

ESFuelCell2011-54507

MULTIVARIATE AND MULTIMODAL WIND DISTRIBUTION MODEL BASED ON KERNEL DENSITY ESTIMATION

Jie Zhang*

Rensselaer Polytechnic Institute
Troy, New York, 12180
Email: zhangj17@rpi.edu

Souma Chowdhury*

Rensselaer Polytechnic Institute
Troy, New York, 12180
Email: chowds@rpi.edu

Achille Messac[†]

Syracuse University
Syracuse, NY, 13244
Email: messac@syr.edu

Luciano Castillo[‡]

Rensselaer Polytechnic Institute
Troy, New York, 12180
Email: castil2@rpi.edu

ABSTRACT

This paper presents a new method to accurately characterize and predict the annual variation of wind conditions. Estimation of the distribution of wind conditions is necessary (i) to quantify the available energy (power density) at a site, and (ii) to design optimal wind farm configurations. We develop a smooth multivariate wind distribution model that captures the coupled variation of wind speed, wind direction, and air density. The wind distribution model developed in this paper also 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 distribution of wind speed and wind direction (bivariate); and (iii) the 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. Ten-year recorded wind data, obtained from the North Dakota Agricultural Weather Network (NDAWN), is used in this paper. We found the coupled distribution to be multimodal. A strong correlation among the wind condition parameters was also observed.

Keywords: Energy, kernel density estimation, multimodal, multivariate, wind distribution, wind power density

INTRODUCTION

Over the last decade, the global installed wind capacity has been growing at an approximate rate of 28% per year [1]. The available energy from a wind resource varies appreciably over one year. The uncertainty in wind resource potential is partially responsible in restraining wind energy from playing a major role in the overall energy market. A determination and the forecasting of the variability of the available energy would serve two important objectives: (i) Analyzing the quality of a wind farm site,

*Doctoral Student, Multidisciplinary Design and Optimization Laboratory, Department of Mechanical, Aerospace and Nuclear Engineering, ASME student member.

[†]Distinguished Professor and Department Chair. Department of Mechanical and Aerospace Engineering, ASME Fellow. Corresponding author.

[‡]Associate Professor, Department of Mechanical Aerospace and Nuclear Engineering.

and (ii) designing an optimum wind farm layout and selecting appropriate 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, which is expressed as

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

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

By far, 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 maximum entropy probability density function [4, 7].

Research Objectives and Motivation

Wind energy sources generally appear in the form of wind farms that consist of multiple wind turbines located in a particular arrangement over 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 simple 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]. Hence, an optimal layout of turbines that ensures maximum farm efficiency is of utmost importance in conceiving a wind farm project.

For a given farm layout, the direction of wind has a strong influence on the wakes created & subsequently on the overall flow pattern in the wind farm. Thus, we believe that 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.

As discussed above, 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 correlation and the modality of the distribution of wind* - such assumptions can lead to approximations that deviate significantly from the actual scenario. In addition, it can be seen from Eqn. (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, we estimated the annual variation in air density 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 rare in the literature.

In this paper, we develop and explore a new method to represent the multivariate (and likely multimodal) distribution of wind conditions. *Multivariate kernel density estimation* [15] 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 II. The ten-year wind data used for the distribution is provided in Section III. Section IV presents the results and discussion for the three scenarios studied. Concluding remarks and future work are given in the last Section.

MULTIVARIATE AND MULTIMODAL WIND DISTRIBUTION (MMWD) MODEL

Kernel Density Estimation (KDE)

KDE, also known as the Parzen-Rosenblatt window method [16, 17], 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 [18]

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

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).

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 [19]. In this paper, we consider $K(x) = (2\pi)^{-d/2} \exp(-\frac{1}{2}x^T x)$, the standard normal throughout. In contrast, the choice of H is crucial in determining the performance of \hat{f} [20].

Optimal Bandwidth Matrix Selection

The most commonly used optimality criterion for selecting a bandwidth matrix is the *Mean Integrated Squared Error* (MISE) [20], which is expressed as

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

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

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

where vec is the vector half operator, given by

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

The general expression of ψ_4 could be found in Wand and Jones [21]. An ideal optimal bandwidth is estimated to be

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

In this paper, we use the multivariate plug-in selector ($PI(H)$) developed by Wand and Jones [22], which is given by

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

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

WIND CONDITION DATA

The wind data used in this paper is obtained from the *North Dakota Agricultural Weather Network* (NDAWN) [14]. We use the daily averaged data for wind speed, wind direction, and air temperature measured at the Baker station (Fig. 1) between the year 2000 and 2009. Table 1 shows the geographical coordinates and the elevation of the station. The measurement information is listed as follows.

