Energy Conversion and Management 72 (2013) 94–102

Contents lists available at SciVerse ScienceDirect

Energy Conversion and Management

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

Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia

N. Masseran [a], A.M. Razali [a,b,*], K. Ibrahim [a,b], M.T. Latif [c]

[a] Centre for Modelling and Data Analysis (DELTA), School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia

[b] Solar Energy Research Institute (SERI), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia

[c] School of Environmental & Natural Resource Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia

ARTICLE INFO

Article history:
Available online 11 April 2013

Keywords:
Circular density
Mixture of von Mises distribution
Statistical model
Wind direction
Wind regime

ABSTRACT

A statistical distribution for describing wind direction provides information about the wind regime at a particular location. In addition, this information complements knowledge of wind speed, which allows researchers to draw some conclusions about the energy potential of wind and aids the development of efficient wind energy generation. This study focuses on modeling the frequency distribution of wind direction, including some characteristics of wind regime that cannot be represented by a unimodal distribution. To identify the most suitable model, a finite mixture of von Mises distributions were fitted to the average hourly wind direction data for nine wind stations located in Peninsular Malaysia. The data used were from the years 2000 to 2009. The suitability of each mixture distribution was judged based on the R^2 coefficient and the histogram plot with a density line. The results showed that the finite mixture of the von Mises distribution with H number of components was the best distribution to describe the wind direction distributions in Malaysia. In addition, the circular density plots of the suitable model clearly showed the most prominent directions of wind blows than the other directions.

1. Introduction

Wind direction is one of the most important factors in evaluating the characteristics of the wind regime of a particular region. Wind direction has a substantial impact on the lives of human beings. Wind direction impacts civil engineering structures, such as towers, bridges and tall buildings [1]. In addition, wind direction has been recognized as an important aspect in the evaluation of wind energy because information about wind direction can complement information about wind speed to aid in drawing conclusions about energy potential. Carta et al. [2] stated that knowledge of wind direction characteristics enables wind turbines to be positioned in such a way as to maximize the amount of captured energy. Unfortunately, we have found that most studies regarding the potential of wind do not describe information about the characteristics of wind direction well, perhaps because of the difficulties involved in the analysis and complex statistical modeling of wind direction compared to wind speed. In this study, we try to overcome some specific problems involved in the statistical model for wind direction data to gain some insight into the characteristics of wind directions in Malaysia.

Several studies on wind direction have been conducted for the purposes of energy modeling and evaluation [2–4]. Carta et al. [3] proposed a finite mixture of the von Mises distribution for modeling the directional wind data from several wind stations in the Canary Islands. The suitability of the distributions was judged according to the coefficient of determination, R^2. Carta and colleagues found that a mixture of von Mises distributions provided a flexible model for studies of wind direction that could be applied to representations of wind direction regimes in regions with several modes or prevailing wind directions. They additionally concluded that as the number of components N of the mixture distribution increases, the value of the R^2 coefficient increases, but that the variations in R^2 were not significant for $N > 6$. Kamisan et al. [4] fit four types of circular distributions, namely the von Mises distribution, the circular uniform distribution, the wrapped-normal distribution and the wrapped-Cauchy distribution to southwesterly Malaysian wind direction data. Two indicators, the mean circular distance and chord length, were used to determine which distributions gave the best fit. Their results indicated that the von Mises distribution gave the best fit for all of the stations under study. Alternatively, Razali et al. [1] determined the best distribution to fit wind direction data from a station at the Universiti Kebangsaan Malaysia (UKM). Three types of circular

* Corresponding author at: Centre for Modelling and Data Analysis (DELTA), School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia. Tel.: +60 3 89215788; fax: +60 3 89254519.

E-mail addresses: mahir@ukm.my, amrazali@yahoo.com (A.M. Razali).

0196-8904/$ - see front matter
http://dx.doi.org/10.1016/j.enconman.2012.11.025

distributions, the von Mises distribution, the generalized von Mises distribution and the wrapped-Cauchy distribution, were fit to the data. The most suitable distribution was determined using a histogram with a density line and a mean circular error. Their results showed that the von Mises distribution fit the data very well compared to the other distributions. Thus, the researchers concluded that the von Mises distribution can be used to forecast wind direction. Conversely, Carta et al. [2] proposed a more flexible joint probability density of wind speed and direction by using the angular–linear distributions obtained by the Normal–Weibull mixture distribution as a marginal distribution for wind speed and the finite mixture of von Mises as a marginal distribution for wind direction. They concluded that the joint distribution proposed is representative of wind directions with several modes. The joint distribution proposed is also representative of the unimodal, bimodal, and bitangential wind speed frequency distributions. However, we believe that their proposed method is rather complex and is quite cumbersome in some applications. Thus, in this study, we try to identify more parsimonious and suitable models of the distributions of wind direction for several wind stations in Malaysia to gain some insight into the wind regime as well as the wind energy potential.

2. Study area and data

Malaysia is a country that lies entirely in the equatorial zone, situated in the southeastern part of Asia. Its geographic coordinates are 2°30′ north latitude and 112°30′ east longitude. Throughout the year, Malaysia experiences a wet and humid climate with daily temperature ranging from 25.5 to 35 °C. The wind that blows across the peninsula as well as Sabah and Sarawak is influenced by the monsoon seasons, namely the southwest monsoon, the northeast monsoon and two short inter-monsoons. The two monsoons that contribute to the rainy seasons are the southwest monsoon, which occurs from May until September, and the northeast monsoon, which occurs from November until March. The later monsoon brings about heavier rainfall on the peninsula, and the most affected areas are in the east and south. Malaysia is a maritime country that is also influenced by the effects of sea breezes and land breezes, especially when the sky is not cloudy. During most afternoons, sea breezes occur with speeds of 10–15 knots. However, at night, the reverse process occurs: weak land breezes occur in the coastal areas [5].

The data used in this study were obtained from the Department of the Environment Malaysia and the Malaysian Meteorological Department. Nine stations were selected for this study: Mersing, Kuala Terengganu Airport, Malacca, Bayan Lepas, Ipoh, Kota Bahru, Balok Baru, Perai and Kangar (shown in Table 1 and Fig. 1). The stations selected in this study were recommended by Masseran et al. [6–8] to be investigated in greater detail with regards to wind energy evaluation and production in Malaysia. In this study, the hourly wind direction data from January 1, 2007 to November 30, 2009 were used. Wind direction data are circular because they are recorded in terms of degrees, from 0° to 360°. However, for modeling, data transformation to radian units can be easily performed.

