Sains Malaysiana 39(3)(2010): 387–393

Finding the Best Circular Distribution for Southwesterly Monsoon Wind Direction in Malaysia

(Menentukan Taburan Berarah Terbaik bagi Data Arah Angin Barat Daya di Malaysia)

NUR ARINA BASILAH KAMISAN, ABDUL GHAPOR HUSSIN* & YONG ZULINA ZUBAIRI

ABSTRACT

In this paper, four types of circular probability distribution were used to evaluate which circular probability distribution gives the best fitting for southwesterly Malaysian wind direction data, namely circular uniform distribution, von Mises distribution, wrapped-normal distribution and wrapped-Cauchy distribution. The four locations chosen were Alor Setar, Langkawi, Melaka and Senai. Two performance indicators or goodness of fit tests which are mean circular distance and chord length were used to test which distribution give the best fitting.

Keywords: Circular probability distribution; chord length; directional data; goodness of fit test; mean circular distance

ABSTRAK

Dalam kajian ini, empat jenis taburan kebarangkalian berarah telah digunakan untuk melihat taburan yang memberikan keserasian yang terbaik dengan data arah angin di Malaysia. Taburan tersebut adalah taburan berarah seragam, taburan von Mises, taburan normal-menggulung dan taburan Cauchy-menggulung. Empat lokasi di Malaysia telah dipilih untuk kajian ini iaitu di Alor Setar, Langkawi, Melaka dan Senai. Dua jenis penentu prestasi iaitu jarak min membulat dan panjang perentas telah digunakan untuk melihat taburan yang memberikan prestasi terbaik.

Kata kunci: Data berarah; jarak min membulat; panjang perentas; taburan kebarangkalian berarah; ujian keserasian terbaik

INTRODUCTION

Wind direction is one of the features that should be considered in building wind turbine and in structural and environmental design analysis. Wind direction is a type of directional data which can be represented as angles measured from one point chosen as the “zero direction”. The starting point and a rotation from this point whether it is clockwise or anticlockwise are taken as the positive direction. Observation on these two dimensional directions are also called circular or directional data (Jamalamadaka & Sengupta 2001). Directional data have many unique and novel features which are different from the standard linear or real line data set. Such distinctive features make directional statistics analysis substantially different from the linear analysis (Mardia & Jupp 1999). Many researchers proposed circular distribution for directional or circular data set.

There are many areas that deal with directional data such as animal navigation in biology, sound waves in physics and vector cardiology in medical science, to name some of them. In this study, wind direction data obtained from the Malaysian Meteorological Department were modelled to fit some circular distribution. This paper aimed to find the best circular distribution that best fitting the southwesterly monsoon wind direction at four different locations in Malaysia. For this, daily wind directions in the year 2005 have been recorded at four different locations namely Alor Setar, Langkawi, Melaka and Senai which are located at the southwest part of Malaysia. The four circular distributions considered in this study were uniform distribution, wrapped-normal distribution, wrapped-Cauchy distribution and von Mises distribution.

EXPERIMENTAL METHODS

CIRCULAR DISTRIBUTION

A circular distribution is a probability distribution whose total probability is concentrated on the circumference of a unit circle (Jamalamadaka & Sengupta 2001). The circular random variables is measured in degrees or radians and the value is in the range of $[0,2\pi)$ or $[-\pi, \pi)$. This section describes briefly some of the circular distribution that will be used in finding the best distribution that fits the Malaysian wind direction data.

Circular Uniform Distribution Circular uniform distribution is the basic distribution on the circle. It is a unique distribution on the circle which is invariant under rotation and reflection (Mardia & Jupp 1999). If the total probability is spread out uniformly on the circumference then we have Circular Uniformity (CU) distribution with the constant density function of:

$$g(\theta) = \frac{1}{2\pi}, \quad 0 \le \theta \le 2\pi. \tag{1}$$

Von Mises Distribution From the point of view of statistical inference, this is the most useful distribution on the circle. This distribution was introduced by von Mises (1918) in order to study the deviations of measured atomic weights from integral values. This distribution is also known as circular normal distribution. A circular random variable θ is said to have a von Mises distribution if it has density function of:

$$g(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, 0 \le \theta \le 2\pi, \tag{2}$$

where $0 \le \mu < 2\pi$ and $\kappa \ge 0$ are mean direction and concentration parameters, respectively. The function I_0 denotes the modified Bessel function of the first kind and order zero and I_0 is defined as:

$$I_0 = \sum_{r=0}^{\infty} \frac{1}{r!^2} \left(\frac{1}{2}\kappa\right)^{2r}. \tag{3}$$