1. Wind speed is measured at 3 meters above the soil surface with an anemometer. The value is the average of all hourly average wind speeds for a 24-hour period from midnight to midnight.
2. Wind direction is the direction from which wind is blowing (degrees clockwise from north) measured at 3 meters above the soil surface ($N = 0^\circ; NE = 45^\circ; E = 90^\circ; SE = 135^\circ; S = 180^\circ; SW = 225^\circ; W = 270^\circ; NW = 315^\circ; \text{etc.}$) with a wind vane. The value is the average of all measured wind directions for a 24-hour period from midnight to midnight.
3. Air temperature is measured at 1.52 meters above the soil surface with a temperature sensor. The value is the average of the maximum and minimum daily air temperatures.

Assuming neutral conditions (negligible thermal effects), the mean velocity in the surface layer (for heights less than 100m) is commonly represented by the log profile [23]. For a known recorded 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 where z_0 is the average roughness length (terrain dependent) in the farm region, and U is the wind speed at a height z . The log-profile given by Eq. 8 was used to determine the wind speed at the hub height from the wind speed data recorded at 3m height.

The density of dry air can be determined using the ideal gas law, expressed as a function of temperature and pressure [24],

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

where ρ is the air density, p represents the absolute pressure; R is the specific gas constant for dry air, which is $287.058 \text{ J}/(\text{kg} \cdot \text{K})$, and T represents absolute temperature. The absolute pressure above sea level is given by [24]

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

where h is the altitude above sea level. The concerned site in this paper is at an altitude of $h = 512 \text{ m}$.

TABLE 1. DETAILS OF NDAWN STATION: BAKER, ND [14]

Parameter	Value
Location	Baker, ND
Latitude	48.167°
Longitude	-99.648°
Elevation	512m



FIGURE 1. BAKER STATION SETUP [14]

CASE STUDY: APPLICATION OF THE MMWD MODEL

The MMWD technique has been applied to three different cases (I, II and III) in order to investigate the extent of wind distribution, as listed below.

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).

In order to validate the effectiveness of the MMWD model, we also investigate and compare the model with standard wind speed distributions [4, 7]: (i) Weibull distribution, (ii) Gamma distribution, (iii) normal distribution, (iv) Lognormal distribution, and (v) Rayleigh distribution. The *probability density function (pdf)* and *cumulative distribution function (cdf)* of each distribution model are listed in Appendix A.

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)* [25]. The maximum likelihood method has the property of maximizing the value of the probability of the randomly observed sample. MLE is adopted in this paper because it usually yields low *mean squared*

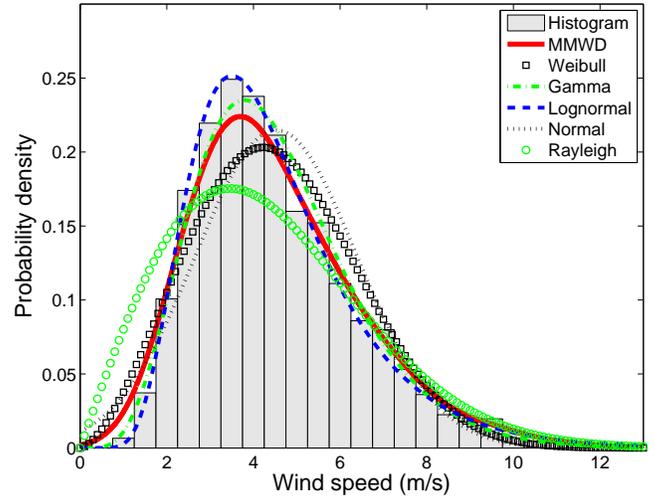


FIGURE 2. WIND SPEED PROBABILITY DENSITY FUNCTIONS

errors (MSE) associated with model parameter estimates than MM estimators for large samples considered here [7]. The expressions of the estimated parameters are listed in Appendix A.

MMWD Case I: Univariate Distribution

In case I, we are estimating the distribution of the wind speed. 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 of the KDE is estimated to be 0.6649 using the optimal bandwidth selection method. Figure 2 shows the distribution estimated by the MMWD model and other distribution models. In Fig. 2, the red line represents the wind speed probability distribution estimated by the MMWD model. The *cdfs* of the probability distributions are shown in Fig. 3.

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.

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 [26]. The choice of quantiles from a theoretical distribution has occasioned much discussion. In this paper, we use

