

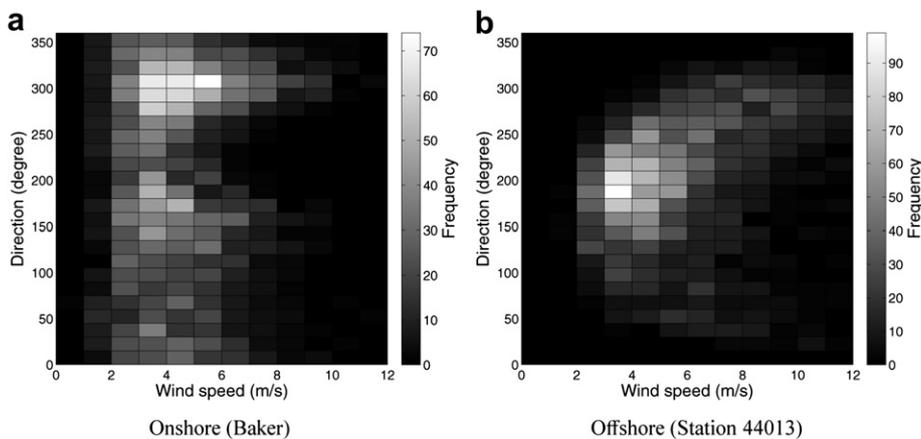


Fig. 1. Histogram of wind speed and wind direction.

2. Wind direction is the direction from which wind is blowing (degrees clockwise from north, $N = 0^\circ$; $NE = 45^\circ$; $E = 90^\circ$; $SE = 135^\circ$; $S = 180^\circ$; $SW = 225^\circ$; $W = 270^\circ$; $NW = 315^\circ$; etc.).
3. Air temperature is measured at 1.52 m above the soil surface at the onshore station (Baker), and at 4 m above the sea level at the offshore station (44013).

Generally, the Wind Power Density (WPD) is evaluated at 10 or 50 m height. In the case of the atmospheric boundary layer, a similarity study can be performed to describe the vertical profiles of turbulence statistics, when fully developed conditions are reached [29]. Assuming neutral conditions (negligible thermal effects), the mean speed in the surface layer (for heights less than 100 m) is commonly represented by the log profile [29]. For a known measured wind speed U_m at a height z_m , the log profile can be expressed as

$$\frac{U}{U_m} = \frac{\ln \frac{z}{z_0}}{\ln \frac{z_m}{z_0}} \quad (8)$$

where z_0 is the average roughness length (terrain dependent) in the farm region. In this paper, the wind speed data at the measured height is still used, which does not affect the distribution of wind conditions. However, wind speed data can be presented at any specific height for commercial wind turbines. In addition to the log profile mentioned in this paper, a variety of other methods are available or under development, such as the wind profile power law, the estimated wind speed profile using adaptive neuro-fuzzy inference system [30–32].

The density of dry air can be determined using the ideal gas law, expressed as a function of temperature and pressure [33],

$$\rho = \frac{p}{R \times T} \quad (9)$$

where ρ is the air density, p represents the absolute pressure; R is the specific gas constant for dry air, which is $287.058 \text{ J}/(\text{kg K})$, and T represents the local temperature in K . The absolute pressure above sea level is given by [33]

$$p = 101325 \left(1 - 2.25577 \times 10^{-5} \times h_a \right)^{5.25588} \quad (10)$$

where h_a is the altitude above sea level. The onshore site at the Baker station is at an altitude of $h_a = 512 \text{ m}$.

The vertical temperature and pressure profiles can also affect the wind power generation estimation. Temperature profiles can be formulated in a way analogous to the wind profile formulations [34]. In addition, a significant amount of research has been done to address temperature and pressure measurement in the literature [35,36]. The measurements and the vertical profiles development of these parameters are not within the scope of this paper.

4. Case study: application of the MMWD model

Using the onshore and the offshore wind data, the MMWD model is applied to three different cases:

1. Case I: Distribution of wind speed (univariate).
2. Case II: Distribution of wind speed and wind direction (bivariate).
3. Case III: Distribution of wind speed, wind direction, and air density (multivariate).

To validate the effectiveness of the MMWD model, the model is also investigated and compared with standard wind speed distributions [4,7]: (i) Weibull distribution, (ii) Gamma distribution, (iii) Normal distribution, (iv) Lognormal distribution, and (v) Rayleigh distribution.

In most of the studies presented in the specialized literature on wind energy and other renewable energy sources, the parameters of the distribution are estimated using classical methods: the *method of moments* (MM), the *maximum likelihood estimators* (MLE) and the *least squares method* (LSM) [37]. The MLE is adopted in this paper because it usually yields lower *mean squared errors* (MSE) associated with model parameter estimates than do MM estimators for the large samples that will be considered here [7].

4.1. MMWD case I: univariate distribution

In case I, the distribution of the wind speed is estimated. Successful modeling of univariate wind speed distribution has been reported in the literature [4,7]. The objective of this case study is to compare the MMWD model with other widely used fitting methods. In this paper, five existing wind distributions are selected, which are Weibull, Gamma, Lognormal, Normal and Rayleigh distributions. The bandwidth h values of the KDE are estimated to be 0.6649 (onshore Baker station [14]) and 0.3241 (offshore station 44013 [28]) using the optimal bandwidth selection method. Fig. 3 shows

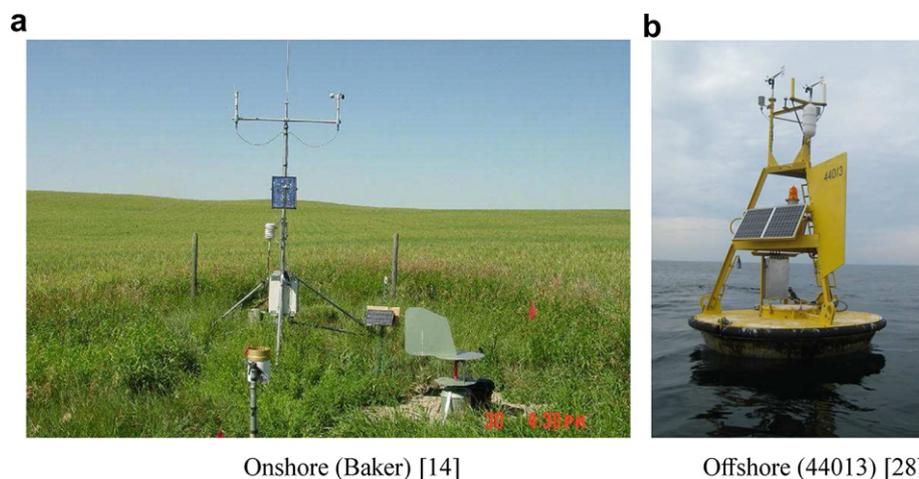


Fig. 2. Station setup.

