

Understanding the Behaviour of Wind Direction in Malaysia during Monsoon Seasons using Replicated Functional Relationship in von Mises Distribution (Pemahaman Tingkah Laku Arah Angin di Malaysia ketika Musim Tengkujuh menggunakan Hubungan Fungsian yang Direplikasi dalam Pengedaran von Mises)

NOR HAFIZAH MOSLIM, NURKHAIRANY AMYRA MOKHTAR, YONG ZULINA ZUBAIRI* & ABDUL GHAPOR HUSSIN

ABSTRACT

In studies of potential wind energy, knowing statistical distribution of wind direction provides useful information in making predictions and gives a better understanding of the behavior of the wind direction. Malaysia experiences two monsoon seasons per year; namely Southwest Monsoon and Northeast Monsoon and in this paper, our interest is to investigate whether the direction of wind data in monsoon seasons can be modelled using replicated LFRM with von Mises distribution. The beauty of this model is that it considers the error terms in both x and y variables. This study considers the bivariate relationship of directional wind data where errors are present in both. Here, we propose a replicated functional relationship model, with the von Mises distribution to describe the relationship of the wind direction data. In the parameter estimation, maximum likelihood method is considered with pseudo-replicated group of the replicated form of the functional relationship. The novelty of this approach is that assumption on the ratio of concentration parameters is no longer deemed necessary. Also, we derive the covariance matrix of the parameters based on Fisher Information. From the Monte Carlo simulation study, small bias measures were obtained, suggesting the viability of the model. Based on the simulation study, it can be concluded that the wind direction of the two monsoons in Malaysia can be modelled using replicated linear functional relationship model.

Keywords: Circular data; Monte Carlo simulation; parameter estimation; von Mises distribution; wind direction data

ABSTRAK

Dalam kajian tentang potensi tenaga angin, mengetahui pengedaran statistik arah angin memberikan maklumat yang berguna dalam membuat ramalan dan memberikan pemahaman yang lebih baik mengenai tingkah laku arah angin. Malaysia mengalami dua musim tengkujuh setiap tahun, iaitu Monsun Barat Daya dan Monsun Timur Laut dan dalam makalah ini, minat kami adalah untuk mengkaji apakah arah data angin pada musim tengkujuh dapat dimodelkan menggunakan LFRM yang direplikasi dengan pengedaran von Mises. Keindahan model ini adalah bahawa ia menganggap istilah kesalahan dalam kedua-dua pemboleh ubah x dan y . Kajian ini mempertimbangkan hubungan bivariat data angin arah dan terdapat kesilapan pada kedua-duanya. Di sini, kami mencadangkan model hubungan fungsian yang direplikasi, dengan pengedaran von Mises untuk menggambarkan hubungan data arah angin. Dalam perkiraan parameter, kaedah kemungkinan maksimum dipertimbangkan dengan kumpulan pseudo-replikasi bentuk replikasi hubungan fungsian. Kebaruan pendekatan ini adalah bahawa anggapan mengenai nisbah parameter kepekatan tidak lagi dianggap perlu. Juga, kami memperoleh matriks kovarians parameter berdasarkan Maklumat Fisher. Daripada kajian simulasi Monte Carlo, ukuran bias kecil diperolehi, menunjukkan keberlangsungan model. Berdasarkan kajian simulasi, dapat disimpulkan bahawa arah angin dua monsun di Malaysia dapat dimodelkan dengan menggunakan model hubungan fungsian linear yang direplikasi.

Kata kunci: Anggaran parameter; data arah angin; data berkeliling; pengedaran von Mises; simulasi Monte Carlo

INTRODUCTION

Wind energy is one of the renewable sources of electricity and has been used throughout the history of mankind for many activities such as moving the ships, grinding grain, and pumping water. With the discovery of electricity from

the early 1900s the wind energy was used for electricity generation (Çevik et al. 2019). Understanding wind direction is important in studies of potential energy from wind. In 2011, Kamisan et al. identified the circular model that gives the best fit for maximal wind speed in Malaysia

for the year 2005. Sanusi et al. (2017) fitted the wind speed and wind direction with Weibull and finite mixture of the von Mises (mvM) respectively for Mersing, Malaysia.