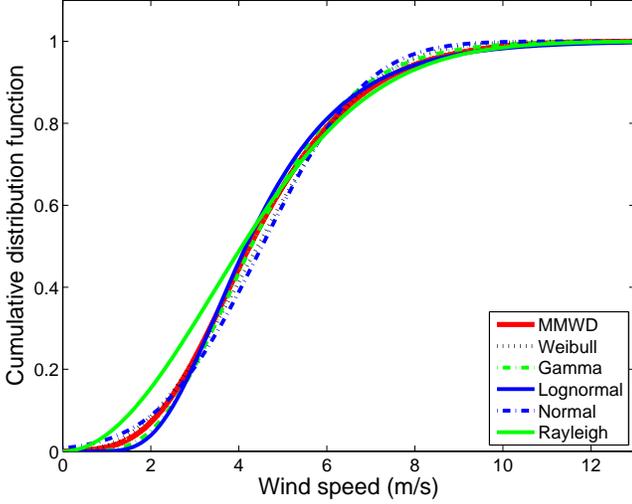


FIGURE 3. WIND SPEED CUMULATIVE DISTRIBUTION FUNCTIONS

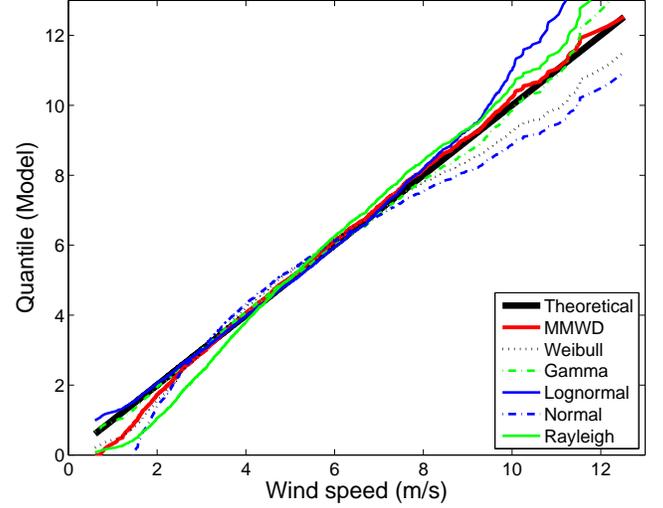


FIGURE 4. WIND SPEED QUANTILE-QUANTILE PLOT OF A SAMPLE OF DATA VERSUS A WEIBULL DISTRIBUTION

the Weibull plotting position [27] in all cases, which is given by

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

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 observed wind speed as the maximum possible wind speed [7]. Figure 4 shows the Q-Q plot of the distributions. The red line is the Q-Q plot of MMWD model. We observe that the MMWD distribution follows the theoretical distribution (45° line $y = x$) closer than other distributions. This observation indicates that the MMWD distribution performs better than other standard distributions for representing the univariate wind speed distribution.

Coefficient of Determination The *coefficient of determination* is a quantity to measure the agreement of a fitting distribution with the recorded data [28]. In the current paper, we evaluate the *coefficient of determination* between the paired sample quantiles. The *coefficient of determination*, R^2 , is the squared of correlation coefficient between the observed and modeled (pre-

dicted) data values, which is expressed as

$$R^2 = \left[\frac{\text{cov}(U, \hat{U})}{\sqrt{\text{var}(U)\text{var}(\hat{U})}} \right]^2 \quad (12)$$

$$= \left\{ \frac{n \sum_{i=1}^n U_i \hat{U}_i - \sum_{i=1}^n U_i \sum_{i=1}^n \hat{U}_i}{\sqrt{\left[n \sum_{i=1}^n U_i^2 - \left(\sum_{i=1}^n U_i \right)^2 \right] \left[n \sum_{i=1}^n \hat{U}_i^2 - \left(\sum_{i=1}^n \hat{U}_i \right)^2 \right]}} \right\}^2$$

where U and \hat{U} are the observed and fitted quantiles; *cov* and *var* means covariance and variance, respectively. The closer the value of R^2 is to one, the more the fitted distribution agrees with observed data, which indicates a better fit of the model.

Table 2 and Fig. 5 show the comparison of the *coefficient of determination*. It can be seen that, the MMWD model has the largest R^2 value. This observation illustrates the strong potential of this technique to provide accurate representation of wind distribution.

Wind Power Density (WPD) Estimation The WPD expressed in Eqn. (1) is estimated using *Monte Carlo integration* method based on the *pdf* of wind speed. *Monte Carlo integration* is an algorithm for the approximate evaluation of definite integral using random numbers. The sample points for integration in this paper are generated using the Sobol's quasirandom sequence generator [29]. 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 algo-

TABLE 2. THE COEFFICIENT OF DETERMINATION, R^2

Distribution model	MMWD	Weibull	Gamma	normal	Lognormal	Rayleigh
R^2	0.99804	0.98011	0.99796	0.98937	0.95383	0.99511