WRAPPED – CAUCHY DISTRIBUTION

The wrapped-Cauchy distribution is obtained by wrapping the Cauchy distribution on the real line with density of

$$f(x) = \left(\frac{1}{\pi}\right) \frac{\sigma}{\sigma^2 + (x-\mu)^2}, \ -\infty < x < \infty \tag{4}$$

around the circle. It has the probability density function

$$g(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2 - 2\rho\cos(\theta - \mu)}, 0 \le \mu < 2\pi \tag{5}$$

where $\rho = e^{-\sigma}$. The equality of the two expressions above is verified by equating the real parts of the geometric series identity $\sum_{k=1}^{\infty} a^k = \frac{a}{1-a}$ with $a = \rho e^{-i(\theta-\mu)}$. The distribution is unimodal and symmetric.

Wrapped Normal Distribution A wrapped-normal distribution is obtained by wrapping a $N(\mu, \sigma^2)$ (distribution around the circle where $\sigma^2 = -2\log\rho$, i.e $\rho = e^{-\sigma^2/2}$. Its probability density function is given by

$$g(\theta; \mu, \rho) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} \exp\left[\frac{-(\theta - \mu - 2\pi m)^2}{2\sigma^2}\right] \tag{6}$$

In particular, the mean direction is μ (mod 2π) and the mean resultant length is ρ. From the theory of theta function, an alternate and more useful representation of this density can be shown to be

$$g(\theta; \mu, \rho) = \frac{1}{2\pi}\left\{1 + 2\sum_{p=1}^{\infty} \rho^{p^2} \cos p(\theta - \mu)\right\}, 0 \le \rho \le 1. \tag{7}$$

It is clear that the density can be adequately approximated by just the first few terms of the infinite series, depending on the value of σ^2. It is unimodal and symmetric about the value of $\theta = \mu$. The wrapped-normal distribution possesses the additive property which is the convolution of two wrapped-normal variables is also wrapped-normal unlike the von Mises distribution.

PERFORMANCE INDICATORS

In this study, there are two performance indicators to measure the goodness of fit of the models used. Because circular data are difference from linear or real line data, the computation of the performance indicator must also taken into account the behaviour of the circular data. Two types of performance indicators were used in this study which are chord length and mean circular distance.

The Chord Length The chord length can be defined by

$$A = \frac{\sum \sin \frac{\pi - |\pi - |\theta_i - \hat{\theta}_i||}{2}}{n} \tag{8}$$

where θ_i is the observed data and $\hat{\theta}_i$ is the predicted data. Small value of mean chord statistics suggests that the data follows the specified distribution.

THE MEAN CIRCULAR DISTANCE

The mean circular distance can be defined by

$$D = \frac{\sum \frac{1}{2}(1 - \cos(\theta_i - \hat{\theta}_i))}{n} \tag{9}$$

where θ_i is the observed data and $\hat{\theta}_i$ is the predicted data. Values that approximately close to zero, gives information that the data follows the specified distribution. These two numerical methods are then calculated by using Matlab software.

FITTING THE WIND DIRECTION OF SOUTHWESTERLY MONSOON

In this paper, we only considered the southwesterly monsoon because all the four locations are located on the west coast side of Malaysia. Below are circular plots for each of the locations. From this plot, we could see the wind spread during southwesterly monsoon.

Figure 1 shows how the wind direction spread during southwesterly monsoon. Most of the points spread on the west and south section of the circular plot. For southwesterly monsoon, we could see from Figure 1 that Alor Setar and Langkawi showed a spread of wind direction from 180° till 320°. This was due to the location of Alor Setar and Langkawi on the northern part of Malaysia and they faced southwesterly monsoon directly every year from May till September. Most of the wind direction for Melaka spread on 150° and 255° angle or we could say that most of the wind direction blows from the south. Melaka is located on western part of Malaysia, however because Melaka location's is in front of Sumatra Island, the wind most likely will blows from the south during southwesterly monsoon. As for Senai, we could see that the wind spreads from to 285°. This is because Senai is located at the southern part of Malaysia and is near to Sumatra Island and Singapore.