The von Mises distribution is said to be the most useful distribution for circular data (Kamisani et al. 2010). The von Mises distribution is analogous to the normal distribution for linear data since it has some similar characteristics with the normal distribution (Caires & Wyatt 2003). The probability distribution function of the von Mises distribution is given by $g(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu))$ where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero, which can be defined by $I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \theta) d\theta$ where μ is the mean direction and κ is the concentration parameter for $0 \leq x < 2\pi$, $0 \leq \mu < \pi$ and $\kappa > 0$.

The weather in Malaysia is characterised by two monsoon regimes, namely, the southwest monsoon from May to September, and the northeast monsoon from November to March (Albani & Ibrahim 2013). In this paper, we investigate the behaviour of wind direction data during northeast monsoon in two stations, namely Kuala Terengganu and Alor Setar over two years period. In doing so, a functional relationship model is considered where we assume that there are unobservable errors. Here, the mean of both errors is assumed zero but have different concentration parameters κ and ν , respectively. It is worthwhile to note that concentration parameter is unestimable unless assumption on the ratio of the error concentration is made (Hussin et al. 2005). If replicates of the data are obtained or when groups of pseudo-replicated based on the unreplicated data are formed, we then have multiples of x and y observations. This model is known as the replicated linear functional relationship model; one example, the replicated linear functional relationship model is applied for linear data where unbiased estimating equations for parameters have been developed using the Student- t distribution for the measurement errors (Galea & de Castro 2017). The motivation of this study is to extend this model to data that is in circular form.

In the parameter estimation using maximum likelihood of the replicated linear functional relationship

model for circular variables by using the von Mises distribution, we proposed a simple algorithm to create the groupings for the pseudo-replicates. In this model, the slope β is set to be equal to one to obtain the desired symmetry of the functional relationship model. We use the method of maximum likelihood estimation for the parameters in the model and derive the covariance matrix of the model based on the Fisher information matrix. To investigate the performance of the model, we carried out the simulation study to obtain the bias measure of the parameter estimates. The proposed method is used to investigate on the annual relationship of wind directions during the two monsoon seasons.

MATERIALS AND METHODS

STUDY AREAS

Malaysia is a country in the Southeast Asia, between latitudes 1° and 7° North, and longitudes 100° and 119° East, that lies in the equatorial zone which its climate is influenced by monsoons. It comprises of two regions, Peninsular Malaysia located at the western part of Malaysia and the Borneo Island located at the eastern part. These regions are separated by 640 km of South China Sea (Chalabi et al. 2006).

Kuala Terengganu is a coastal town located in the state named Terengganu in Peninsular Malaysia with high average wind speed (Albani & Ibrahim 2013; Khatib et al. 2012). It has a tropical rainforest climate under the Köppen climate classification with constant moist throughout the year (Ibrahim et al. 2015).

Another study area is Alor Setar, the state capital of Kedah, where it is located at the north region of Peninsular Malaysia. Department of Irrigation and Drainage Malaysia reported that a total area of 209 km² of Kedah is prone to flood, which may affect about 124,000 people. Due to tropical monsoon climate, Kedah has a long, wet season and known as the 'rice bowl' of Malaysia for its rice yield (Eli et al. 2012). Figure 1 shows the location of Kuala Terengganu and Alor Setar, and Table 1 shows the reading of the latitude, longitude and the elevation of the study areas.



FIGURE 1. The location of Kuala Terengganu and Alor Setar in Malaysia

(Source: https://simple.wikipedia.org/wiki/Template:Location_map_Malaysia#/media/File:Malaysia_relief_location_map.jpg)

TABLE 1. The reading of the latitude, longitude and the elevation of the study areas

Stations	Latitude	Longitude	Elevation
Kuala Terengganu	5° 23' N	103° 06' E	5.2 m
Alor Setar	6° 12' N	100° 24' E	3.9 m

THE REPLICATED LINEAR FUNCTIONAL RELATIONSHIP MODEL

The un-replicated linear functional relationship model (LFRM) proposed by Caires and Wyatt (2003) is $Y = X + \alpha \pmod{2\pi}$ with $x_i = X_i + \delta_i$ and $y_i = Y_i + \varepsilon_i$ where $i = 1, 2, \dots, n$ for some rotation parameter α . The random errors δ_i and ε_i are assumed to be independently distributed with von Mises distribution with $\delta_i \sim VM(0, \kappa)$ and $\varepsilon_i \sim VM(0, \nu)$, respectively.