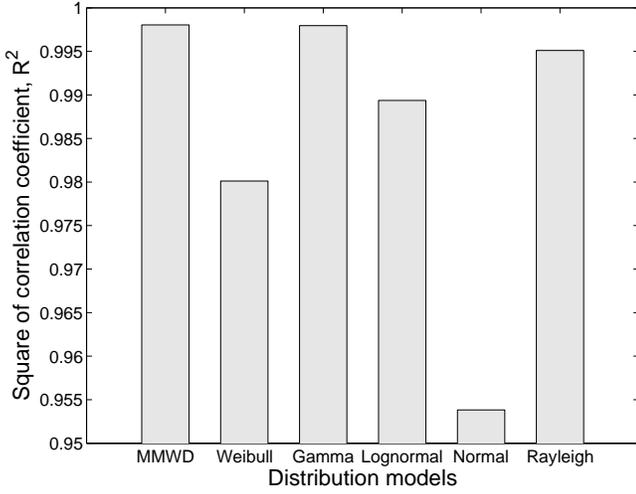


FIGURE 5. THE COEFFICIENT OF DETERMINATION, R^2

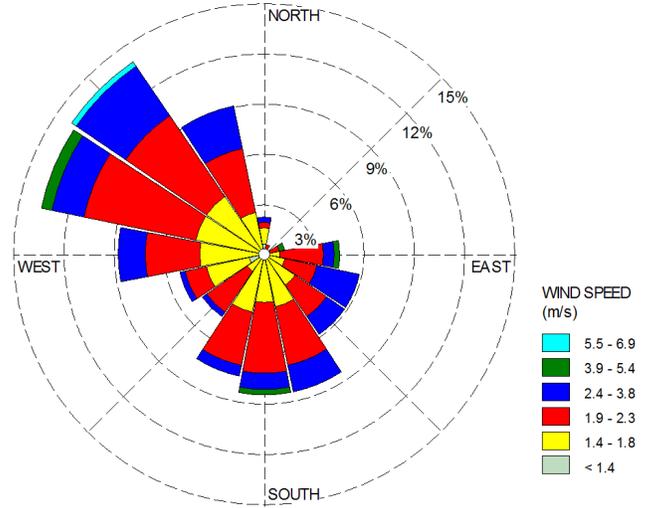


FIGURE 8. WIND ROSE AT THREE METERS HEIGHT

gorithm for generating Sobol sequences can be found in Bratley and Fox, Algorithm 659 [30]. The approximated WPD is then expressed as

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

where $\Delta U = U_{max}/N_p$

where N_p is the number of sample size. In this case study, the density of the air is set to a reference value of 1.2 kg/m^3 at the station. The WPD is then estimated to be 87.83 W/m^2 based on the ten-year wind data.

MMWD Case II: Bivariate Distribution

The effectiveness of the MMWD model was validated in case I. Case II investigates the coupled distribution of the wind speed and wind direction. Figures 6(a) and 7(a) represent the estimated wind velocity distribution for the ten-year wind data (2000-2009); Figures 6(b) and 7(b) represent the distribution for the recorded data in the year 2009. Interestingly we observe that the estimated probability distributions are multimodal in nature.

Wind Rose A wind rose is a graphical tool used by meteorologists to provide a succinct illustration of how wind speed and wind direction are typically distributed at a particular 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). Figure 8 shows the estimated wind rose for this location based on the ten-year wind data. We observe that winds from the North-west and the South dominate over the whole year. Minimal wind is observed from the Northeast direction.

Wind Power Density (WPD) Estimation In this case study, the density of the air is also set to a reference value of 1.2 kg/m^3 . Then, the approximated WPD is expressed as

$$\begin{aligned}
 WPD &= \int_{0^\circ}^{360^\circ} \int_0^{U_{max}} \frac{1}{2} \rho U^3 f(U, \theta) dU d\theta \\
 &\simeq \sum_{i=1}^{N_p} \frac{1}{2} \rho U_i^3 f(U_i, \theta_i) \Delta U \Delta \theta
 \end{aligned} \quad (14)$$

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

The WPD is estimated to be 87.83 W/m^2 . The WPD for each cardinal direction, estimated using the *Monte Carlo integration*