Table 1
Geographic coordinates and altitudes for each station.

Stations	Latitude	Longitude
Mersing	2°27′N	105°50′E
Kuala Tereggανu	5°23′N	103°06′E
Malacca	2°16′N	102°15′E
Bayan Lepas	5°18′N	100°16′E
Ipoh	4°34′N	101°06′E
Kota Bahru	6°10′N	102°17′E
Balok Baru	3°57′N	103°22′E
Perai	5°23′N	100°24′E
Kangar	6°25′N	100°11′E

3. Wind direction sensor

The wind direction sensor used by the Department of the Environment Malaysia to collect hourly wind direction data for each station was provided by Met One Instruments. The 020C Wind Direction Sensor provides azimuth data for use in micrometeorological measurements related to operational studies and research. The lightweight airfoil vane is directly coupled to a single precision potentiometer. This sensor is especially useful when a low starting threshold, a high damping ratio, or a short delay distance is required. Fig. 2 shows the 020C Wind Direction Sensor model, while Table 2 describes the specifications of the 020C Wind Direction Sensor, including its characteristics, its accuracy and its range value, for more details, please refer to [9].

4. Methods

Several types of circular distributions have been used to model the wind direction data of particular regions, for example, the von Mises distribution, the generalized von Mises distribution, the finite mixture of von Mises, the wrapped-Cauchy, the circular uniform distribution and the wrapped-normal distribution, among many others. However, the von Mises and the finite mixture of von Mises are among the most commonly used in modeling wind directional data; for examples, see [1–4,10–14] and many more. A finite mixture of von Mises is a flexible model for dealing with wind direction data that have several modes [3]. Thus, in this study, the single and the mixture of the von Mises distributions were used as candidate models for the wind direction data in Malaysia before a further analysis was performed to investigate the potential of wind energy.

4.1. The finite mixture of von Mises distributions

The von Mises distribution is a probability distribution function whose total probability is concentrated in the circumference of a unit circle. It was introduced by von Mises in 1918, and its importance and its similarities to the Normal distribution have been emphasized by Gumbel et al. [15]. From the point of view of statistical inference, the von Mises distribution is the most commonly used model for modeling circular data. Let θ be a random variable representing wind direction in radians units. Thus, the probability density function for a single von Mises distribution can be written as

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta-\mu)} \tag{1}$$

where θ is a random variable representing the wind direction in radians, $0 \leqslant \mu < 2\pi$ is the mean direction and $\kappa \geqslant 0$ is a concentration parameter. while $I_0(\kappa)$ in the normalizing constant is the modified Bessel function of the first kind and of order zero, given by

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

The cumulative distribution function for the von Mises mixture distribution is given by

$$F(\theta;\mu,\kappa) = \frac{\left\{\theta I_0(\kappa) + 2\sum_{p=1}^{\infty} \frac{I_p(\kappa)\sin p(\theta-\mu)}{p}\right\}}{2\pi I_0(\kappa)} \tag{3}$$

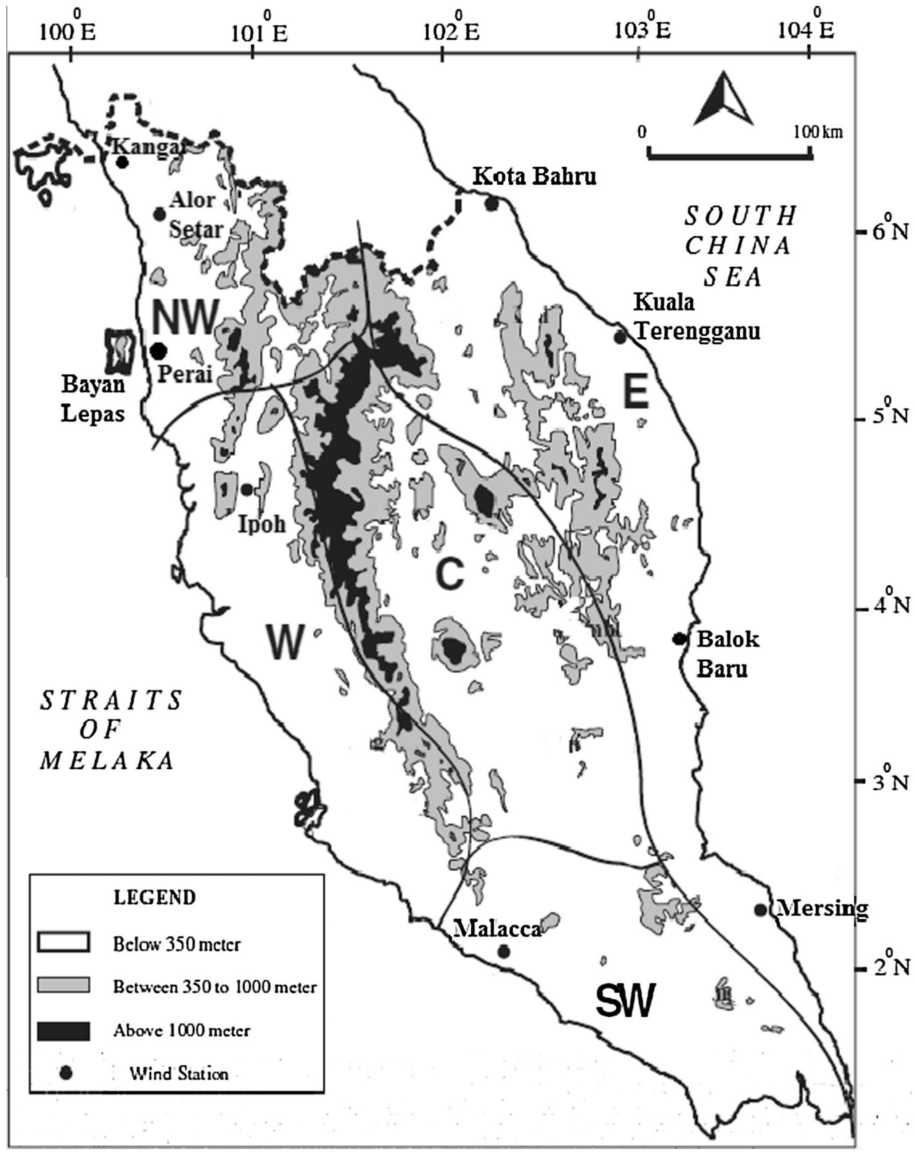

Fig. 1. Locations of wind stations in Peninsular Malaysia.

