A New Circular Distribution and Its Application to Wind Data

Xu Qin (Corresponding author)

Faculty of Science and State Key Laboratory for Manufacturing Systems Engineering

Xi'an Jiaotong University, 28 Xian Ning Street, Xi'an 710049, China

Tel: 86-29-8267-3324 E-mail: jiayouxuxu@gmail.com

Jiangshe Zhang

Faculty of Science and State Key Laboratory for Manufacturing Systems Engineering

Xi'an Jiaotong University, 28 Xian Ning Street, Xi'an 710049, China

Tel: 86-29-8266-5961 E-mail: jszhang@mail.xjtu.edu.cn

Xiaodong Yan

Key Laboratory of Regional Climate-Environment Research for Temperature East Asia

Institute of Atmospheric Physics, Chinese Academy of Science

Qi Jia Huo Zi, Beijing 100029, China

Tel: 86-10-8299-5275 E-mail: yxd@tea.ac.cn

The research is financed by Asian Development Bank. No. 2006-A171(Sponsoring information)

Abstract

A new circular probability distribution which is based on GvM (generalization of the von Mises) is proposed. The new distribution is used to construct a joint probability distribution which is applied to fit joint distribution of linear and circular variables such as wind speed and wind direction. The results of several numerical experiments show that compared with the existing distribution models, the new circular distribution and the new constructed joint distribution in the paper can provide higher degree of the fit for the wind data under study.

Keywords: Joint probability density, Singly Truncated from below Normal Weibull mixture distribution, Von Mises, Wind speed, Wind direction

1. Introduction

In recent years, the study of wind characteristics has aroused universal concern. What we focus on is the distributions of wind data which include wind speed and wind direction. In available literature on wind energy and other renewable energy sources, different distribution models have been proposed to fit wind speed and wind direction data (Carta, 2009; Carta, 2007; Gatto, 2008; Carta, 2008a; Mooney, 2003; Gatto, 2007). For wind speed which is a linear variable, a Singly Truncated from below Normal Weibull mixture distribution (TNW) (Carta, 2007) and a finite mixture of Weibull distribution (Carta, 2009) are mostly used by authors. Wind direction data are circular data which have no magnitude and can be conveniently represented as points on the circumference of a unit circle centered at the origin or as unit vectors connecting the origin to these points. So the circular probability distribution models are used to fit distribution of wind direction. The von Mises (or circular normal) distribution was the most commonly used circular distribution (Carta, 2008a; Mooney, 2003; Gatto, 2007). However, the literature about joint distributions of wind speed and wind direction is scarce in the specialized literature on wind energy and other renewable sources as stated by José A. Carta (Carta, 2008b). McWilliams et al proposed isotropic Gaussian model in 1979 and then Weber generalized this model to obtain anisotropic Gaussian model (McWilliams, 1979; Weber, 1991). The two models took as their starting points certain hypotheses which made them inconvenient in practice. In 2008, Carta and Ramírez proposed a very flexible joint probability density function of wind speed and direction. In that paper, authors used the angular-linear distribution model proposed by Johnson and Wehrly (Johnson, 1978) to construct joint distributions from information about the shape of the marginal distributions. The joint probability distribution defined by Johnson and Wehrly is

$$f_{V,\Theta}(v,\theta) = 2\pi g(\zeta) f_V(v) f_\Theta(\theta), \quad 0 \le \theta < 2\pi, \ -\infty \le v < \infty \tag{1}$$

where g(·) is the pdf of the circular variable ζ, given by

$$\zeta = 2\pi[F_V(v) - F_\Theta(\theta)]$$

And v represents the wind speed and $v \ge 0$ in this paper. θ represents wind direction. In meteorology wind direction is measured in a clockwise direction using the North as starting point (the y-axis). In (Carta, 2008b) a TNW-pdf and a finite mixture of von Mises distribution were used to fit the marginal distribution of wind speed and wind direction respectively. And a finite mixture of von Mises distribution was used to fit the circular variable ζ.

In this paper, we propose a new circular probability distribution based on the GvM model proposed by Gatto (Gatto, 2007) to fit the circular variable ζ. The probability density function of GvM is given by

$$f_{\mathrm{GvM}}(\theta;\mu_1,\mu_2,\kappa_1,\kappa_2) = \frac{1}{2\pi G_0(\delta,\kappa_1,\kappa_2)}\exp\{\kappa_1\cos(\theta-\mu_1)+\kappa_2\cos 2(\theta-\mu_2)\}$$

for $\theta \in [0, 2\pi)$, $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi)$, $\delta = (\mu_1 - \mu_2) \bmod \pi$, $\kappa_1, \kappa_2 > 0$ and where the normalizing constant is given by

$$G_0 = \frac{1}{2\pi}\int_0^{2\pi}\exp\{\kappa_1\cos\theta + \kappa_2\cos 2(\theta+\delta)\}d\theta \tag{2}$$

In this paper, we add a new parameter to GvM model and obtain a new circular distribution model.