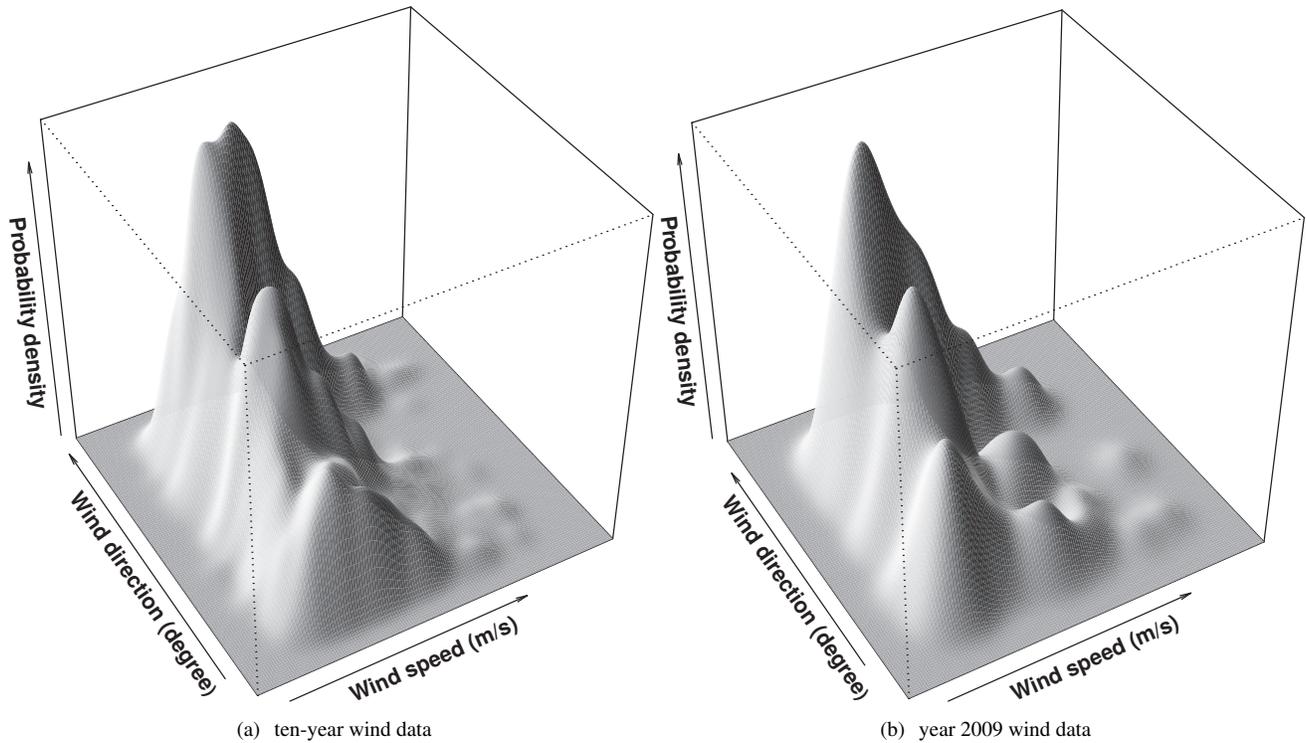


FIGURE 6. DISTRIBUTION OF WIND SPEED AND WIND DIRECTION

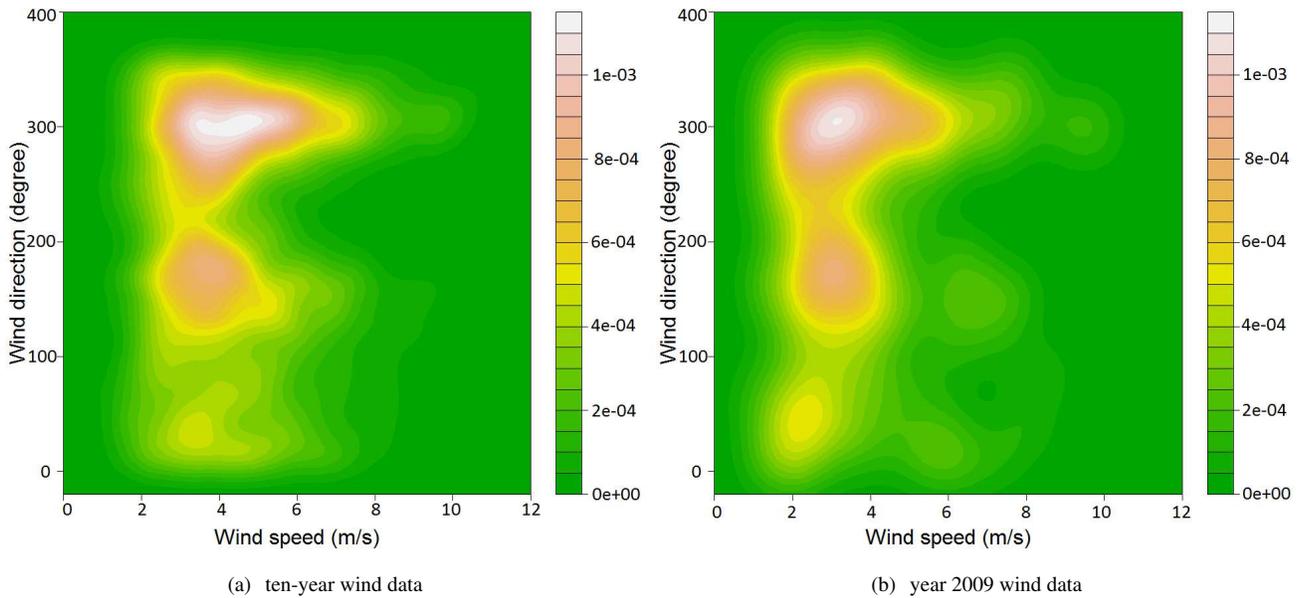


FIGURE 7. DISTRIBUTION OF WIND SPEED AND WIND DIRECTION (CONTOUR PLOT)