However, in some applications of wind direction modeling, the observed wind data cannot be represented by a unimodal distribution. To overcome this problem, a finite von Mises mixture distributions (FVMM) which comprises a sum of H von Mises probability distributions, has been proposed. This distribution is given by

$$f(\theta;\mu_h,\kappa_h,\omega_h)=\sum_{h=1}^{H}\omega_h\frac{1}{2\pi I_0(\kappa_h)}e^{\kappa_h\cos(\theta-\mu_h)} \tag{4}$$

where θ is a random variable representing the wind direction in radians units, μ_h, κ_h is composed of a mean direction parameter and concentration parameter, respectively, for $h = 1,2,\ldots,H$ components of the von Mises distribution, while ω_h is a mixing parameter of nonnegative quantities that sum to one, given by

$$0\leqslant\omega_h\leqslant 1\quad\text{and}\quad\sum_{h=1}^{H}\omega_h=1\quad\text{for}\quad(h=1,2\ldots,H) \tag{5}$$

The cumulative distribution function for the von Mises mixture distribution is given by

$$F(\theta;\mu_h,\kappa_h,\omega_h)=\sum_{h=1}^{H}\omega_h\frac{\left\{\theta I_0(\kappa_h)+2\sum_{p=1}^{\infty}\frac{I_p(\kappa_h)\sin p(\theta-\mu_h)}{p}\right\}}{2\pi I_0(\kappa_h)} \tag{6}$$

Conversely, let $\mathbf{x}'=[\cos\theta_i,\ \sin\theta_i]'$ be circular data in terms of rectangular coordinates, with θ as a random variable representing the wind direction in radians unit, as mention in Eqs. (1) and (4). Thus, the von Mises mixture distribution can also be written as

$$f(\mathbf{x};\boldsymbol{\mu}_h,\kappa_h)=\sum_{h=1}^{H}\omega_h\frac{1}{2\pi I_0(\kappa_h)}e^{(\kappa_h\mathbf{x}'\boldsymbol{\mu}_h)} \tag{7}$$

where $||\boldsymbol{\mu}|| = 1$, and $\kappa > 0$. This type of von Mises mixture distribution describes the model for circular data in terms of rectangular coordinates. Readers who are interested in a discussion of the von Mises and the von Mises mixture distributions are referred to [16–18].

4.2. The expected maximization algorithm for parameter estimation

Banerjee et al. [18] provided the solution for the parameter estimate of a von Mises mixture distribution in Eq. (7) using an

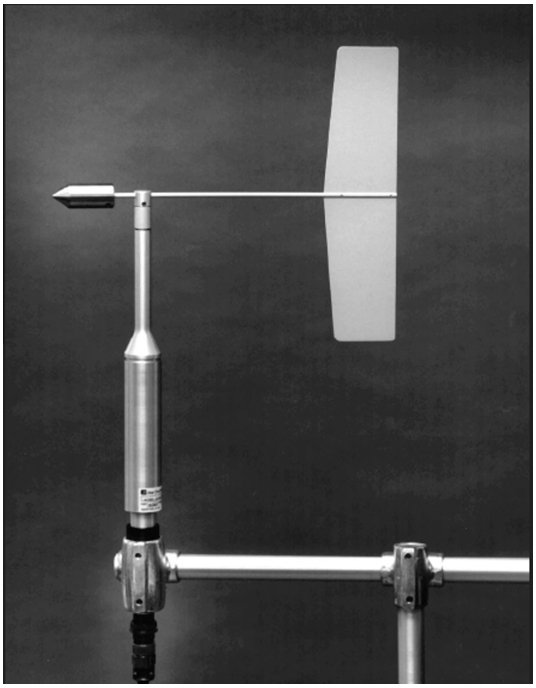

Fig. 2. The 020C Wind Direction Sensor model [9].

Table 2
020C Wind Direction Sensor specifications [9].

Performance characteristics	
Azimuth	Electrical 0–357° Mechanical 0–360°
Threshold	0.5 mph (0.22 m/s)
Linearity	b of full scale
Accuracy	±3° Resolution <0.1°
Damping ratio	Standard 0.6 (magnesium tail) (meet EPA specification)
Delay distance	Less than 3 ft (91 cm)
Temperature range	−50 °C to +65 °C (−58°F to +149°F)
Electrical characteristics	
Power requirements	12 VDC at 10 mA, 12 VDC at 250 mA for internal heater
Output signal selectable	a. 0–5 V for 0–360° b. 0–2.5 V for 0–360°
Output impedance	100 Ω Maximum
Physical characteristics	
Weight	1.5 lbs (0.68 kg)
Finish	Clear anodized aluminium
Cable and mounting	
PN 1957 Mounting	Cable assembly; specify length in feet or meters PN191 Crossarm assembly (contains orientation lock)

expectation maximization (EM) algorithm as an alternative method for finding the maximum likelihood estimates. Let $\boldsymbol{\alpha}_h = (\boldsymbol{\mu}_h, \kappa_h)$ denote the parameter of the von Mises mixture distribution from Eq. (7) for $1 < h < H$. Thus, the von Mises mixture distribution can simply be written as

$$f(\mathbf{x};\boldsymbol{\Theta}) = \sum_{h=1}^{H} \omega_h f_h(\mathbf{x}|\boldsymbol{\alpha}_h) \tag{8}$$

where $\Theta = (\omega_1, \omega_1, \ldots, \omega_H, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_H)$. According to Banerjee et al. [18], to generate a random sample from this mixture distribution, the h-th von Mises distribution will be randomly chosen with probability ω_{hk}. Then, the random samples from $f_h(\mathbf{x}|\boldsymbol{\alpha}_\mathbf{h})$ will be generated. Let, $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}'$ be the observed data following a von Mises mixture distribution from Eq. (8), and let $Z = \{z_1, z_2, \ldots, z_n\}'$ be the corresponding set of hidden random variables that indicate a particular von Mises distribution from which a sample is generated. Particularly, $z_h = h$ if $\mathbf{x}_i$ is generated from $f_h(\mathbf{x}|\boldsymbol{\alpha}_h)$. Thus, the log-likelihood can be written as

$$\log f(X, Z|\boldsymbol{\Theta}) = \sum_{i=1}^{H} \log(\omega_{z_i} f_{z_i}(\mathbf{x}_i|\boldsymbol{\alpha}_{z_i})) \tag{9}$$