RESULTS AND DISCUSSION

PARAMETERS ESTIMATE

The wind direction data for all four locations were fitted into particular circular distributions and each of the locations were plotted to make comparison between the distributions. Unlike the real line distributions, circular distributions do not require particular parameters for different distribution. The mean angle, μ is given by:

$$\mu = \begin{cases} \tan^{-1}\left(\dfrac{S}{C}\right) & S > 0, C > 0 \\ \tan^{-1}\left(\dfrac{S}{C}\right) + \pi & C < 0 \\ \tan^{-1}\left(\dfrac{S}{C}\right) + 2\pi & S < 0, C > 0 \end{cases} \tag{10}$$

where $C = \sum_{i=1}^{n} \cos\theta_i$ and $S = \sum_{i=1}^{n} \sin\theta_i$.

The mean resultant length of circular distribution, ρ is given by:

$$\rho = (\bar{C}^2 + \bar{S}^2)^{\frac{1}{2}} \tag{11}$$

where $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} \cos\theta_i$ and $\bar{S} = \frac{1}{n}\sum_{i=1}^{n} \sin\theta_i$.

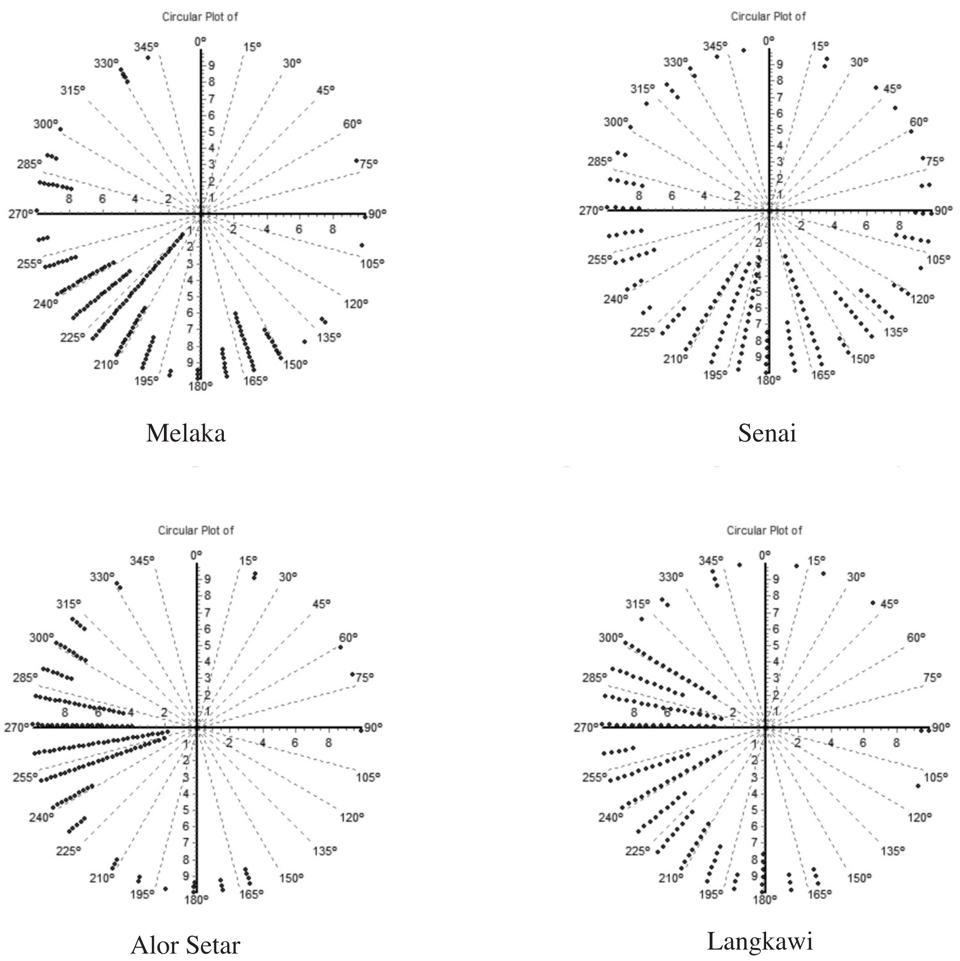

FIGURE 1. Circular plots of southwesterly monsoon for all four locations

The concentration parameter of circular distribution, κ is given by:

$$\kappa = \begin{cases} 2\rho + \rho^3 + \frac{5}{3}\rho^5 & \rho < 0.53 \\ -0.4 + 1.39\rho + \frac{0.43}{1-\rho} & 0.53 \le \rho < 0.85 \\ \left(\rho^3 - 4\rho^2 + 3\rho\right)^{-1} & \rho \ge 0.85 \end{cases} \tag{12}$$

Table 1 shows the parameters estimated for all four circular distributions. By using Oriana software, the parameters were obtained for each location.