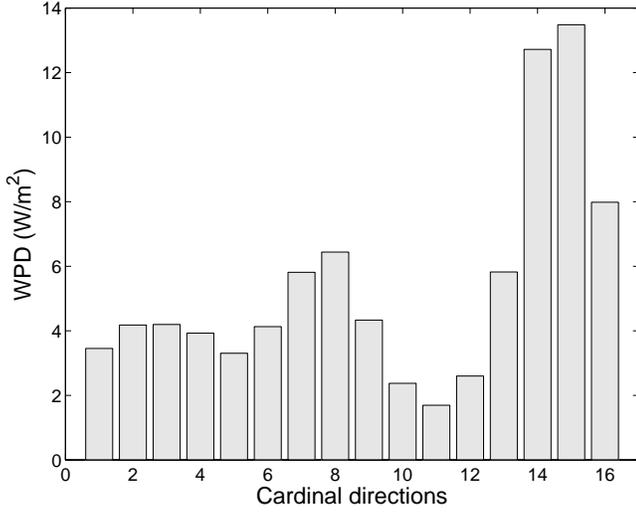
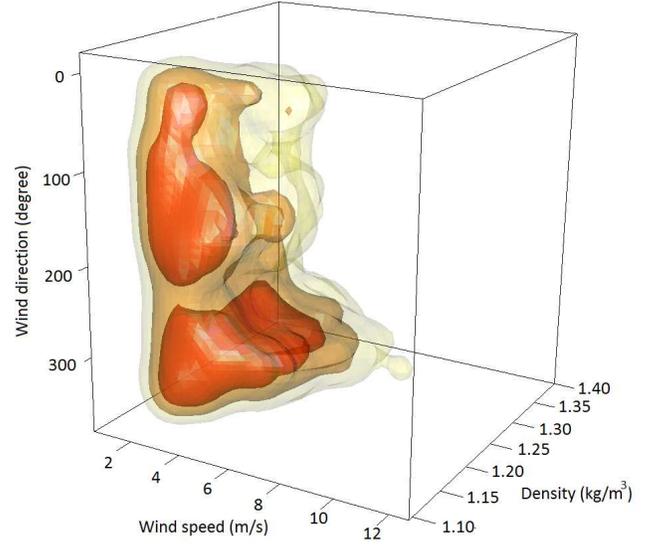
method, is shown in Table 3 and Fig. 9. We observe that the wind from northwest has relatively high WPD in this region (as also seen from the wind rose).

MMWD Case III: Multivariate Distribution

In case III, we model the multivariate probability distribution of wind speed, wind direction, and air density. It can be seen from Eqn. (1) that the WPD is strongly related to all the three

TABLE 3. WPD FOR CARDINAL DIRECTIONS

Direction	1	2	3	4	5	6	7	8
$WPD(W/m^2)$	3.45	4.18	4.20	3.93	3.30	4.13	5.81	6.44
Direction	9	10	11	12	13	14	15	16
$WPD(W/m^2)$	4.33	2.37	1.69	2.60	5.82	12.72	13.48	7.99

**FIGURE 9. WPD FOR CARDINAL DIRECTIONS****FIGURE 10. DISTRIBUTION OF WIND SPEED, WIND DIRECTION, AND AIR DENSITY**

factors.

Figure 10 shows the distribution 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. 10. This observation is the foundation of our hypothesis - *an accurate representation of wind data (for power prediction & farm design) requires multivariate distribution models.*

Wind Power Density (WPD) Estimation The approximated WPD estimated using *Monte Carlo integration* method can be expressed as

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

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

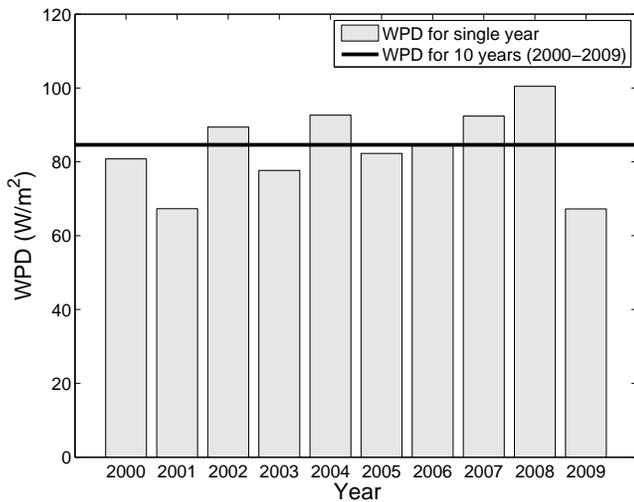
The WPD is estimated to be $84.60W/m^2$ using the ten-year

wind data (2000-2009), which is represented by the straight line in Fig. 11. In order to show the variation of wind conditions, the WPD for each single year is evaluated, which is shown in Table 4 and Fig. 11. We can see that the WPD varies significantly over years. The variation is calculated to be 39.24% using Eqn. (16), which indicates significant uncertainty of wind conditions. It is important to recognize that uncertainty anywhere in the system will potentially lead to uncertainty everywhere - thereby eventually impacting the overall power output of the wind farm. Careful modeling and characterization of these uncertainties, together with their propagation into the overall system, will allow for the credible quantification of the overall wind farm power output. Such uncertainty characterization should be an important direction for future research.