Suppose that the posterior distribution, $p(h|\mathbf{x}_i, \Theta)$, of the hidden variables $Z|(X, \Theta)$ is known. From here, the expectation step (E-step) is computed as the expectation of the log-likelihood evaluated using the current estimate for the parameters. In the maximization step (M-step), the parameter estimates are computed by maximizing the expected log-likelihood obtained in the E-step. In brief, the parameter estimates for the mixture von Mises are given by

$$\hat{\omega}_h = \frac{1}{n}\sum_{i=1}^{n} p(h|\mathbf{x}_i, \boldsymbol{\Theta}) \tag{10}$$

$$\hat{\boldsymbol{\mu}}_h = \frac{\mathbf{r}_h}{\|\mathbf{r}_h\|} \tag{11}$$

$$\frac{I_{d/2}(\hat{\kappa}_h)}{I_{d/2-1}(\hat{\kappa}_h)} = \frac{\|\mathbf{r}_h\|}{\sum_{i=1}^{N} p(h|\mathbf{x}_i, \boldsymbol{\Theta})} \tag{12}$$

where $\mathbf{r}_h = \sum_{i=1}^{n} \mathbf{x}_i p(h|\mathbf{x}_i, \boldsymbol{\Theta})$. Readers that would like a detailed discussion of the parameters estimates for the von Mises mixture distribution using the EM algorithm can refer to [17–19].

4.3. Evaluating goodness of fit using the R^2 coefficient

In this study, the R^2 coefficient was used to evaluate the goodness of fit for each fitted model. A large value of R^2 indicates a mixture distribution that fits the data well. R^2 has been used for goodness of fit comparisons because it quantifies the correlation between the observed probabilities and the predicted probabilities, based on a particular distribution. The R^2 coefficient is determined as

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{F}_i - \overline{F})^2}{\sum_{i=1}^{n}(\widehat{F}_i - \overline{F})^2 + \sum_{i=1}^{n}(F_i - \widehat{F}_i)^2} \tag{13}$$

where F_i is a set of empirical cumulative probabilities, $\widehat{F}_i$ is a set of estimated cumulative probabilities for the suitable model and $\overline{F} = \frac{\sum_{i=1}^{n}\widehat{F}_i}{n}$. The estimated cumulative probabilities of $\widehat{F}_i$ were derived from the cumulative distribution function of the proposed model. A large R^2 indicates a better model fit of $\widehat{F}_i$ to the empirical cumulative probabilities F_i. The R^2 coefficient has been used by several researchers to measure goodness of fit, see [2,3,20–22]. The plot of the histogram with a density line is also shown to support our conclusion.

4.4. Evaluating goodness of fit using the mean absolute percentage error

The mean absolute percentage error (MAPE) is an index that is very useful in measuring the accuracy of a fitted model in statistics. The MAPE is calculated by computing the difference between the observed data and the data simulated from the model, which can be written as

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\theta_i - \hat{\theta}_i}{\theta_i}\right|$$

where θ_i is the observed wind direction data and $\hat{\theta}_i$ is the simulated data from the fitted von Mises model. The absolute value in this calculation is summed for every fitted data point in time and divided again by the number of fitted data points n. Multiplying this figure by 100 transforms it into a percentage error. A small MAPE value indicates a mixture distribution that fits the data well. The results for the MAPE are compared with the R^2 coefficients.

5. Results and discussion

As mentioned above, the objective of this study was to identify the most suitable mixture distribution for the wind direction at the Mersing station to gain insight into the wind regime at this location. Firstly, the Mersing station was chosen as an example to show in detail how the analysis was carried out; the same method was followed for the other stations. Table 3 shows the parameter estimates of the von Mises mixture distribution (H = 1, 2,...,8) based on the EM algorithm method for the Mersing station. The parameter means, μ, are defined in terms of rectangular coordinates, $\boldsymbol{\mu}' = [\cos\mu, \sin\mu]'$. However, transforming the results into degrees, $0^\circ \leqslant \mu < 180^\circ$, may be more appropriate, as shown in Table 4. Fig. 3 represents the fitted von Mises mixture distribution (H = 1, 2,...,8) for the wind direction at the Mersing station. Here, the single von Mises distribution (H = 1) failed to accurately model the wind direction data at the Mersing station. However, when the number of components N of the mixture distribution increased, the FVMM models adequately fit the observed data. Alternatively, for H = 6, 7 and 8 the fitted FVMM models have an approximately similar (to each other) accuracy for modeling the observed data. To support our conclusion, the R^2 coefficient and the MAPE were used to evaluate the goodness of fit for each fitted model

Fig. 4a shows the R^2 coefficient for each fitted FVMM model. The R^2 values increase significantly for H = 1, 2, 3, 4 and 5. From the results obtained, the relevance of the increasing value of R^2 to the precise modeling of the actual modality of the wind direction histogram is clear because the R^2 measurements indicate how much each of the FVMM models can be used to describe the observed data. Thus, a higher value of R^2 will provide an analysis of the data model that is more accurate. For example, the R^2 for an FVMM with H = 1 component is approximately 0.93. This value of R^2 is quite high. In terms of regression analysis, a value of $R^2 > 0.7$ is usually considered good enough for data modeling and forecasting. However, in terms of fitting a statistical distribution, modeling and analysis need to be more precise. Thus, a value of $R^2 > 0.7$ is not sufficient, particularly in the case of multi-modal data. This argument is substantiated by the case of an FVMM model with H = 1 component. Because the value of R^2 for the FVMM with $H = 1$ component is approximately 0.93, most of the data can still be modeled. However, in terms of precision, some of the modality of the data cannot be modeled in an accurate way. Thus, by increasing the number of components of the FVMM model, the value of R^2 also increases, and consequently, with the highest values of R^2 provided by the FVMM models with H = 6, 7 and 8 components, most of the data, including the modality of the data, can be modeled accurately.To examine in more detail the role of the R^2 value in determining the best model for wind direction data, it is interesting to make a comparison between FVMM (H = 5) and FVMM (H = 6, 7, 8). From Fig. 3, it is clear that the densities are very similar for H > 5. The R^2 value for the FVMM model with H = 5 components is approximately 0.982, and by increasing the components to H = 6, 7, and 8, the R^2 value increases to be more than 0.999. However, the variations in R^2 are not pronounced for H = 6, 7 and 8, as shown in Fig. 4a. It is difficult to measure the significant differences between FVMM (H = 5) and FVMM (H = 6, 7, 8) based on the figure, however, R^2 clearly shows that FVMM (H = 6, 7, 8) is the best model for the data: more than 99.9% of the actual data can be modeled in a precise way with FVMM (H = 6, 7, 8), while 98.2% of the data can be modeled with FVMM (H = 5). Because our objective is to select the best model, FVMM (H = 6) is more preferable than FVMM (H = 5), particularly in modeling the modality of the data. By this argument, it is clear that the results determined by R^2 are very important for supporting and strengthening every decision made regarding the best model selected from the histograms with a density line. Thus, we conclude that the most suitable model for the wind direction at the Mersing station is FMVM with H = 6 components. Because the 'best' model for describing the wind direction at the Mersing station has been determined, it is reasonable to extract some valuable information from the model. Fig. 4b shows a circular density plot for the mixture of von Mises with H = 6 components. Fig. 4b clearly shows that most of the wind was blowing from the north–north-east, the west-south-west and some from the east-south-east. The most prominent wind direction for the Mersing station corresponded to the parameter μ, 215–240°, 6–12° and 103°, as shown in Table 4. The other directions show an approximately uniform dispersion, which indicates that the wind direction