Table 1
Details of the two stations.

Parameter	Onshore station [14]	Offshore station [28]
Location	Baker, ND	Station 44013, MA
Latitude	48.167° N	42.346° N
Longitude	99.648° W	70.651° W
Elevation	512 m	Sea level
Anemometer height	3 m above site elevation	5 m above site elevation
Air temp height	1.52 m above site elevation	4 m above site elevation
Water depth	None	61 m

the distributions estimated by the MMWD model and other distribution models. In Fig. 3(a) and (b), the solid lines represent the wind speed probability distributions estimated by the MMWD model. It is seen from Fig. 3(a) that, the probability distribution curves of MMWD, Lognormal and Gamma match the histogram the best. From Fig. 3(b), it is observed that the probability distribution curves of MMWD and Lognormal match the histogram the best. The cdf's of the probability distributions are shown in Fig. 4.

observed wind speed as the maximum possible wind speed [7]. Fig. 5(a) shows the Q–Q plot of the distributions at the Baker station. The red line is the Q–Q plot of the MMWD model. It is observed that the MMWD distribution follows the theoretical distribution (45° line $y = x$) more closely than other distributions. This observation indicates that the MMWD distribution performs better than other standard distributions in representing the univariate wind speed distribution. Similarly, the MMWD distribution also performs best at the offshore site (Station 44013), as seen from Fig. 5(b).

4.1.2. Coefficient of determination

The coefficient of determination is a measure of the agreement between an estimated distribution and the recorded data [41]. The coefficient of determination between the paired sample quantiles is evaluated. The coefficient of determination, R^2 , is the squared of correlation coefficient between the observed and modeled (predicted) data values, which is expressed as

$$R^2 = \left[\frac{\text{cov}(U, \hat{U})}{\sqrt{\text{var}(U)\text{var}(\hat{U})}} \right]^2 = \left\{ \frac{n \sum_{i=1}^n U_i \hat{U}_i - \sum_{i=1}^n U_i \sum_{i=1}^n \hat{U}_i}{\sqrt{\left[n \sum_{i=1}^n U_i^2 - \left(\sum_{i=1}^n U_i \right)^2 \right] \left[n \sum_{i=1}^n \hat{U}_i^2 - \left(\sum_{i=1}^n \hat{U}_i \right)^2 \right]}} \right\}^2 \quad (12)$$

The goodness-of-fit of the various fitted distributions to the wind speed data is evaluated using the coefficient of determination (R^2) associated with the resulting Quantile–Quantile (Q–Q) plots.

4.1.1. Quantile–Quantile (Q–Q) plot

In statistics, a Q–Q plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other [38]. The choice of quantiles from a theoretical distribution has significant discussion in the literature [39]. In this paper, the Weibull plotting position [40] is used in all cases, which is given by

$$p_i = i/(n + 1) \text{ where } i = 1, 2, \dots, n \quad (11)$$

The Weibull plotting position always gives an unbiased estimate of the observed cumulative probability regardless of the underlying distribution considered, and does not estimate the highest

where U and \hat{U} are the observed and fitted quantiles, respectively; cov and var mean covariance and variance, respectively. The closer the value of R^2 is to one, the more the fitted distribution agrees with the observed data.

Table 2 shows the comparison of the coefficient of determination at both onshore and offshore sites. It is observed that the MMWD model has the largest R^2 values at both stations. This observation illustrates the strong potential of this technique to provide accurate representations of wind distribution. This accurate MMWD wind distribution model would be also useful for wind resource assessment and wind farm layout optimization.

4.1.3. Wind Power Density (WPD) estimation

The WPD expressed in Eq. (1) is estimated using the Monte Carlo integration method based on the pdf of the wind speed. The sample points for integration are generated using the Sobol's quasirandom

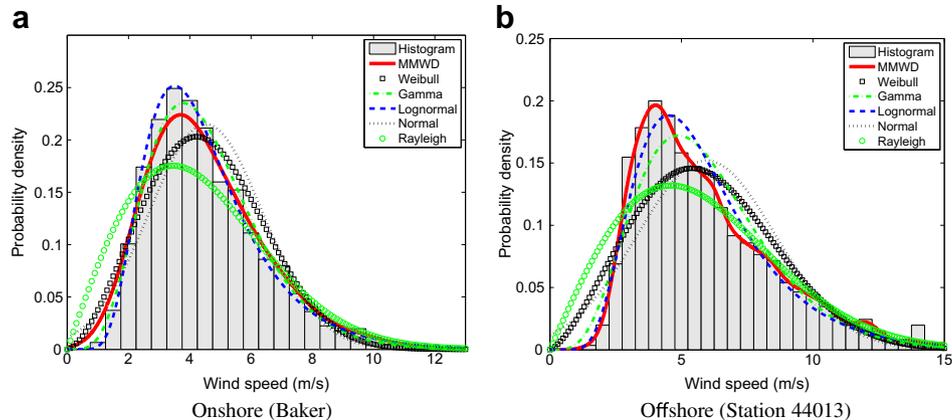


Fig. 3. Wind speed probability density functions.

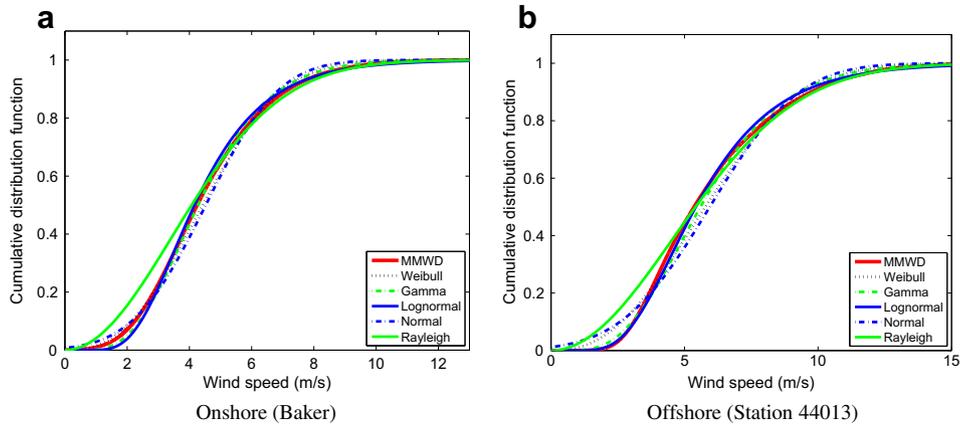


Fig. 4. Wind speed cumulative distribution functions.

sequence generator [42]. Sobol sequences use a base of two to form successively finer uniform partitions of the unit interval, and then reorder the coordinates in each dimension. The algorithm for generating Sobol sequences can be found in Bratley and Fox, Algorithm 659 [43]. The approximated WPD is then expressed as

$$\begin{aligned}
 \text{WPD} &= \int_0^{U_{max}} \frac{1}{2} \rho U^3 f(U) dU \\
 &\approx \sum_{i=1}^{N_p} \frac{1}{2} \rho U_i^3 f(U_i) \Delta U \\
 \text{where } \Delta U &= U_{max}/N_p
 \end{aligned} \tag{13}$$

where N_p is the sample size. In this case study, the density of the air is set to a reference value of 1.2 kg/m^3 at both stations. The WPD at the measured height of the Baker station is estimated to be 87.83 W/m^2 based on the ten-year wind data; and the WPD at the measured height of the offshore site (Station 44013) is estimated to be 211.52 W/m^2 based on the ten-year wind data.