A TNW model and a vMM model in this paper are still applied to fit wind speed and wind direction data just like what were used in the joint model proposed by Carta and Ramírez. But different from the existing joint model, we use the new proposed circular distribution to fit the circular variable ζ in this paper. Consequently we obtain a new joint distribution of wind speed and wind data. For convenience, the existing joint distribution model proposed by Carta and Ramírez and the new joint distribution model proposed in this paper are denoted by T-V-V and T-V-N respectively.

This paper is organized as follows: The models for wind data are given in Section 2. The estimation method of parameters is described in Section 3. Section 4 gives the numerical results. We apply the two joint distribution models to four lattice-point wind data sequences extracted from the NCEP daily reanalysis data at 700Pha in 2005. In addition, a comparison between the new T-V-N model proposed in this paper and the existing T-V-V model is made. Finally, Section 5 provides some concluding remarks.

2. The models

2.1 Model of wind speed

In this paper, we use TNW model to fit wind speed data.

2.1.1 TNW model

The TNW model takes into account the frequency of null winds and, therefore, can represent wind regimes with high percentages of null wind speeds. The form of TNW-pdf is

$$f_{\mathrm{TNW}}(v) = \frac{\omega}{I(\mu,\sigma)}Z(v,\mu,\sigma) + (1-\omega)\frac{\alpha}{\beta}(\frac{v}{\beta})^{\alpha-1}\exp[-(\frac{v}{\beta})^{\alpha}], \quad 0 \le v < \infty$$

where $Z(v,\mu,\sigma)$ and $I(\mu,\sigma)$ are given by

$$Z(v,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{(v-\mu)^2}{2\sigma^2}]$$

$$I(\mu,\sigma) = \frac{1}{\sigma}\int_0^{\infty} Z(v,\mu,\sigma)\mathrm{d}v$$

where β is a scale parameter with the same unit as the random variable and α is a shape parameter; μ and σ are parameters with the same units as the random variable. In this paper, the unit of wind speed data is m/s; ω is the weight in the mixture of the Singly Truncated from below Normal distribution ($0 \le \omega \le 1$).

2.2 Models of circular variable

We use two types of density probability function to represent circular variable in this paper. One is the finite mixture of von Mises distribution (vMM) and the other is the new proposed model based on GvM model.

2.2.1 vMM model

A random variable Θ has a von Mises (vM) distribution, if its probability density function is defined by

$$f_{\mathrm{vM}}(\theta;\kappa,\mu) = \frac{1}{2\pi I_0(\kappa)}\exp[\kappa\cos(\theta-\mu)] \quad 0 \le \theta < 2\pi$$

where $\kappa \ge 0$ and $0 \le \mu < 2\pi$ are parameters. The distribution is unimodal and is symmetrical about $\theta = \mu$. The parameter μ is the mean direction and the parameter κ is known as the concentration parameter. Here, $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero and is given by

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi}\exp(\kappa\cos\theta)d\theta = \sum_{k=0}^{\infty}\frac{1}{(k!)^2}(\frac{\kappa}{2})^{2k}$$

The vMM model is comprised of a weighted sum of M von Mises probability densities $f_{\text{vM}}(\theta;\kappa_j,\mu_j)$. The density function of vMM is given by

$$f_{\text{vMM}}(\theta) = \sum_{j=1}^{M} \omega_j f_{\text{vM}}(\theta;\kappa_j,\mu_j) = \sum_{j=1}^{M} \frac{\omega_j}{2\pi I_0(\kappa_j)} \exp[\kappa_j \cos(\theta-\mu_j)] \quad 0 \le \theta < 2\pi$$

where ω_j is the jth mixture weight,

$$0 \le \omega_j \le 1 \ \ (j = 1, 2, \ldots, M) \ \ \text{and} \ \ \sum_{j=1}^{N} \omega_j = 1$$

2.2.2 The new model

In this paper we add one parameter m to GvM model, equation (2) and propose a new model which is given by

$$f_{\text{New}}(\theta;\mu_1,\mu_2,\kappa_1,\kappa_2,m) = \frac{1}{2\pi N_0(\mu_1,\mu_2,\kappa_1,\kappa_2,m)} \exp\{\kappa_1\cos(\theta-\mu_1) + \kappa_2\cos(m\theta-\mu_2)\}$$

for $\theta \in [0, 2\pi)$, $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, 2\pi)$, $\kappa_1, \kappa_2 \ge 0$, $m \ge 0$ and where the normalizing constant is given by

$$N_0 = \frac{1}{2\pi}\int_0^{2\pi} \exp\{\kappa_1\cos(\theta-\mu_1) + \kappa_2\cos(m\theta-\mu_2)\}d\theta$$

2.3 Joint distribution models of wind speed and wind direction

In this paper, different distributions of linear variable and circular variable are used to construct the joint distribution models of wind speed V and wind direction Θ. We use the equation (1) to represent the joint distribution of wind speed and wind direction.