Table 3
The parameter estimates for the FVMM (H = 1, 2, 3, 4, 5, 6, 7 and 8) based on the EM algorithm.

FVMM	Parameter estimates			
	μ'		κ	ω
H = 1	−0.1557523	−0.9877961	0.477	1
H = 2	0.9582439	0.2859522	0.5156	0.6487694
	−0.5699040	−0.8217113	11.431	0.3512306
H = 3	−0.5238709	−0.8517977	22.747	0.2699097
	−0.7521964	−0.6589390	0.1788	0.5835522
	0.9784459	0.2065034	33.990	0.1465381
H = 4	−0.7348117	−0.6782712	3.028	0.2511905
	0.8617943	0.5072579	0.465	0.4275728
	−0.4928733	−0.8701011	40.87	0.1927872
	0.9792641	0.2025880	40.78	0.1284494
H = 5	−0.1772620	0.9841637	2.084	0.1468458
	0.9873568	−0.1585139	1.636	0.2220188
	0.9769764	0.2133474	41.24	0.1277784
	−0.4876439	−0.8730426	42.96	0.1850011
	−0.7300458	−0.6833982	3.066	0.3183558
H = 6	−0.4885167	−0.8725546	39.91	0.1987266
	−0.8165384	−0.5772912	3.58	0.2373109
	0.9769764	0.2133474	44.64	0.1175061
	0.9943873	0.1058016	2.34	0.1886927
	−0.07033087	−0.99752372	1.52	0.1332170
	−0.2408133	0.9705715	3.01	0.1245467
H = 7	0.9784459	0.2065034	46.59	0.11328076
	0.9543903	−0.2985617	1.15	0.08397336
	0.7736686	−0.6335905	1.18	0.08699198
	0.8906607	0.4546685	2.45	0.13801148
	−0.4893890	−0.8720656	39.95	0.19849876
	−0.7976502	−0.6031204	2.68	0.31353498
	−0.1527882	−0.9882590	14.03	0.06570868
H = 8	−0.99823590	−0.05937241	4.59	0.11383551
	0.8225021	0.5687621	1.53	0.10857548
	0.9790610	0.2035672	43.22	0.12334809
	−0.109171	−0.994023	6.66	0.22184295
	0.6277124	−0.7784453	1.85	0.12345291
	−0.3720380	0.9282175	51.4	0.04379492
	−0.4841478	−0.8749862	49.4	0.15657467
	0.8225021	0.5687621	1.53	0.10857548

Table 4
The parameter μ for each mixture distribution in terms of degrees.

h	$H = 1$	$H = 2$	$H = 3$	$H = 4$	$H = 5$	$H = 6$	$H = 7$	$H = 8$
	μ	μ	μ	μ	μ	μ	μ	μ
1	261.06	16.635	238.42	222.73	100.20	240.77	11.903	183.42
2		235.25	221.22	30.510	350.86	215.28	342.62	34.68
3			11.89	240.47	12.33	12.297	320.71	11.75
4				11.66	240.80	6.091	27.058	232.78
5					223.12	265.96	240.72	308.87
6						103.96	217.12	111.86
7							261.23	241.05
8								34.68

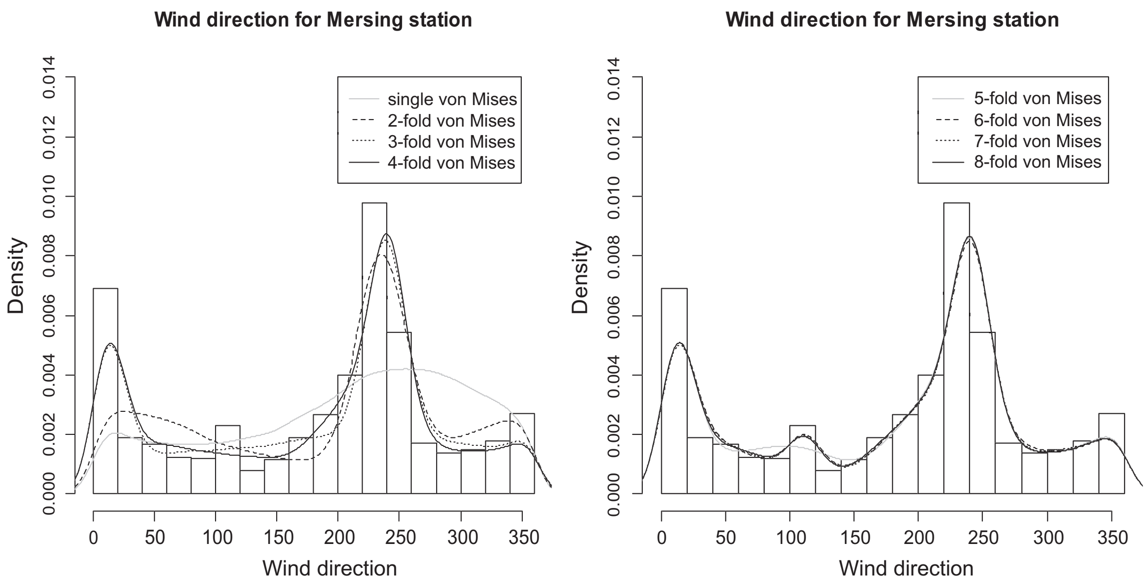

Fig. 3. The FVMM (H = 1, 2, 3, 4, 5, 6. 7 and 8) for wind direction in Mersing.