$$\% \text{ variation} = \frac{WPD_{max} - WPD_{min}}{WPD_{avg}} \quad (16)$$

TABLE 4. WPD FOR EACH SINGLE YEAR

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
WPD(W/m^2)	80.78	67.31	89.39	77.66	92.64	82.27	84.44	92.37	100.45	67.25

**FIGURE 11. WPD FOR EACH SINGLE YEAR (FROM 2000 TO 2009)**

CONCLUDING REMARKS AND FUTURE WORK

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. We explored univariate, bivariate and trivariate wind distributions 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). A strong correlation was observed among wind speed, wind direction, and air density. In addition, the case study 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. Such an unassuming stochastic modeling approach can provide a strong foundation to the development of methods to represent the variation in other 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 planing of commercial scale wind farms should further establish the true potential of this method.

ACKNOWLEDGMENT

Support from the National Found Awards CMMI-0533330, CMII-0946765, and CBET-0730922 is gratefully acknowledged.

REFERENCES

- [1] GWEC, 2008. The global wind 2008 report. Tech. rep., Global Wind Energy Council.
- [2] Burton, T., Sharpe, D., Jenkins, N., and Bossanyi, E., 2001. *Wind energy handbook*. Wiley.
- [3] Manwell, J., McGowan, J., and Rogers, A., 2010. *Wind Energy Explained: Theory, Design and Application*, 2 ed. Wiley.
- [4] Carta, J., Ramírez, P., and Velázquez, S., 2009. "A review of wind speed probability distributions used in wind energy analysis case studies in the canary islands". *Renewable and Sustainable Energy Reviews*, **13**(5), pp. 933–955.
- [5] Seguro, J., and Lambert, T., 2000. "Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis". *Journal of Wind Engineering and Industrial Aerodynamics*, **85**(1), pp. 75–84.
- [6] Dorvlo, A., 2002. "Estimating wind speed distribution". *Energy Conversion and Management*, **43**(17), pp. 2311–2318.
- [7] Morgan, E., Lackner, M., Vogel, R., and Baise, L., 2011. "Probability distributions for offshore wind speeds". *Energy Conversion and Management*, **52**(1), pp. 15–26.
- [8] Sorensen, P., and Nielsen, T., 2006. "Recalibrating wind turbine wake model parameters - validating the wake model performance for large offshore wind farms". In European Wind Energy Conference and Exhibition.
- [9] Katic, I., Hojstrup, J., and Jensen, N., 1986. "A simple model for cluster efficiency". In Proceedings of European Wind Energy Conference and Exhibition.
- [10] Lackner, M. A., and Elkinton, C. N., 2007. "An analytical framework for offshore wind farm layout optimization". *Wind Engineering*, **31**(1), pp. 17–31.
- [11] Vega, R., 2008. "Wind directionality: A reliability-based approach". PhD thesis, Texas Tech University, Lubbock, TX, August.
- [12] Carta, J., Ramírez, P., and Bueno, C., 2008. "A joint probability density function of wind speed and direction for wind energy analysis". *Energy Conversion and Management*, **49**(6), pp. 1309–1320.
- [13] Erdem, E., and Shi, J., 2011. "Comparison of bivariate distribution construction approaches for analysing wind speed

- and direction data”. *Wind Energy*, **14**(1), pp. 27–41.
- [14] NDAWN, 2010. The north dakota agricultural weather network. <http://ndawn.ndsu.nodak.edu/>.
- [15] Simonoff, J., 1996. *Smoothing Methods in Statistics*, 2 ed. Springer.
- [16] Rosenblatt, M., 1956. “Remarks on some nonparametric estimates of a density function”. *The Annals of Mathematical Statistics*, **27**(3), pp. 832–837.
- [17] Parzen, E., 1962. “On estimation of a probability density function and mode”. *The Annals of Mathematical Statistics*, **33**(3), pp. 1065–1076.
- [18] Jones, M., Marron, J., and Sheather, S., 1996. “A brief survey of bandwidth selection for density estimation”. *Journal of the American Statistical Association*, **91**(433), pp. 401–407.
- [19] Epanechnikov, V., 1969. “Non-parametric estimation of a multivariate probability density”. *Theory of Probability and its Applications*, **14**, pp. 153–158.
- [20] Duong, T., and Hazelton, M., 2003. “Plug-in bandwidth matrices for bivariate kernel density estimation”. *Nonparametric Statistics*, **15**(1), pp. 17–30.
- [21] M.P. Wand, M. J., 1995. *Kernel Smoothing*. CRC Press.
- [22] Wand, W., and Jones, M., 1994. “Multivariate plug-in bandwidth selection”. *Computational Statistics*, **9**, pp. 97–116.
- [23] Crasto, G., 2007. Numerical simulations of the atmospheric boundary layer. Tech. rep., Universit degli Studi di Cagliari, Cagliari, Italy, February.
- [24] Kutz, M., 2006. *Mechanical engineers’ handbook*, 3 ed. Wiley.
- [25] Canavos, G., 1984. *Applied Probability and Statistical Methods*. Little, Brown and Company.
- [26] Gibbons, J., and Chakraborti, S., 2003. *Nonparametric Statistical Inference*. CRC Press.
- [27] Looney, S., and Gullledge, T., 1985. “Use of the correlation coefficient with normal probability plots”. *The American Statistician*, **39**(1), pp. 75–79.
- [28] Forrester, A., Sóbester, A., and Keane, A., 2008. *Engineering Design via Surrogate Modelling: a Practical Guide*. J. Wiley.
- [29] Sobol, I., 1976. “Uniformly distributed sequences with an additional uniform property”. *USSR Computational Mathematics and Mathematical Physics*, **16**(5), pp. 236–242.
- [30] Bratley, P., and Fox, B., 1988. “Algorithm 659: Implementing sobol’s quasirandom sequence generator”. *ACM Transactions on Mathematical Software*, **14**(1), pp. 88–100.