4.1.4. Wind farm power generation estimation

This power generation model is adopted from Chowdhury et al. [15]. The power generated by a wind farm is an intricate function of the configuration and location of the individual wind turbines. The flow pattern inside a wind farm is complex, primarily due to the wake effects and the highly turbulent flow. The power generated by

a wind farm (P_{farm}) comprised of N wind turbines is evaluated as a sum of the powers generated by the individual turbines, which is expressed as [15]

$$P_{farm} = \sum_{j=1}^N P_j \tag{14}$$

The detailed formulation of the power generation model can be found in Ref. [15].

A rectangular wind farm of given dimensions, consisting of 9 turbines, is considered in this paper. The GE-1.5-MW-XLE [44] turbine is used in the case studies. The features of this turbine are provided in Table 3.

To further investigate the practical usefulness of the MMWD model, we compare the Weibull distribution and the MMWD model for estimating: (i) the WPD at the onshore Baker station; and (ii) the annual averaged power generation of the 9-turbine wind farm (based on the ten-year wind data). Table 4 shows the values of the WPD and power generation. We observe that the farm power generation estimated using the MMWD model is 5% lower than that estimated using the Weibull distribution; however, the WPD given by the MMWD model is higher than that given by the Weibull distribution. To explain this unexpected finding, we illustrate the WPD and power generation values contributed by different wind speeds, i.e. $\frac{1}{2} \rho U^3 f(U) \Delta U$ (Fig. 6(a)) and $P_{farm}(U) f(U) \Delta U$ (Fig. 6(b)), respectively. A higher WPD estimate is generally expected to translate into a higher estimate of wind farm power generation. The

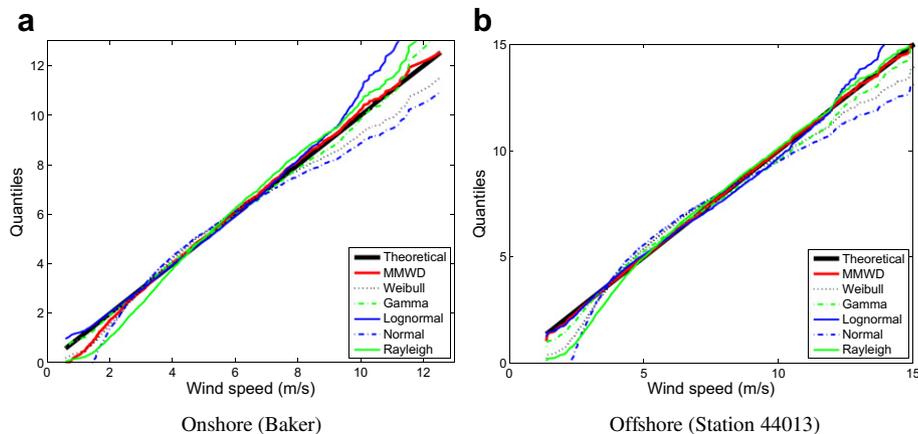


Fig. 5. Wind speed Quantile–Quantile plot of a sample of data versus a Weibull distribution.

Table 2
The coefficient of determination, R^2 .

Distribution model	Onshore (Baker)	Offshore (Station 44013)
MMWD	0.99804	0.99978
Weibull	0.98011	0.97110
Gamma	0.99796	0.99121
normal	0.98937	0.98687
Lognormal	0.95383	0.92921
Rayleigh	0.99511	0.98681

seemingly counterintuitive observation in the reported comparison can be attributed to the effect of the cut-out wind speed of the concerned turbine. The cut-out wind speed of the GE-1.5-MW-XLE turbine is 20.0 m/s at the hub height (80.0 m), which is equivalent to 10.2 m/s at the recorded data height of 3.0 m (height at which the wind distribution is estimated), according to the log profile. Hence, even if the MMWD estimates higher probabilities of wind speeds above 10.0 m/s (approx.) than given by Weibull distribution, it does not contribute to increased farm power generation, since no power is generated by the turbines when the wind speed is above the cut-out speed.

These observations also indicate that the Wind Power Density (WPD) might not be the best way to evaluate the resource potential of a wind farm site in practice. It is important to appreciate that the actual/effective resource potential is subject to (and not independent of) the power characteristics of currently available commercial turbines. More comprehensive measures of wind resource potential are therefore necessary to facilitate better planning of wind energy projects.

4.2. MMWD case II: bivariate distribution

The effectiveness of the MMWD model was validated in Case I. Case II investigates the joint distribution of the wind speed and wind direction. Figs. 7(a) and 8(a) represent the estimated wind velocity distribution for the onshore site (Baker station [14]). Figs. 7(b) and 8(b) represent the distribution for the offshore site (Station 44013 [28]). Interestingly, it is observed that the estimated probability distribution of the onshore wind data is multimodal in nature as expected; and the estimated offshore wind distribution can be treated practically as unimodal, which is actually common for offshore winds.

4.2.1. Wind rose

A wind rose is a graphical tool used by meteorologists to provide a succinct illustration of how the wind speed and the wind direction are distributed at a location. Sixteen cardinal directions are used in this illustration. In this illustration, North corresponds to 0° or 360° , East to 90° , South to 180° and West to 270° . The cardinal directions are ranked ($d_c = 1, 2, \dots, 16$) in the clockwise order ($d_c = 1$ for North, $d_c = 5$ for East, $d_c = 9$ for South, and $d_c = 13$ for West). Fig. 9 shows the estimated wind roses for the two site locations based on the ten-year wind data. For the onshore site, it is observed that winds from the northwest and the south dominate over the whole year. Minimal wind is observed from the northeast direction.

Table 3
Features of the GE-1.5-MW-XLE turbine [44].

Turbine feature	1.5-MW-XLE
Rated power (P_{r0})	1.5 MW
Rated wind speed (U_{r0})	11.5 m/s
Cut-in wind speed (U_{ino})	3.5 m/s
Cut-out wind speed (U_{out0})	20.0 m/s
Rotor diameter (D_0)	82.5 m
Hub height (H_0)	80.0 m

Table 4
WPD and farm power generation comparisons between Weibull and MMWD.