In the model as proposed by Caires and Wyatt (2003), it is worthwhile to note that one needs the ratio of concentration parameter $\lambda = \frac{\nu}{\kappa}$ to estimate the concentration parameters of δ_i and ε_i . The shortcoming of this model is that an assumption has to be made on the ratio value. In this paper, we extend the un-replicated LFRM of Caires and Wyatt to become a replicated LFRM. By doing so, the assumption on the ratio of the concentration parameters is no longer deemed necessary when estimating the concentration parameters of δ_{ij} and ε_{ik} , respectively.

This new replicated LFRM which considers replicated observation has practical implications as replicate reduces the chance an observed effect was a fluke (Cumming 2006). We consider that there are replicated observations of X_i and Y_i occurring in p sub-groups. All of the data of x and y variables are divided into p -subgroups with m elements such that $p \times m = n$, where p is the maximum divisor of n and $p \leq m$. Then, the data are arranged in the form of $p \times m$ matrix. Measurements x_{ij} ($j = 1, \dots, m_i$) are made on X_i and measurements y_{ik} ($k = 1, \dots, m_i$) are made on Y_i , where $0 \leq x_{ij}, y_{ik} < 2\pi$. The model is given by $Y = \alpha + X \pmod{2\pi}$, where α is the rotation parameter and the variables considered for this model are $x_{ij} = X_i + \delta_{ij}$ and $y_{ik} = Y_i + \varepsilon_{ik}$.

The error terms of $\delta_i \sim VM(0, \kappa)$ and $\varepsilon_i \sim VM(0, \nu)$, are considered for the variables x and y , respectively. In this case, the error terms δ_{ij} and ε_{ik} are independently distributed with the von Mises distribution. The concentration parameters for the error terms δ_{ij} and ε_{ik} are κ and ν , respectively. Concentration parameter influences

the von Mises distribution inversely as variance affects the normal distribution (Caires & Wyatt 2003).

PARAMETER ESTIMATION OF THE REPLICATED LINEAR FUNCTIONAL RELATIONSHIP MODEL FOR CIRCULAR VARIABLES

In this section, the parameters of replicated LFRM, X_i , the rotation parameter α , the concentration parameters κ and ν , are estimated by using the method of maximum likelihood.

The log likelihood equation of the von Mises distribution is given by

$$\log L(\alpha, \kappa, \nu, X; x, y) = -NM \log 2\pi - N \log I_0(\kappa) - M \log I_0(\nu) + \kappa \sum_{i=1}^p \sum_{j=1}^n \cos(x_{ij} - X_i) + \nu \sum_{i=1}^p \sum_{k=1}^m \cos(y_{ik} - \alpha - X_i) \quad (1)$$

where N is the total sample size of variable x and M is the total sample size of variable y . We may assume that the total sample size of x and y variables are equal and thus, $N = M$.

To estimate X_i we find the first derivative of X_i with respect to $\log L$ and set it equals to 0.

$$\frac{\partial \log L}{\partial X_i} = \kappa \sum_{j=1}^m \sin(x_{ij} - X_i) + \nu \sum_{k=1}^m \sin(y_{ik} - \alpha - X_i) = 0 \quad (2)$$

X_i may be solved iteratively by some 'initial guess'. Suppose \hat{X}_{i0} is an initial estimate of \hat{X}_i . Then,

$$x_{ij} - \hat{X}_i = x_{ij} - \hat{X}_{i0} + \hat{X}_{i0} - \hat{X}_i = (x_{ij} - \hat{X}_{i0}) + \Delta_i \quad \text{where } \Delta_i = \hat{X}_{i0} - \hat{X}_i. \quad (3)$$