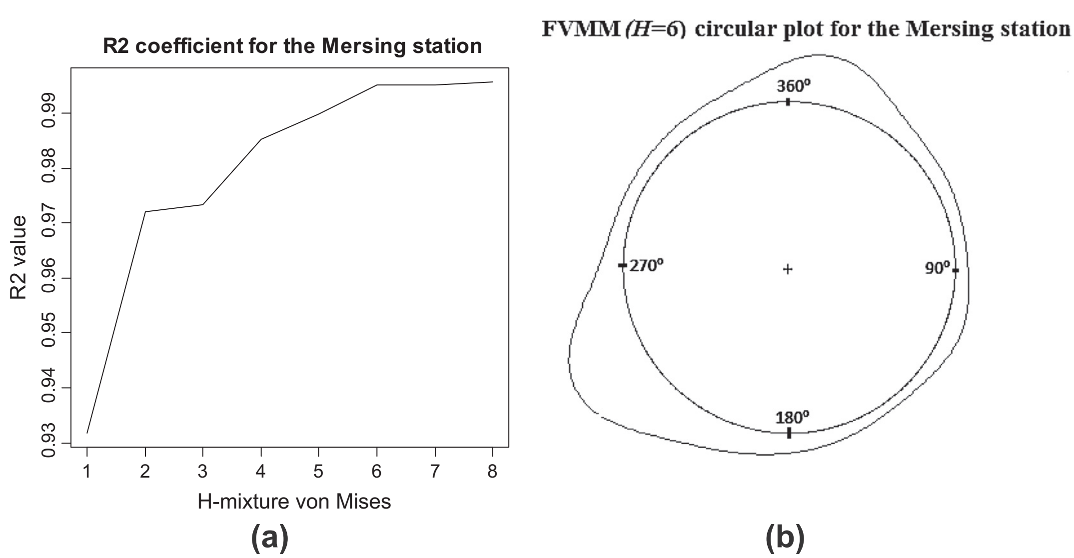

Fig. 4. (a) R^2 value for each fitted FVMM model. (b) Circular density plot for the FVMM (H = 6) model at the Mersing station.

at the Mersing station is not distributed uniformly; instead, some directions are more dominant than others. Using the same procedure, the most suitable mixture distributions of the wind directions for the other stations have been determined. Table 5 shows the values derived based on the R^2 coefficient for each station. From Table 5, it can be observed that increasing the number of von Mises components in the mixture distribution will increase the value of the R^2 coefficient, which indicates a better fit to the data. Table 6 shows the results of the MAPE analysis for a better comparison with the R^2 values for each station.

Table 5
Evaluating the goodness-of-fit for the finite mixture of von Mises distributions based on the R^2 coefficient.

Stations	R^2 value ($H = 1$)	R^2 value ($H = 2$)	R^2 value ($H = 3$)	R^2 value ($H = 4$)	R^2 value ($H = 5$)	R^2 value ($H = 6$)	R^2 value ($H = 7$)	R^2 value ($H = 8$)
K.Terengganu	0.917277	0.996750	0.996326	0.997213	0.997039	0.998149	0.998262	0.997443
Malacca	0.931651	0.997128	0.997286	0.997442	0.995389	0.995614	0.995971	0.995984
Bayan Lepas	0.936980	0.983166	0.985322	0.990281	0.991268	0.999286	0.999188	0.999261
Ipoh	0.944612	0.997505	0.998112	0.999776	0.999750	0.999784	0.999878	0.999690
Kota Bahru	0.925208	0.980947	0.991921	0.998010	0.998242	0.998783	0.999412	0.999136
Balok Baru	0.986861	0.987625	0.992229	0.999880	0.999665	0.999837	0.999888	0.999943
Perai	0.945639	0.990890	0.998476	0.999446	0.999918	0.999865	0.999956	0.999946
Kangar	0.939972	0.999397	0.999760	0.999861	0.999397	0.999852	0.999964	0.999890

Table 6
Evaluating the goodness-of-fit for the finite mixture of von Mises distributions based on the MAPE.

Stations	MAPE ($H = 1$)	MAPE ($H = 2$)	MAPE ($H = 3$)	MAPE ($H = 4$)	MAPE ($H = 5$)	MAPE ($H = 6$)	MAPE ($H = 7$)	MAPE ($H = 8$)
Mersing	6.482	3.142	3.211	3.138	2.112	2.005	2.016	2.021
K.Terengganu	7.327	2.169	2.113	2.071	1.662	1.636	1.734	1.726
Malacca	5.221	2.455	2.119	2.168	2.142	2.114	2.103	2.106
Bayan Lepas	9.322	3.652	3.568	3.414	3.417	2.411	2.383	2.387
Ipoh	6.265	3.433	3.242	3.324	2.770	2.351	2.127	2.134
Kota Bahru	9.696	6.248	3.925	2.898	2.242	2.191	1.211	1.131
Balok Baru	7.822	4.103	2.671	1.912	2.110	1.658	1.666	1.665
Perai	8.676	3.142	2.882	2.860	2.814	2.232	2.044	1.931
Kangar	6.320	1.572	1.771	1.719	1.711	1.715	1.658	1.192