GRAPHICAL APPROACHES

For each location, a new data set will be generated by using the parameters obtained above for all four circular distributions. For each location, a comparison were made graphically by using cumulative distribution (cdf) plot between all four circular distributions. Location's cdf plot were the standard cdf plot. The plot which follows the location's cdf plot closely was said to be the best circular distribution that fit with the location. Figure 2, 3, 4 and 5 show four cdf plots of the four locations.

The von Mises plot is the closest plot that follows the location's plot compared to others location which says that all four locations wind direction fit best with the von Mises distribution. Other than cdf plot, we also used quantile-quantile (Q-Q) plot in order to see if our data fits with von Mises distribution. Q-Q plot can be used to compare a data set against a theoretical distribution (Henry 2002). If the plot is on a straight line, then data with the von Mises distribution.

Figure 6 shows the Q-Q plots for the four locations and We could see that most of the plots are on the straight line. Although the Alor Setar, plot is over estimated at the centre of the straight line but basically the plot was located near the straight line and on the straight line. From this Q-Q plots in Figure 6, we could conclude that the wind direction fit well with von Mises distribution.

NUMERICAL APROACHES

To further support the graphical assessment by Q-Q plot, a measure using numerical on the goodness of fit was also used. We proposed two methods which are mean circular distance and chord length. Table 2 shows the results for both performance indicators.

Consistent with the results of the cdf plots, the numerical computation also showed small values of von Mises distribution for all four locations compared to the other three distributions (Table 2). The difference of the

TABLE 1. Result of parameter estimates for southwesterly monsoon

Parameters Locations	N	Mean angle, μ (radian)	Concentration parameter, κ	Mean resultant length, ρ
Alor Setar	152	4.5406	2.853	0.799
Langkawi	152	4.4395	2.001	0.699
Melaka	153	3.7691	2.225	0.732
Senai	151	3.3210	1.184	0.511

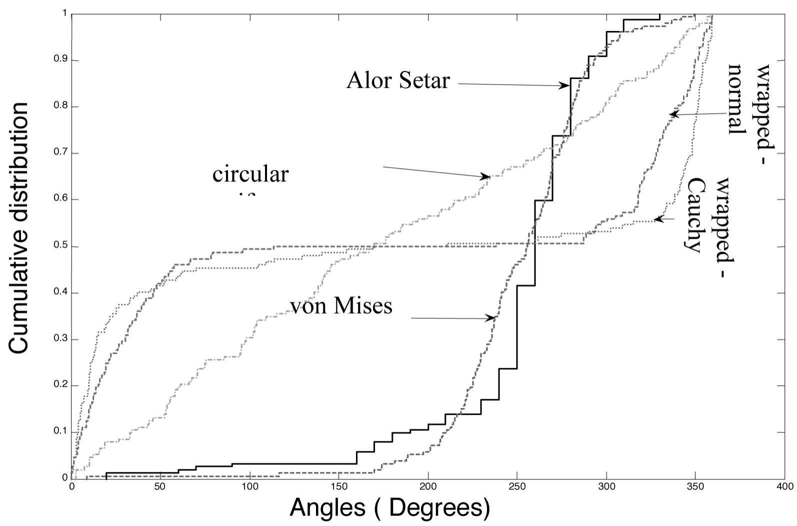

FIGURE 2. The cumulative distribution function plot of Alor Setar

TABLE 2. Performance indicators result of southwesterly monsoon wind direction

Performance Indicator	Chord Length			
Location	Alor Setar	Langkawi	Melaka	Senai
Circular Uniform	0.6465	0.5843	0.6400	0.6427
Wrapped – Cauchy	0.7260	0.6894	0.8150	0.7505
Wrapped – Normal	0.7088	0.6935	0.8268	0.7353
Von Mises	0.3237	0.4334	0.4108	0.5439
Performance Indicator	Mean Circular Distance			
Circular Uniform	0.5095	0.4446	0.5176	0.5160
Wrapped – Cauchy	0.5697	0.5434	0.7185	0.6425
Wrapped – Normal	0.5557	0.5589	0.7326	0.6289
Von Mises	0.1726	0.2647	0.2550	0.3959