Performance	MMWD	Weibull
WPD (W/m^2)	87.83	86.15
Power generation (W)	7.41e6	7.78e6

For the offshore site, it is observed that winds from the west and the southwest dominate over the whole year. Minimal wind is observed from the north direction. This information plays an important role in the farm layout design and the selection of turbine types.

4.2.2. Wind Power Density (WPD) estimation

In this case study, the density of the air is set to a reference value of 1.2 kg/m^3 . The approximated WPD is expressed as

$$\begin{aligned} \text{WPD} &= \int_{0^\circ}^{360^\circ} \int_0^{U_{\max}} \frac{1}{2} \rho U^3 f(U, \theta) dU d\theta \\ &\approx \sum_{i=1}^{N_p} \frac{1}{2} \rho U_i^3 f(U_i, \theta_i) \Delta U \Delta \theta \end{aligned} \quad (15)$$

where $\Delta U \Delta \theta = U_{\max} \times 360^\circ / N_p$

From Eq. (15), the WPD is estimated to be 87.83 W/m^2 at the Baker station [14] and to be 211.53 W/m^2 at the Station 44013 [28]. The WPD for each cardinal direction, estimated using the *Monte Carlo integration* method, is shown in Fig. 10. It is observed that (i) the wind from northwest has relatively high WPD at the onshore site; and (ii) the wind from West has relatively high WPD at the offshore site.

4.3. MMWD case III: multivariate distribution

In case III, the multivariate probability distribution of the wind speed, the wind direction, and the air density is modeled. It is seen from Eq. (1) that the WPD is directly dependant on these three factors.

Fig. 11 shows the distributions of wind speed, wind direction and air density in a series of nested three-dimensional contours. The contours are at 25% (dark color), 50% (medium color) and 75% (light color), which are upper percentages of highest density regions. A strong correlation among the three wind condition parameters is evident from Fig. 11. This observation is the foundation of the hypothesis that - *an accurate representation of wind data for power prediction and farm design requires multivariate distribution models.*

4.3.1. Wind Power Density (WPD) estimation

The WPD is estimated using the *Monte Carlo integration* method, and is expressed as

$$\begin{aligned} \text{WPD} &= \int_{0^\circ}^{360^\circ} \int_0^{U_{\max}} \int_{\rho_{\min}}^{\rho_{\max}} \frac{1}{2} \rho U^3 f(U, \theta, \rho) d\rho dU d\theta \\ &\approx \sum_{i=1}^{N_p} \frac{1}{2} \rho_i U_i^3 f(U_i, \theta_i, \rho_i) \Delta U \Delta \theta \Delta \rho \end{aligned} \quad (16)$$

where $\Delta U \Delta \theta \Delta \rho = U_{\max} \times 360^\circ \times (\rho_{\max} - \rho_{\min}) / N_p$

For the onshore site (Baker station [14]), the WPD is estimated to be 84.60 W/m^2 using the ten-year wind data (2000–2009), which is represented by the horizontal line in Fig. 12(a). To show the variation of wind conditions, the WPD for each single year is evaluated, which is also shown in Fig. 12(a). It is observed that the WPD varies significantly over years. The variation is evaluated to be 39.24% using Eq. (17) below, which indicates significant uncertainty of wind conditions.

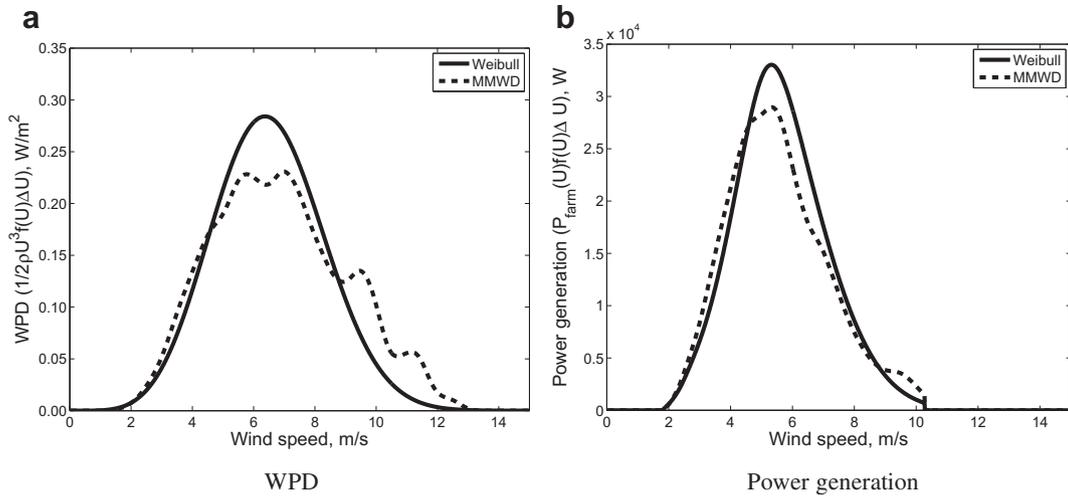


Fig. 6. The comparison of the Weibull and MMWD (Baker).

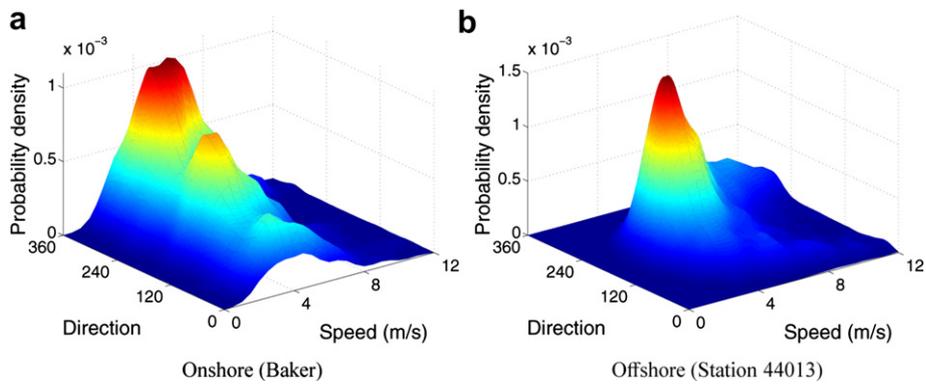


Fig. 7. Distribution of wind speed and wind direction.

For the offshore site (Station 44013 [28]), the WPD is estimated to be 225.88 W/m² using the ten-year wind data (2000–2009), which is represented by the horizontal line in Fig. 12(b). The WPD for each single year is also shown in Fig. 12(b). It is also observed that the WPD varies significantly over years. The variation is evaluated to be 53.46% using Eq. (17), which indicates significant uncertainty of offshore wind conditions.