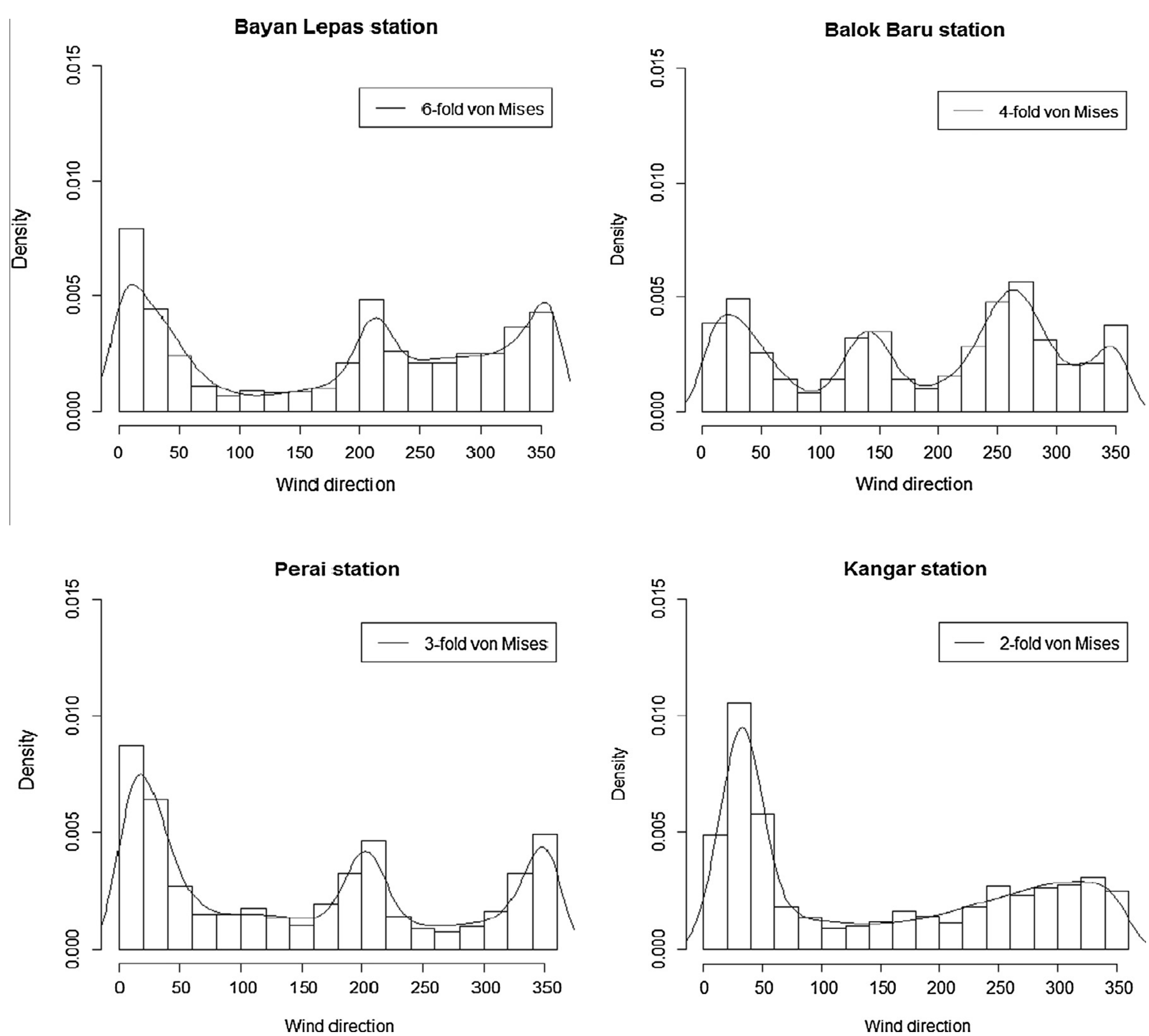

Fig. 5. The FVMMs for wind directions in Bayan Lepas ($H = 6$), Balok Baru ($H = 4$), Perai ($H = 3$) and Kangar ($H = 2$).

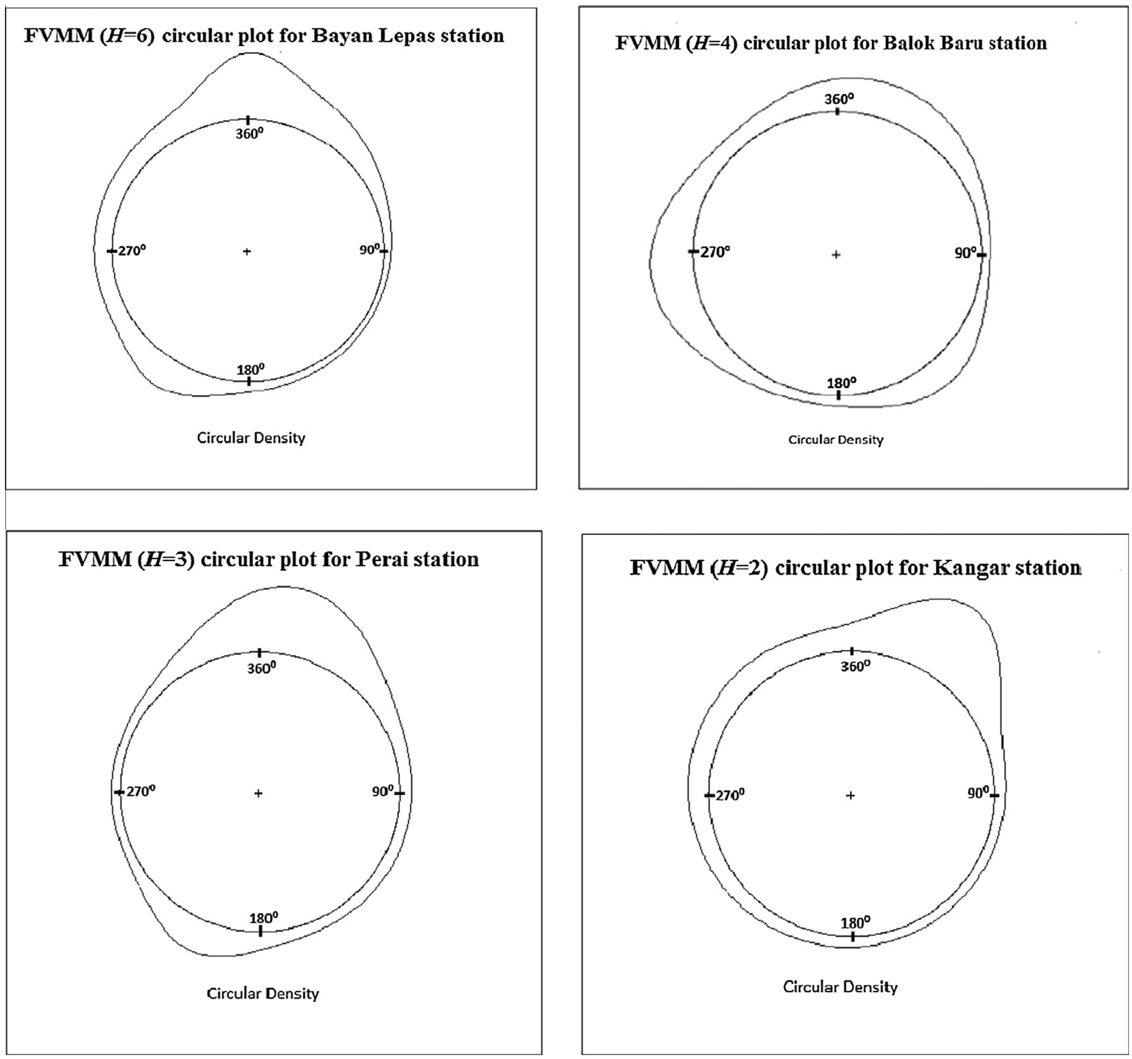

Fig. 6. The FVMM circular plots for Bayan Lepas ($H = 6$), Balok Baru ($H = 4$), Perai ($H = 3$) and Kangar ($H = 2$).