From n sample wind speed and direction data, n values are calculated of the variable ζ, through

$$\zeta_i = \begin{cases} 2\pi(F_V(v_i) - F_\Theta(\theta_i)), & F_V(v_i) \ge F_\Theta(\theta_i) \\ 2\pi(F_V(v_i) - F_\Theta(\theta_i)) + 2\pi, & F_V(v_i) < F_\Theta(\theta_i) \end{cases}$$

such a way that $0 \le \zeta_i < 2\pi$ where $i = 1, ..., n$. In this paper the vMM distribution and the new proposed distribution are used to fit this sample of ζ.

With these mentioned linear and circular distributions, we get two joint distribution models: T-V-V, T-V-N.

3. Estimation of parameters

The parameters of all the models used in this paper are all estimated by maximum likelihood (ML) method.

For vMM models in this paper, we apply the EM algorithm to the computation of the MLE. The EM algorithm is an iterative method for calculating MLE from incomplete data. The application of the EM algorithm ensures that the likelihood values increase monotonically (McLachlan, 2004).

The parameters of the TNW and the new proposed circular distribution models are estimated by ML method in which the maximization problem is regarded as a nonlinear programming with only inequality constraints and is solved numerically by the classical interior-point method (Byrd, 2000; Waltz, 2006).

4. Numerical experiments

In this part, we will carry numerical experiments for the T-V-V model and T-V-N model. All programs are written in Matlab. The extraction of data used in this paper resorts to GrADS software. We choose four lattice-point wind sequences extracted from NCEP daily reanalysis data at 700Pha in 2005. These lattice points are (100^oE, 22.5^oN), (100^oE, 25^oN), (102.5^oE, 27.5^oN), (105^oE, 27.5^oN) whose location is shown in Figure 1. The geopotential height of 700Pha locates the troposphere. The horizontal components, v_x and v_y, of the wind speed at the four lattice points under study are plotted in Figure 2. Table 1 lists some statistical description of these series where the computations of the circular mean direction, the correlation coefficient of linear variable and circular variable, the estimated prevailing wind direction, are mentioned by Jammalamadaka (2001), Mardia (1976) and McWilliams et al (1980) respectively.

We apply the TNW model and the vMM model to fit wind speed and wind direction data respectively. For the circular variable ζ, the vMM and the new model are used to fit its distribution in turn. All the parameters are estimated by ML method. The estimation results are shown in Table 2. In order to compare the fitted results of the two circular distributions for ζ, the mean circular error, MCE is adopted (Razali, 2008). For circular data, MCE is an analogy to the mean standard error, MSE. MCE is given by

$$MCE = \frac{1}{n}\sum_{i=1}^{n}[1 - \cos(F_n(x_i), F(x_i))]^2$$

where n is the number of observations, that is 365 in this paper. $F(x)$ is a given cumulative distribution. F_n is the empirical distribution function and for n *iid* observation X_j. And it is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leq x}$$

where $I_{X_i \leq x}$ is the indicator function, equal to 1 if $X_i \leq x$ and equal to 0 otherwise. The lower MCE is, the greater the fit. Table 3 lists the results of MCE for ζ of the four wind sequences. It can be seen that the new proposed circular distribution has lower values of MCE than the vMM distribution for each sequence under study.

In this paper, we construct two joint distribution models of wind speed and wind direction which are the existing model T-V-V and the new model T-V-N. For comparison of the two models, we take the coefficient of determination (R^2) as performance index. The coefficient of determination is computed as:

$$R^2 = 1 - \frac{SSE}{\sum_{i=1}^{N}(P_i - \overline{FF})^2}$$

where the sum of squares due to error (SSE) is computed as:

$$SSE = \sum_{i=1}^{N}(P_i - FF_i)^2$$

$\overline{FF}$ is the mean of FF_i values, P_i, $i = 1, ..., N$, are the values of the empirical cumulative relative frequencies and the FF_i, $i = 1, ..., N$, are the theoretical values using the fitted distribution functions. The higher R^2 is, the greater the fit. The results of comparison are shown in Table 4. It can be seen that, for all the lattice points analyzed, the proposed joint distribution T-V-N has a higher degree of fit to the wind data than the existing joint distribution T-V-V.