APPENDIX A: PROBABILITY DISTRIBUTION FUNCTIONS

Weibull Distribution

The 2-parameter Weibull distribution is the most widely accepted distribution for wind speed. The Weibull *pdf* and *cdf* are

expressed as

$$f(x; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right] \quad (17)$$

and

$$F(x; \alpha, \beta) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right] \quad (18)$$

where $x \geq 0$.

The estimated shape parameter $\hat{\beta}$ can be solved using an iterative procedure, given by

$$\hat{\beta} = \left[\frac{\sum_{i=1}^n (x_i^{\hat{\beta}} \ln x_i)}{\sum_{i=1}^n x_i^{\hat{\beta}}} - \frac{1}{n} \sum_{i=1}^n \ln x_i \right]^{-1} \quad (19)$$

The estimated scale parameter $\hat{\alpha}$ can be solved using

$$\hat{\alpha} = \left(\frac{1}{n} \sum_{i=1}^n x_i^{\hat{\beta}} \right)^{\frac{1}{\hat{\beta}}} \quad (20)$$

Gamma Distribution

The Gamma *pdf* and *cdf* are expressed as

$$f(x; k, \theta) = x^{k-1} \frac{\exp\left(-\frac{x}{\theta}\right)}{\theta^k \Gamma(k)} \quad (21)$$

and

$$F(x; k, \theta) = \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)} \quad (22)$$

where $x \geq 0$ and $k, \theta > 0$. The estimated parameter \hat{k} can be obtained by solving

$$\ln(\hat{k}) - \psi(\hat{k}) = \ln\left(\frac{1}{n} \sum_{i=1}^n x_i\right) - \frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (23)$$

where $\psi(\hat{k}) = \Gamma'(\hat{k})/\Gamma(\hat{k})$ is the digamma function. The estimated scale parameter $\hat{\theta}$ can be solved using

$$\hat{\theta} = \frac{1}{\hat{k}n} \sum_{i=1}^n x_i \quad (24)$$

Normal Distribution

The Normal *pdf* and *cdf* are expressed as

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (25)$$

and

$$F(x; \mu, \sigma) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \right] \quad (26)$$

The estimated parameter $\hat{\mu}$ and $\hat{\sigma}$ are expressed as

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \hat{\sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (27)$$

Lognormal Distribution

The Lognormal *pdf* and *cdf* are expressed as

$$f(x; \mu_L, \sigma_L) = \frac{1}{x\sqrt{2\pi\sigma_L^2}} \exp\left[-\frac{(\ln(x) - \mu_L)^2}{2\sigma_L^2}\right] \quad (28)$$

and

$$F(x; \mu_L, \sigma_L) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{\ln(x) - \mu_L}{\sigma_L\sqrt{2}}\right) \right] \quad (29)$$

The estimated parameter $\hat{\mu}_L$ and $\hat{\sigma}_L$ are expressed as

$$\hat{\mu}_L = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \hat{\sigma}_L = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_L)^2 \quad (30)$$

Rayleigh Distribution

The Rayleigh *pdf* and *cdf* are expressed as

$$f(x; b) = \frac{x}{b^2} \exp\left(-\frac{x^2}{2b^2}\right) \quad (31)$$

and

$$F(x; b) = 1 - \exp\left(-\frac{x^2}{2b^2}\right) \quad (32)$$

where $x \geq 0$. The estimated parameter \hat{b} is given by

$$\hat{b} = \left(\frac{1}{2n} \sum_{i=1}^n x_i^2 \right)^{1/2} \quad (33)$$