KDE can often lead to overfitting of probability distributions. As a result, the prediction of long term (20 years) wind conditions using 10 years data might introduce greater uncertainties. Careful

modeling and characterization of these uncertainties, together with their propagation into the overall system, will allow for a more comprehensive quantification of the overall wind farm power output. Uncertainty characterization is a direction for future research.

$$\eta = \frac{WPD_{max} - WPD_{min}}{WPD_{avg}} \times 100\% \tag{17}$$

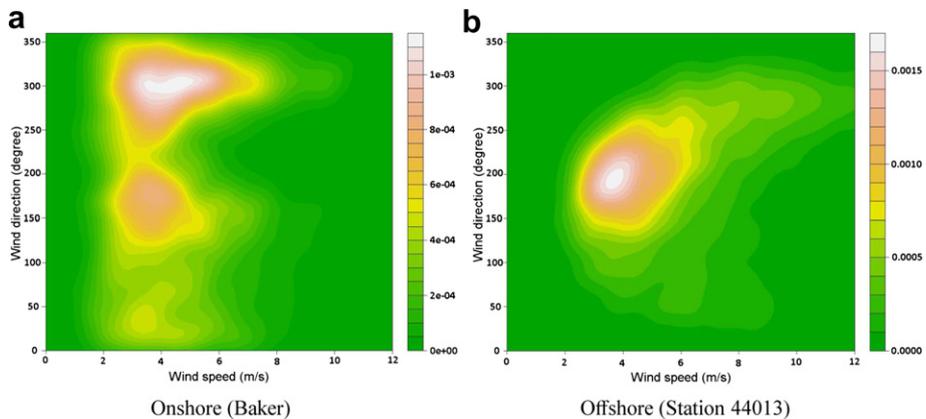


Fig. 8. Distribution of wind speed and wind direction (contour plot).

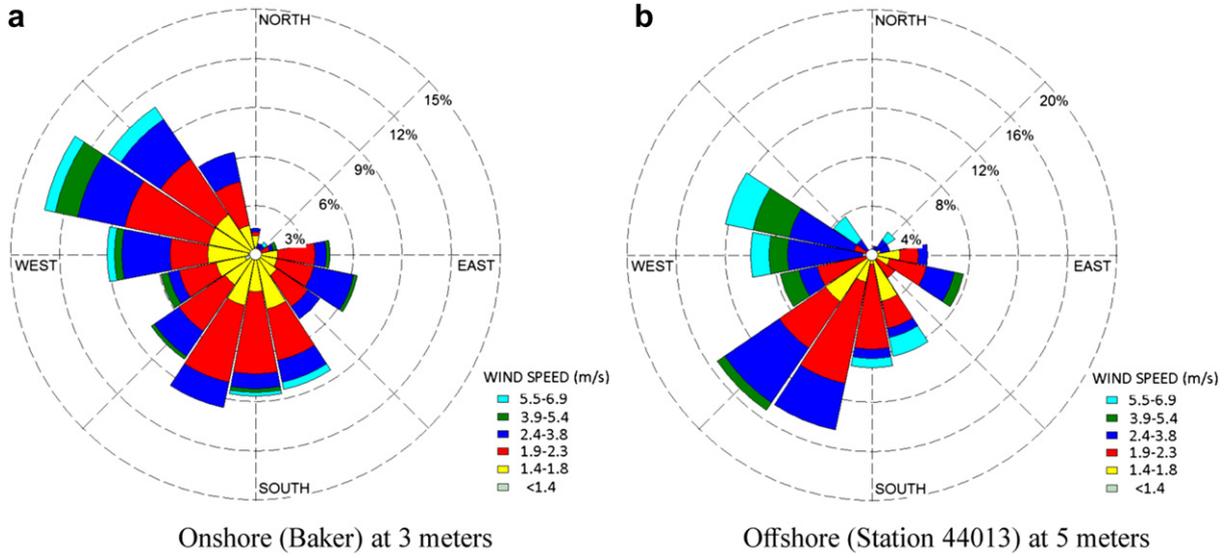


Fig. 9. Wind rose.

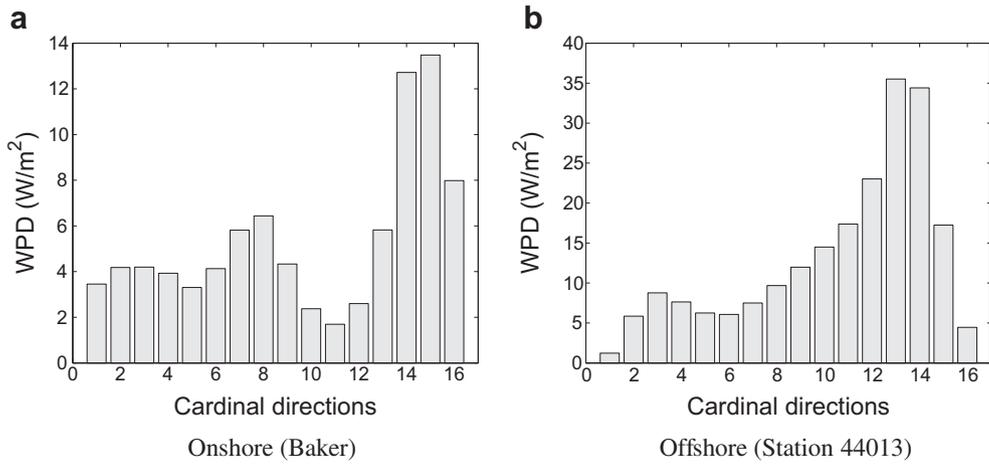


Fig. 10. WPD for cardinal directions.

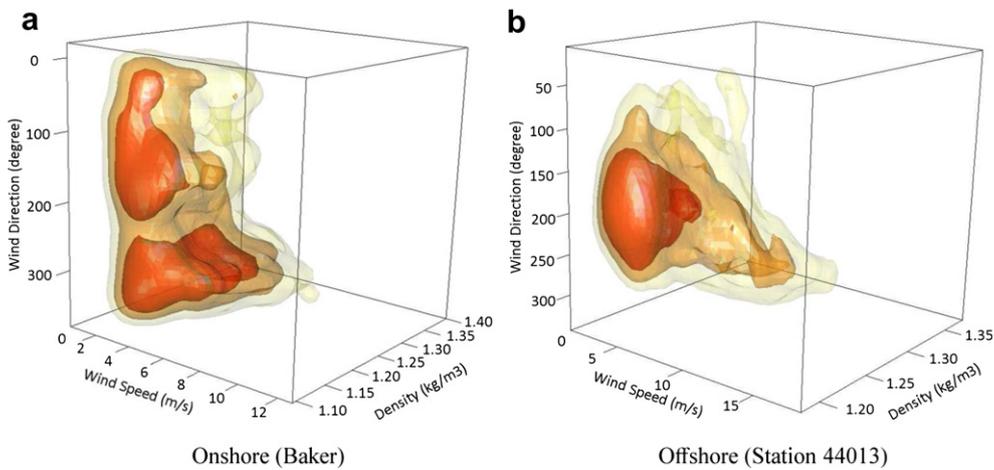


Fig. 11. Distribution of wind speed, wind direction, and air density.