5. Conclusion

In this paper, we propose a new circular distribution and apply it to construct a joint distribution of wind speed and wind direction. The parameters of the models used in paper are estimated by ML method. The new joint model and the existing model are used to fit four wind sequences extracted from NCEP daily reanalysis data set. In virtue of MCE and R^2 we compare the fitted results of the new and existing models. We conclude that, in the case of the four wind sequences under study, the joint distribution model with the new proposed circular distribution in this paper provides better fits in all the cases analyzed than those obtained with the model used in the specialized literature on wind energy.

References

Byrd, R. H., Gilbert, J. C. & Nocedal, J. (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89, 149-85.

Carta, J.A., Bueno, C. & Ramírez, P. (2008a). Statistical modelling of directional wind speeds using mixtures of von Mises distribution: Case study. *Energy Conversion & Management*, 49, 897-907.

Carta, J.A. & Ramírez, P. (2007). Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. *Energy Conversion & Management*, 48, 281-291.

Carta, J. A., Ramírez, P. & Bueno, C. (2008b). A joint probability density function of wind speed and direction for wind energy analysis. *Energy Conversion & Management*, 49, 1309-1320.

Carta, J.A., Ramírez, P. & Velázquez, S. (2009). A Review of wind speed probability distributions used in wind energy analysis case studies in the Canary Islands. *Energy Conversion & Management*, 13, 933-955.

Gatto, R. (2008). Some computational aspects of the generalized von Mises distribution. *Statistics and Computing*, 18, 321-331.

Gatto, R. & Jammalamadaka, S. R. (2007). The generalized von Mises distribution. *Statistical Methodology*, 4, 341-353.

Jammalamadaka, S. R. & Gupta, A. S. (2001). *Tropic in circular statistics*. World Scientific Publishing.

Johnson, R. A. & Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73, 602-606.

Mardia, K. V. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63, 403-405.

McLachlan, G. & Peel, D. (2004). *Finite mixture models*. New Jersy: John Wiley & Sons.

McWilliams, B., Newmann, M. M. & Sprevak, D. (1979). The probability distribution of wind velocity and direction. *Wind Energy*, 3, 269-273.

McWilliams, B. & Sprevak, D. (1980). The estimation of the parameters of the distribution of wind speed and direction. *Wind Energy*, 4, 227-238.

Mooney, J. A., Helms P. J. & Jolliffe I.T. (2003). Fitting mixture of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics* & *Data Analysis*, 41, 505-513.

Razali, A. M., Ahmad, A., aharim, A. & et al. (2008). Fitting the probability distribution to wind direction data in Bangi. *Seminar on Engineering Mathematics*.

Waltz, R. A., Morales, J. L., Nocedal, J. & et al. (2006). An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107, 391-408.

Weber, R. (1991). Estimator for the standard deviation of wind direction based on moments of the Cartesian components. *Journal of Meteorology Science*, 30, 1341-1353.

Table 1. Summarizing statistics for wind series used in this paper

	wind speed(m/s)			wind direction(rad)		wind speed and direction
sample	maximum	minimum	mean	mean	prevailing direction	correlation coefficient r^2
(100^{o}E, 22.5^{o}N)	17.26	0.32	7.73	1.34	1.39	0.19
(100^{o}E, 25^{o}N)	17.75	0.11	8.42	1.30	1.40	0.58
(102.5^{o}E, 27.5^{o}N)	18.86	0.75	7.71	1.12	1.21	0.43
(105^{o}E, 27.5^{o}N)	20.04	0.32	7.90	1.13	1.18	0.15

Table 2. Parameter estimation values of models

	wind speed					wind direction			ζ					
	TNW					vMM			T-V-V			T-V-N		
sample	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\omega}$	$\hat{\mu}_j$	$\hat{\kappa}_j$	$\hat{\omega}_j$	$\hat{\mu}_j$	$\hat{\kappa}_j$	$\hat{\omega}_j$	$\hat{\mu}_j$	$\hat{\kappa}_j$	$\hat{m}$
(100^{o}E, 22.5^{o}N)	11.21	2.73	1.88	6.24	0.39	1.45	7.36	0.72	0.44	0.27	0.51	1.65	0.37	26.67
						5.40	0.78	0.28	1.55	1.18	0.49	3.47	0.12	
(100^{o}E, 25^{o}N)	12.17	2.60	2.16	6.65	0.40	1.45	11.97	0.78	1.60	0.53	0.56	1.42	0.33	28.91
						6.28	1.23	0.22	6.28	00.06	0.44	2.58	0.03	
(102.5^{o}E, 27.5^{o}N)	14.04	1.86	2.19	7.97	0.09	1.23	10.80	0.79	1.28	0.97	0.63	1.23	0.67	37.72
						6.28	1.05	0.21	6.28	0.17	0.37	4.49	0.09	
(105^{o}E, 27.5^{o}N)	9.92	4.36	1.87	7.59	0.35	1.21	10.89	0.71	2.14	0.31	0.57	1.84	0.71	8.64
						6.28	0.46	0.29	1.76	1.50	0.43	5.11	0.20	