We may also have

$$y_{ik} - \hat{\alpha} - \hat{X}_i = (y_{ik} - \hat{\alpha} - \hat{X}_{i0}) + \Delta_i \quad (4)$$

Thus, the partial derivative equation above becomes

$$\sin(x_{ij} - \hat{X}_{i0} + \Delta_i) + \sin(y_{ik} - \hat{\alpha} - \hat{X}_{i0} + \Delta_i) = 0 \quad (5)$$

For small Δ , then $\cos \Delta_i \approx 1$ and $\sin \Delta_i \approx \Delta_i$. Hence the equation is simplified (approximately) to become

$$\hat{X}_{i1} \approx \hat{X}_{10} + \frac{\sum_{j=1}^m \sin(x_{ij} - \hat{X}_{i0}) + \frac{\nu}{\hat{\kappa}} \sum_{k=1}^m \sin(y_{ik} - \hat{\alpha} - \hat{X}_{i0})}{\sum_{j=1}^m \cos(x_{ij} - \hat{X}_{i0}) + \frac{\hat{\nu}}{\hat{\kappa}} \sum_{k=1}^m \cos(y_{ik} - \hat{\alpha} - \hat{X}_{i0})} \quad (6)$$

To estimate α , we find the first derivative of α with respect to $\log L$ and set it equals to zero.

$$\frac{\partial \log L}{\partial \alpha} = \nu \sum_{i=1}^p \sum_{k=1}^m \sin(y_{ik} - \alpha - X_i) = 0 \quad (7)$$

$$\frac{\cos \hat{\alpha}}{\sin \hat{\alpha}} = \frac{\sum_{i=1}^p \sum_{k=1}^m \cos(y_{ik} - \hat{X}_i)}{\sum_{i=1}^p \sum_{k=1}^m \sin(y_{ik} - \hat{X}_i)} \quad (8)$$

$$\hat{\alpha} = \tan^{-1} \left\{ \frac{\sum_{i=1}^p \sum_{k=1}^m \sin(y_{ik} - \hat{X}_i)}{\sum_{i=1}^p \sum_{k=1}^m \cos(y_{ik} - \hat{X}_i)} \right\} \quad (9)$$

Therefore, we obtain

$$\hat{\alpha} = \begin{cases} \tan^{-1} \left\{ \frac{S}{C} \right\} & \text{when } S > 0, C > 0 \\ \tan^{-1} \left\{ \frac{S}{C} \right\} + \pi & \text{when } C < 0 \\ \tan^{-1} \left\{ \frac{S}{C} \right\} + 2\pi & \text{when } S < 0, C > 0 \end{cases} \quad (10)$$

where $S = \sum_{i=1}^n \sin(y_{ik} - \hat{X}_i)$ and $C = \sum_{i=1}^n \cos(y_{ik} - \hat{X}_i)$.

To estimate k , we find the first derivative of k with respect to $\log L$ and set it equals to zero.

$$\frac{\partial \log L}{\partial \kappa} = -N \frac{I_0^1(\kappa)}{I_0(\kappa)} + \sum_{i=1}^p \sum_{j=1}^m \cos(x_{ij} - X_i) = 0. \quad (11)$$

$$\text{Let } A(k) = \frac{I_0^1(\kappa)}{I_0(\kappa)} = \frac{I_1(\kappa)}{I_0(\kappa)} \quad (12)$$

where $I_0(\kappa)$ and $I_1(\kappa)$ are the asymptotic power series for the Bessel functions.

$$\text{Thus, } A(k) = \frac{1}{N} \left\{ \sum_{i=1}^p \sum_{j=1}^m \cos(x_{ij} - X_i) \right\}. \quad (13)$$

$$\text{Therefore, } \hat{\kappa} = A^{-1} \left(\frac{1}{N} \left\{ \sum_{i=1}^p \sum_{j=1}^m \cos(x_{ij} - \hat{X}_i) \right\} \right) \quad (14)$$

where A^{-1} is the inverse function of the ratio of the first and zeroth order Bessel functions of the first kind. This

function is used to compute the maximum likelihood estimate of the concentration parameter of a von Mises distribution.