The results for the MAPE analysis are show in Table 6, in which a small value of the MAPE indicates a FVMM that fits the data well. That MAPE analysis leads to a conclusion similar to that of the R^2 analysis for each station. By increasing the number of components of the FVMM model for each station, the value of the MAPE for model fitting decreases. The decreasing value of the MAPE for each particular model indicates a model fit that is more precise, particularly in terms of the modality of the histogram data. Thus, based on the results in Tables 5 and 6, we conclude that the most suitable model for the wind direction for the Bayan Lepas station is FVMM with $H = 6$ components, Kota Bahru and Balok Baru with $H = 4$ components, Perai with $H = 3$ components and for other stations, with $H = 2$ components. Based on the number of suitable mixture components, the fitted von Mises mixture distributions and the circular density plots for several selected stations are show in Figs. 5 and 6. From Figs. 5 and 6, the circular plots for the Bayan Lepas and Perai stations show that most of the wind was blowing from the north and the south–south–west, with the most prominent wind directions within 340–24° and 200¯250°. For the Balok Baru station, the circular plot shows that the most dominant wind directions are the north (340–45°), the east–south–east to south–south–east (120–160°) and the south–south–west to west–north–west (200–300°). Alternatively, for the Kangar station, the circular plot clearly shows that the wind blows most prominently in the north to northeast direction (0–50°). Thus, knowing the most prominent direction of the wind blows for a region contributes significant information to the process of planning or forecasting for wind energy generation, air pollution assessment, climate change, the construction sector, the agricultural sector, maritime activities, meteorology and many more. In wind energy evaluation in particular, information regarding the model and the prominent wind direction will enable wind turbines to be positioned in such a way as to maximize the captured energy.

6. Conclusions

Our study showed that finite von Mises mixture distributions with H numbers of components are adequate for describing the distributions of wind directions for the wind stations studied. This conclusion was also supported by R^2 coefficients MAPE analysis. However, by plotting the model in a circular plot, it is clear that some directions of wind blows are more dominant compared to other directions. Based on our results, we suggest that a more comprehensive analysis involving more stations be conducted in the future to obtain a better understanding of wind directions as well as the characteristics of wind regimes and energy potential in Malaysia. Thus, more efforts must be made to identify suitable locations before erecting wind turbines.

Acknowledgments

The authors are indebted to the staff of the Department of Environment Malaysia and Malaysian Meteorology Department for providing wind direction data. This research would not have been possible without the sponsorship of the Universiti Kebangsaan Malaysia and the Ministry of Higher Education, Malaysia (Grant Number UKM-GUP-2011-213).

References

[1] Razali AM, Ahmad A, Zaharim A, Sopian K. Statistical analysis of wind directional data. In: The 3rd international conference on mathematics and statistics (ICoMS-3), 5–6 August 2008, Indonesia.
[2] Carta JA, Ramirez P, Bueno C. A joint density function of wind speed and direction for wind energy analysis. Energy Convers Manage 2008;49:1309–20.
[3] Carta JA, Bueno C, Ramirez P. Statistical modeling of directional wind speeds using mixture of von Mises distributions: case study. Energy Convers Manage 2008;49:897–907.
[4] Kamisan NAB, Hussin AB, Zubairi YZ. Finding the best circular distribution for southwesterly monsoon wind direction in Malaysia. Sains Malaysiana 2010;39:387–93.
[5] Masseran N, Razali AM, Ibrahim K, Wan Zin WZ. Evaluating the wind speed persistence for several wind stations in Peninsular Malaysia. Energy 2012;37:649–56.
[6] Masseran N, Razali AM, Ibrahim K, Wan Zin WZ, Zaharim A. On spatial analysis of wind energy potential in Malaysia. WSEAS Trans Math 2012;11:467–77.
[7] Masseran N, Razali AM, Ibrahim K. An analysis of wind power density derived from several wind speed density functions: the regional assessment on wind power in Malaysia. Renew Sustain Energy Rev 2012;16:6476–87.
[8] Masseran N, Razali AM, Ibrahim K, Wan Zin WZ, Zaharim A. On spatial estimation of wind energy potential in Malaysia. In: 5th international conference on applied mathematics, simulation, modelling (ASM'11), July 2011, Corfu Island.
[9] Met One Instrument. Wind Energy. <http://www.metone.com/windpower.php>, 010C-020C; 2010.
[10] Acoltzi TM, Martinez AT, Perez JC. The surface wind direction as a function of the wind at 850 mb in Laguna Verde zone, Veracruz, Mexico. Atmosfera 1989;2:181–7.
[11] Azmani M, Reboul S, Choquel JB, Benjelloun M. A recursive, change point estimate of the wind speed and direction. In: IEEE 7th international conference on computational cybernetics, November 26–29, Palma de Mallorca, Spain; 2009.
[12] Shieh GS, Johnson RA. Inferences based on a bivariate distribution with von Mises marginals. Ann Inst Statist Math 2005;57:789–802.
[13] Dobigeon N, Tourneret JY. Joint segmentation of wind speed and direction using a hierarchical model. Comput Statist Data Anal 2007;51:5603–21.
[14] Williams B, Christensen WF, Reese CS. Pollution sources direction identification: embedding dispersion models to solve an inverse problem. Environmetrics 2011;22:962–74.
[15] Gumbel EJ, Greenwood JA, Durand D. The circular normal distribution: theory and tables. J Am Statist Assoc 1953;48:131–52.
[16] Jammalamadaka SR, SenGupta A. A topics in circular statistics. Singapore: World Scientific Publishing; 2001.
[17] Mooney JA, Helms PJ, Jolliffe IT. Fitting mixture of von Mises distributions: a case study involving sudden infant death syndrome. Comput Statist Data Anal 2003;41:505–13.
[18] Banerjee A, Dhillon IS, Ghosh J, Sra S. Clustering on the unit hypershere using von Mises–Fisher distributions. J Mach Learn Res 2005;6:1345–82.
[19] R Development Core Team R. A language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria; 2012, ISBN 3-900051-07-0, URL <http://www.R-project.org>.
[20] Morgan EC, Lackner M, Vogel RM, Baise LG. Probability distributions for offshore wind speeds. Energy Convers Manage 2011;52:15–26.
[21] Carta JA, Ramirez P. Analysis of two-component mixture Weibull statistics for estimation wind speed distributions. Renew Energy 2007;32:518–31.
[22] Brano VL, Orioli A, Ciulla G, Culotta S. Quality of wind speed fitting distribution for urban area of Palermo, Italy. Renew Energy 2011;36:1026–39.