Table 3. The comparison of fitted results for circular variable ζ

	MCE	
sample	TNW	New
(100^{o}E, 22.5^{o}N)	0.0427	**0.0421**
(100^{o}E, 25^{o}N)	0.0465	**0.0446**
(102.5^{o}E, 27.5^{o}N)	0.0460	**0.0446**
(105^{o}E, 27.5^{o}N)	0.0423	**0.0415**

Table 4. The comparison of fitted results for joint distribution models

sample	R^2 T-V-V	R^2 T-N-V
(100^oE, 22.5^oN)	0.9987	**0.9989**
(100^oE, 25^oN)	0.9990	**0.9992**
(102.5^oE, 27.5^oN)	0.9981	**0.9983**
(105^oE, 27.5^oN)	0.9971	**0.9976**

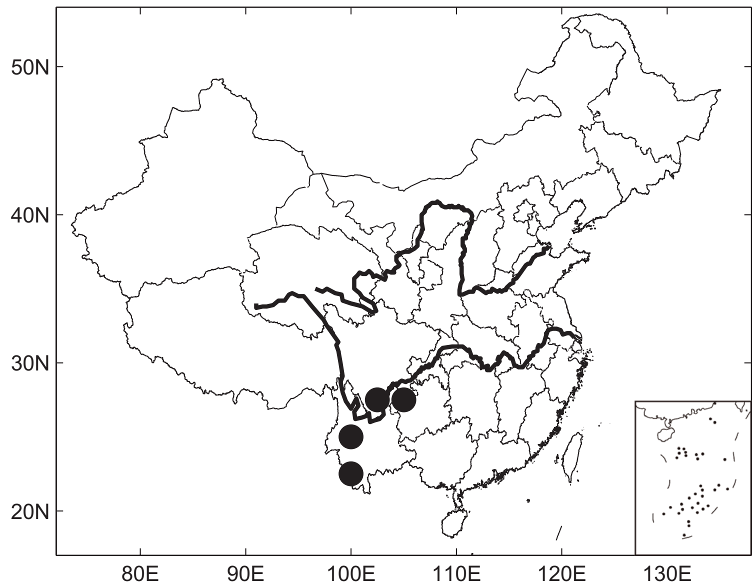

Figure 1. The location of the lattice points used

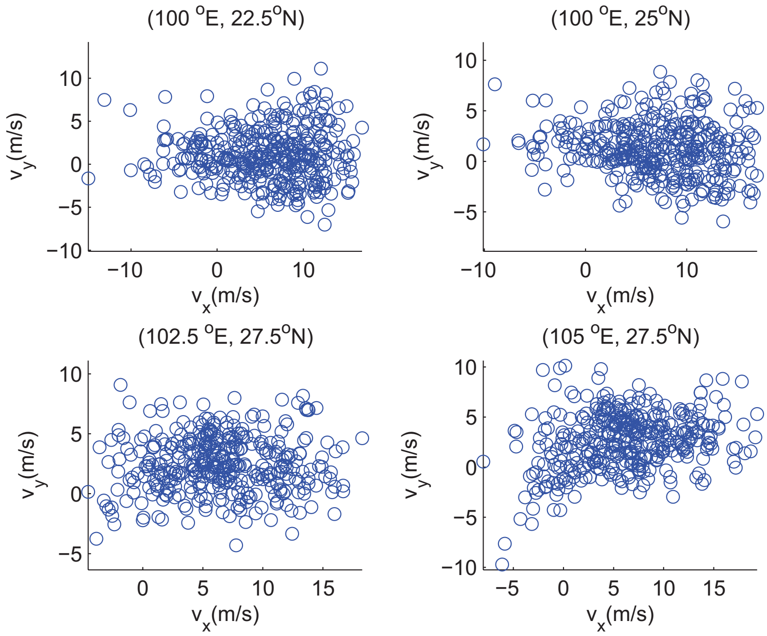

Figure 2. Horizontal components of the wind speed