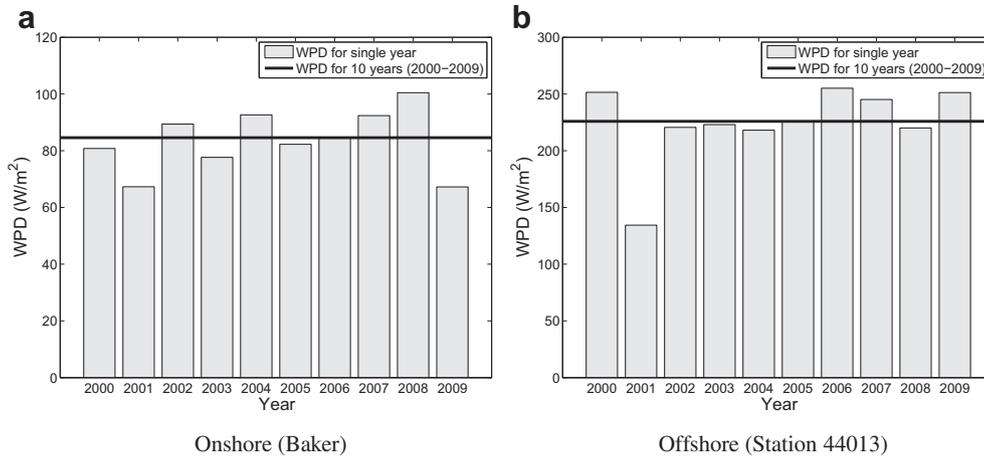


Fig. 12. WPD for each single year (from 2000 to 2009).

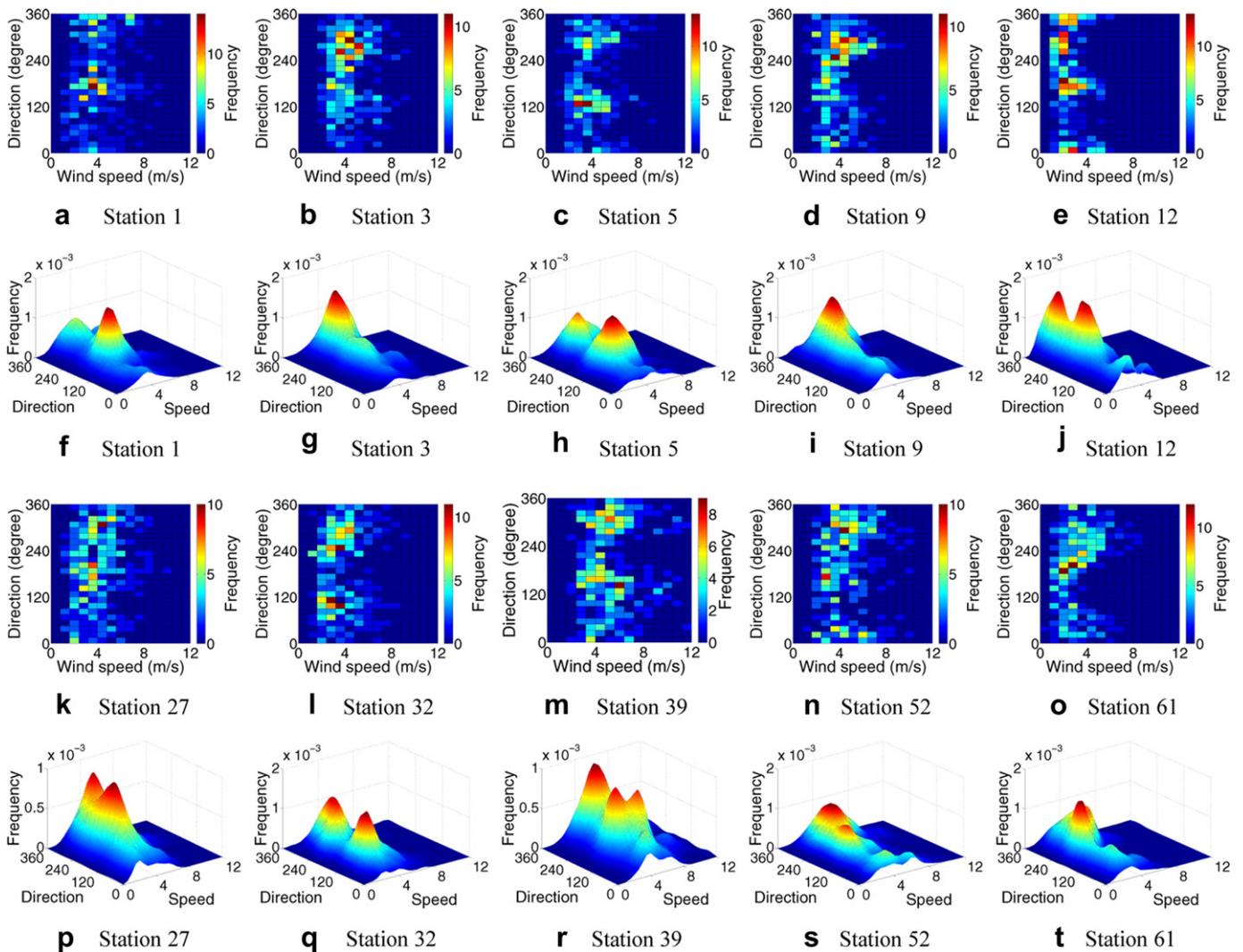


Fig. 13. Histograms and MMWD estimations of wind speed and wind direction (onshore).

where WPD_{avg} refers the average WPD over ten years; WPD_{max} and WPD_{min} represent the maximum and minimum WPD values, respectively.

4.4. Discussing the flexibility of the MMWD model

The MMWD model was applied to 100 wind sites (onshore and offshore) to evaluate the effectiveness, the flexibility and the wide applicability of the model. Two sites (one onshore and one offshore) were investigated in detail above. The other 98 sites consist of 71 onshore sites and 27 offshore sites in the USA. The onshore wind data is obtained from the North Dakota Agricultural Weather Network (NDAWN) [14]; and all the 72 stations available from the NDAWN website are selected. The offshore wind data is obtained from the National Data Buoy Center (NDBC) [28]. The 27 offshore sites consist of: (i) 7 sites in Northeast USA; (ii) 5 sites in Northwest USA; (iii) 5 sites in Southeast USA; (iv) 5 sites in Southwest USA; and (v) 5 sites in the Great Lakes region. The daily averaged wind speed and direction data from the year 2011 is used for the 98 wind sites. These entire 100 sites are expected to offer a wide variety of wind patterns.