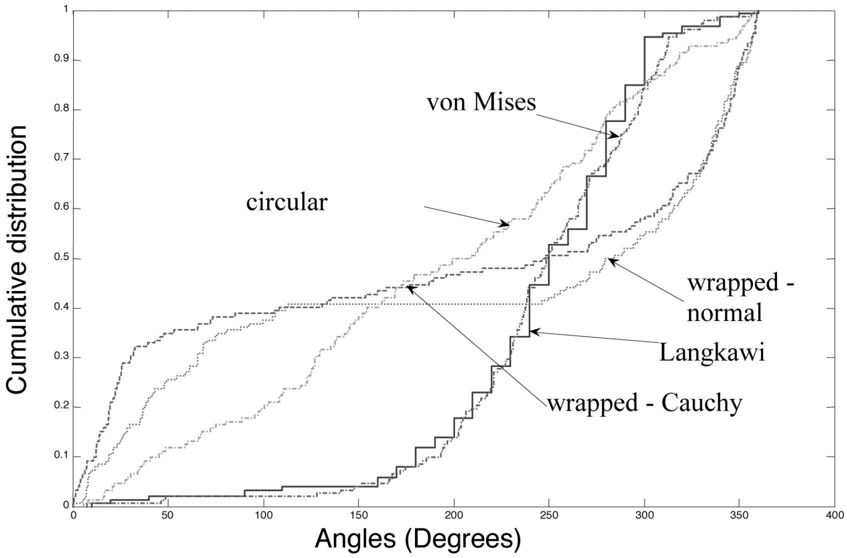

FIGURE 3. The cumulative distribution function plot of Langkawi

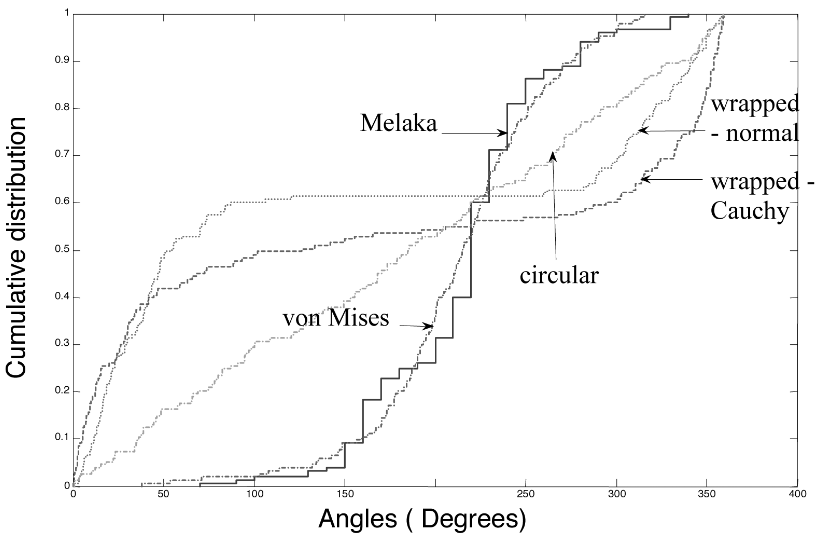

FIGURE 4. The cumulative distribution function plot of Melaka

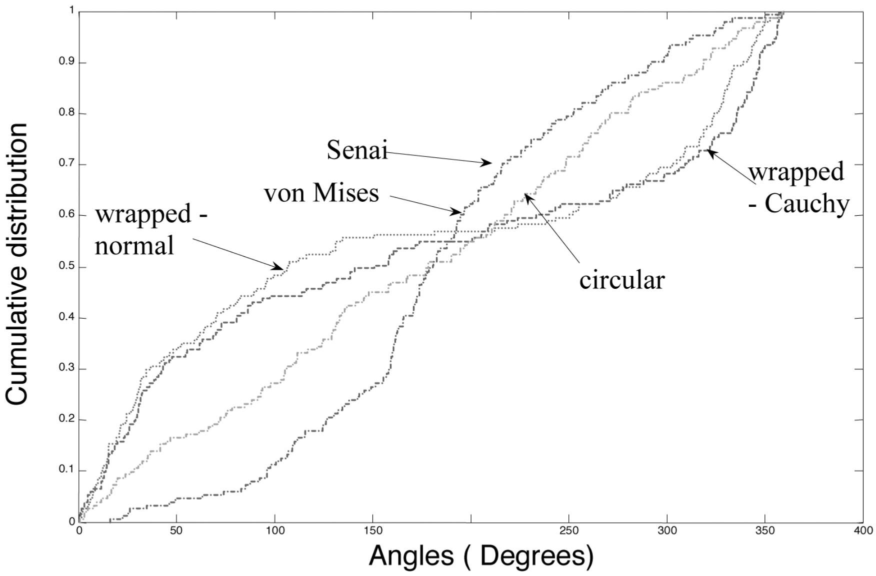

FIGURE 5. The cumulative distribution function plot of Senai

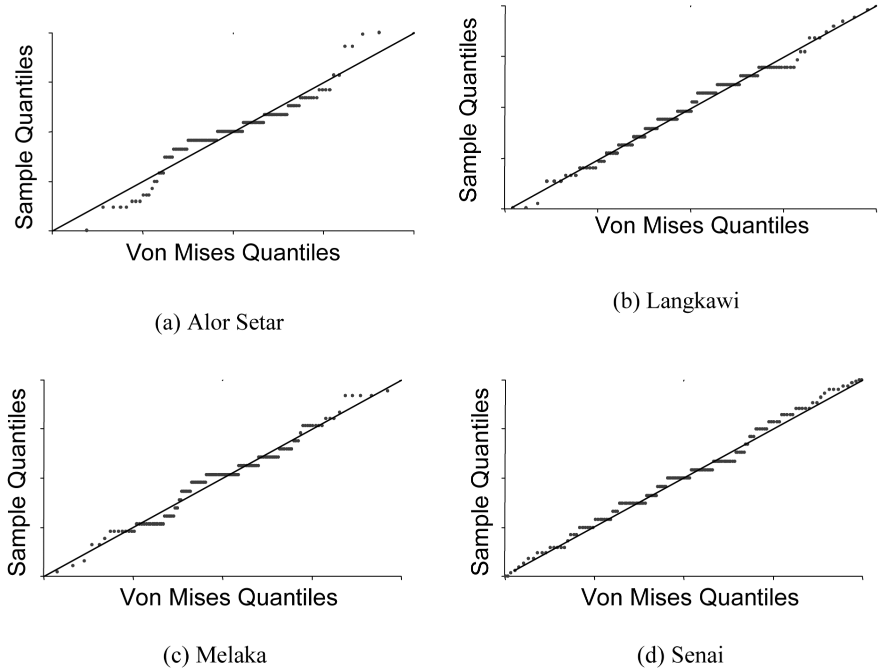

FIGURE 6. Q-Q plot for all four locations

values between the circular distribution for each location was quite large because the value of these two performance indicators should not be larger than 1. Although the difference between the von Mises distribution and other distributions were large, it is of interest to carry-out the statistical test to look at any significant difference between the von Mises distribution and the other distribution as shown in the performance indicators. The test was significantly different at 95% level which suggests that the performance indicator obtained for the von Mises distribution was significantly less compare to others.

Hence, using graphical and numerical approaches we have shown that the southwesterly monsoon wind direction fit best with the von Mises distribution.

CONCLUSION

Our study showed that von Mises distribution is the closest to the location's cdf plot. This result was also supported by the Q-Q plot results in Figure 6. Two performance indicators were proposed in this paper to see the goodness-of-fit and the result in Table 2 shows that von Mises distribution gives the best fit for all four locations and for both performance indicators too. Von Mises distribution gave the smallest value and the differences between the other three distributions were quite obvious for all four locations. To support this performance indicators result, we also conducted a t-test to show that there is a significant difference between the performance indicator for von Mises distribution and the other distributions. Therefore, von Mises distribution was the best circular distribution to describe the southwesterly monsoon wind direction.

REFERENCES

Henry, C.T. 2002. *Testing for Normality*. New York: Marcel Dekker.

Jamalamadaka, S.R. & SenGupta A. 2001. *Topics in Circular Statistics*. Singapore: World Scientific Publishing Co. Pte. Ltd.

Mardia, K.V. & Jupp, P. E. 1999. *Directional Statistics*. England: John Wiley, Chichester.

Centre for Foundation Studies in Science
University of Malaya
50603 Kuala Lumpur
Malaysia

*Corresponding author; email: ghapor@um.edu.my

Received: 12 June 2009
Accepted: 27 August 2009