For the sake of brevity, the results at 10 onshore and 10 offshore stations are provided to show the diversity of distribution behavior. Figs. 13 and 14 show differing wind patterns for the 10

onshore and 10 offshore stations, respectively. The two-dimensional figures represent the histograms of the recorded wind data; the three-dimensional figures represent the wind distributions estimated by the MMWD model. Each MMWD estimation plot corresponds to the histogram plot immediately above it. It is observed that the MMWD model accurately captures the joint distribution of wind speed and direction for the differing wind patterns.

It is observed from Fig. 13(a)–(t) that the wind distributions at some of the onshore stations are strongly multimodal — specifically, stations 1, 5, 12, 27, 32 and 39. On the contrary, the majority of the studied offshore stations (Fig. 14(a)–(t)) present practically unimodal distributions. Although such a small set of stations may not adequately represent the generic scenario, it can be said that offshore sites are more likely to have unimodal wind distributions. Further investigation of onshore and offshore atmospheric boundary layers and their variations can provide more insight into such distinct characteristics. The prevailing wind speeds vary approximately between 3 m/s and 5 m/s at the 10 onshore stations; while the prevailing wind speeds vary approximately between 4 m/s and 11 m/s at the 10 offshore stations. Especially for the offshore stations 9 and 10 (Fig. 14(n) and (o)), the prevailing wind speeds vary in a broad range. Interestingly, for the onshore sites with multimodal wind distributions, the modes are separated primarily

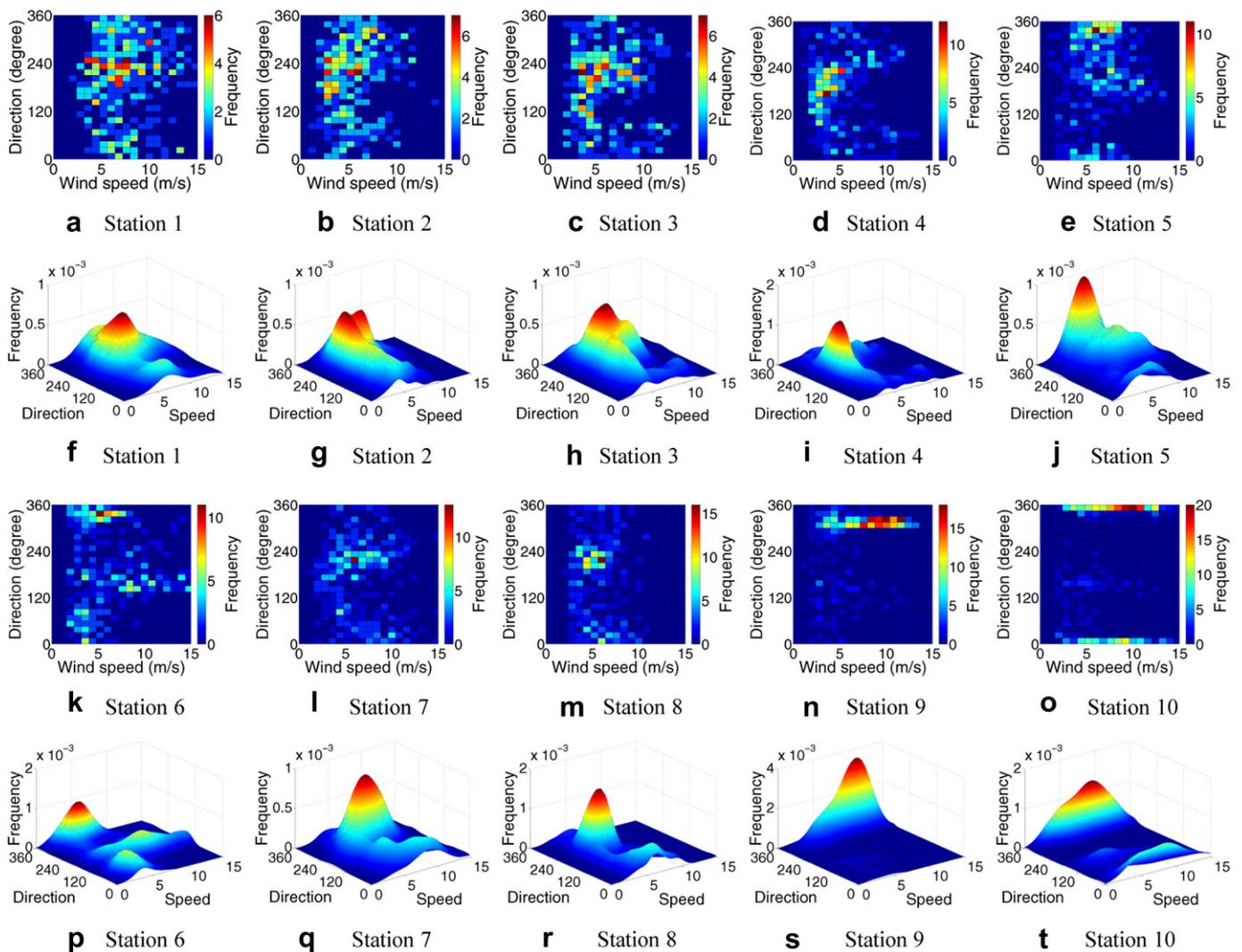


Fig. 14. Histograms and MMWD estimations of wind speed and wind direction (offshore).

by wind direction. This observation indicates the presence of two or more principal wind directions.

The north direction angle is defined as 0° for the data presented in this paper; and the wind direction angle increases clockwise. Ideally, the wind from the 360° should be identical with the wind from the 0° . It is seen from Fig. 14(o) and (t) that the principal wind direction at that site is near 0° (North direction); and the plot of the MMWD model yields two modes near the 0 and the 360° . In this case, the induced artificial separation of 0° and 360° directions has not introduced significant discrepancies in the wind probability values. However, further research is necessary to allow the MMWD model to better represent the periodic nature of wind direction data.

5. Concluding remarks

This paper developed a *Multivariate and Multimodal Wind Distribution* (MMWD) model to represent the distribution of wind conditions (speed, direction and air density) using recorded data. Univariate, bivariate and trivariate wind distributions were explored using the MMWD model.

The performance of the MMWD model was measured using the *coefficient of determination* (R^2) associated with the resulting *Quantile–Quantile* (Q–Q) plots. The effectiveness and the reliability of the MMWD model were successfully validated using the ten-year wind data from the *North Dakota Agricultural Weather Network* (NDAWN) and the *National Data Buoy Center* (NDBC). A strong correlation was observed among wind speed, wind direction, and air density. In addition, the MMWD model was applied to 100 wind sites; and the results illustrated a multimodal wind distribution. These observations corroborate the need to develop wind distribution models that can capture both the multivariate and multimodal nature of recorded data. The results also showed differing wind distribution patterns between the onshore and offshore sites. Such a nonparametric stochastic modeling approach can be used to represent the variation in other intermittent natural energy resources as well.

Specifically, the MMWD model could be helpful for evaluating the wind resource potential for farm siting. The implementation of the MMWD model in optimal planning of commercial scale wind farms will further establish the true potential of this methodology. Explicit consideration of the uncertainties in wind conditions, in addition to the mean long term variations (developed in this paper) will provide further insights into the exploration of wind resource potential.

Acknowledgments

Support from the National Science Foundation Awards CMMI-1100948 and CMMI-0946765 is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

References

- [1] BTM, International wind energy development world market update 2010, Tech. rep., BTM Consult (March 2011).
- [2] Burton T, Sharpe D, Jenkins N, Bossanyi E. Wind energy handbook. Wiley; 2001.
- [3] Manwell J, McGowan J, Rogers A. Wind energy explained: theory, design and application. 2nd ed. Wiley; 2010.
- [4] Carta J, Ramírez P, Velázquez S. A review of wind speed probability distributions used in wind energy analysis case studies in the canary islands. Renewable and Sustainable Energy Reviews 2009;13(5):933–55.
- [5] Seguro J, Lambert T. Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis. Journal of Wind Engineering and Industrial Aerodynamics 2000;85(1):75–84.
- [6] Dorvlo A. Estimating wind speed distribution. Energy Conversion and Management 2002;43(17):2311–8.
- [7] Morgan E, Lackner M, Vogel R, Baise L. Probability distributions for offshore wind speeds. Energy Conversion and Management 2011;52(1):15–26.
- [8] Sorensen P, Nielsen T. Recalibrating wind turbine wake model parameters - validating the wake model performance for large offshore wind farms. In: European wind energy conference and exhibition; 2006. Athens, Greece.
- [9] Katic I, Hojstrup J, Jensen N. A simple model for cluster efficiency. In: Proceedings of European wind energy conference and exhibition. Rome: Italy; 1986.
- [10] Lackner MA, Elkinton CN. An analytical framework for offshore wind farm layout optimization. Wind Engineering 2007;31(1):17–31.
- [11] Vega R. Wind directionality: a reliability-based approach, Ph.D. thesis, Texas Tech University: Lubbock, TX; August 2008.
- [12] Carta J, Ramírez P, Bueno C. A joint probability density function of wind speed and direction for wind energy analysis. Energy Conversion and Management 2008;49(6):1309–20.
- [13] Erdem E, Shi J. Comparison of bivariate distribution construction approaches for analysing wind speed and direction data. Wind Energy 2011;14(1):27–41.
- [14] NDSU. The North Dakota agricultural weather network, <http://ndawn.ndsu.nodak.edu/>; 2011.
- [15] Chowdhury S, Zhang J, Messac A, Castillo L. Unrestricted wind farm layout optimization (uwfloo): investigating key factors influencing the maximum power generation. Renewable Energy 2012;38(1):16–30.
- [16] Zhang J, Chowdhury S, Messac A, Castillo L. Multivariate and multimodal wind distribution model based on kernel density estimation, In: ASME 2011 5th International Conference on Energy Sustainability, Washington, DC; 2011.
- [17] Chowdhury S, Zhang J, Messac A, Castillo L. Developing a flexible platform for the optimal design of commercial-scale wind farms, In: ASME 2011 5th International Conference on Energy Sustainability, Washington, DC; 2011.
- [18] Simonoff J. Smoothing methods in statistics. 2nd ed. Springer; 1996.
- [19] Nfaoui H, Bahraoui J, Darwish A, Sayigh A. Wind energy potential in morocco. Renewable Energy 1991;1(1):1–8.
- [20] Rosenblatt M. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 1956;27(3):832–7.
- [21] Parzen E. On estimation of a probability density function and mode. The Annals of Mathematical Statistics 1962;33(3):1065–76.
- [22] Jones M, Marron J, Sheather S. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 1996; 91(433):401–7.
- [23] Epanechnikov V. Non-parametric estimation of a multivariate probability density. Theory of Probability and Its Applications 1969;14:153–8.
- [24] Duong T, Hazelton M. Plug-in bandwidth matrices for bivariate kernel density estimation. Nonparametric Statistics 2003;15(1):17–30.
- [25] Wand MJMP. Kernel smoothing. CRC Press; 1995.
- [26] Wand W, Jones M. Multivariate plug-in bandwidth selection. Computational Statistics 1994;9:97–116.
- [27] Turlach BA. Bandwidth selection in kernel density estimation: a review, CORE and Institut de Statistique.
- [28] NOAA. Noaa's national data buoy center, <http://www.ndbc.noaa.gov/>; 2011.
- [29] Crasto G. Numerical simulations of the atmospheric boundary layer, tech. rep. Cagliari, Italy: Università degli Studi di Cagliari; February 2007.
- [30] Mohandes M, Rehman S, Rahman S. Estimation of wind speed profile using adaptive neuro-fuzzy inference system (anfis). Applied Energy 2011;88: 4024–32.
- [31] He H, Cao Y, Cao Y, Wen J. Ensemble learning for wind profile prediction with missing values. Neural Computing & Applications 2011. <http://dx.doi.org/10.1007/s00521-011-0708-1>.
- [32] Pena A, Gryning S, Hasager C. Measurements and modelling of the wind speed profile in the marine atmospheric boundary layer. Boundary-layer Meteorology 2008;479–95.
- [33] Kutz M. Mechanical engineers' handbook. 3rd ed. Wiley; 2006.
- [34] Panofsky HA, Dutton JA. Atmospheric turbulence: models and methods for engineering applications. Wiley; 1984.
- [35] Tong W, editor. Wind power generation and wind turbine design. WIT Press; 2010.
- [36] Schmitt C, Piazzoni A, Serralunga P, Ruille B, Goretzki B. Effects of varying atmospheric stratification on vertical wind profile and energy yield prediction in complex terrain sites. In: EWEA 2011 annual event. Brussels: Belgium; 2011.
- [37] Canavos G. Applied probability and statistical methods, little. Brown and Company; 1984.
- [38] Gibbons J, Chakraborti S. Nonparametric statistical inference. CRC Press; 2003.
- [39] Makkonen L. Bringing closure to the plotting position controversy. Communications in Statistics – Theory and Methods 2008;37(3):460–7.
- [40] Looney S, Gullede T. Use of the correlation coefficient with normal probability plots. The American Statistician 1985;39(1):75–9.
- [41] Forrester A, Sobester A, Keane A. Engineering design via surrogate modelling: a practical guide. J. Wiley; 2008.
- [42] Sobol I. Uniformly distributed sequences with an additional uniform property. USSR Computational Mathematics and Mathematical Physics 1976;16(5): 236–42.
- [43] Bratley P, Fox B. Algorithm 659: implementing sobol's quasirandom sequence generator. ACM Transactions on Mathematical Software 1988;14(1):88–100.
- [44] GE GE. Energy 1.5 MW wind turbine brochure, general electric, <http://www.gepower.com/>; 2010.