To estimate ν , we find the first derivative of ν with respect to $\log L$ and set it equals to zero.

$$\frac{\partial \log L}{\partial \nu} = -M \frac{I_0^1(\nu)}{I_0(\nu)} + \sum_{i=1}^p \sum_{k=1}^m \cos(y_{ik} - \alpha - X_i) = 0 \quad (15)$$

$$\text{Let } A(\nu) = \frac{I_0^1(\nu)}{I_0(\nu)} = \frac{I_1(\nu)}{I_0(\nu)} \quad (16)$$

$$\text{Thus, } A(\nu) = \frac{1}{M} \left\{ \sum_{i=1}^p \sum_{k=1}^m \cos(y_{ik} - \alpha - X_i) \right\} \quad (17)$$

$$\text{Therefore, } \hat{\nu} = A^{-1} \left(\frac{1}{M} \left\{ \sum_{i=1}^p \sum_{k=1}^m \cos(y_{ik} - \alpha - X_i) \right\} \right) \quad (18)$$

THE COVARIANCE MATRIX OF THE REPLICATED LFRM

Using the first and second derivative of the parameters, we derived the covariance matrix of the model based on Fisher Information matrix $F = \begin{bmatrix} B & 0 & E \\ 0 & C & O \\ E^T & 0 & D \end{bmatrix}$

$$\text{where } B \text{ is a } p \times p \text{ matrix; } B = \begin{bmatrix} \hat{\kappa}NA(\hat{\kappa}) + \hat{\nu}MA(\hat{\nu}) & \dots & 0 \\ p & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\hat{\kappa}NA(\hat{\kappa}) + \hat{\nu}MA(\hat{\nu})}{p} \end{bmatrix} \quad (19)$$

$$C \text{ is a } 2 \times 2 \text{ matrix; } C = \begin{bmatrix} NA'(\hat{\kappa}) & 0 \\ 0 & MA'(\hat{\nu}) \end{bmatrix} \quad (20)$$

$$D \text{ is an } 1 \times 1 \text{ matrix; } D = [\hat{\nu}MA'(\hat{\nu})] \quad (21)$$

$$E \text{ is a } p \times 1 \text{ matrix; } E = \begin{bmatrix} \hat{\nu}MA'(\hat{\nu}) \\ p \\ \vdots \\ \hat{\nu}MA'(\hat{\nu}) \\ p \end{bmatrix} \quad (22)$$

The variance of the parameters may be solved by

$$\text{covariance} \begin{bmatrix} \hat{\kappa} \\ \hat{\nu} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} C^{-1} & 0 \\ 0 & (D - E^T B^{-1} E)^{-1} \end{bmatrix} \quad (23)$$

$$\text{where } C^{-1} = \begin{bmatrix} [NA'(\hat{\kappa})]^{-1} & 0 \\ 0 & [MA'(\hat{\nu})]^{-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{NA'(\hat{\kappa})} & 0 \\ 0 & \frac{1}{MA'(\hat{\nu})} \end{bmatrix} \quad (24)$$

$$(D - E^T B^{-1} E)^{-1} = \frac{p[\hat{\kappa}NA'(\hat{\kappa}) + \hat{\nu}MA'(\hat{\nu})]}{\hat{\nu}MA'(\hat{\nu})p[\hat{\kappa}NA'(\hat{\kappa}) + \hat{\nu}MA'(\hat{\nu})] - p[\hat{\nu}MA'(\hat{\nu})]^2} \quad (25)$$

Therefore, the covariance matrix of the parameters becomes

$$\text{cov} \begin{bmatrix} \hat{\kappa} \\ \hat{\nu} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \frac{1}{NA'(\hat{\kappa})} & 0 & 0 \\ 0 & \frac{1}{MA'(\hat{\nu})} & 0 \\ 0 & 0 & \frac{p[\hat{\kappa}NA'(\hat{\kappa}) + \hat{\nu}MA'(\hat{\nu})]}{\hat{\nu}MA'(\hat{\nu})p[\hat{\kappa}NA'(\hat{\kappa}) + \hat{\nu}MA'(\hat{\nu})] - p[\hat{\nu}MA'(\hat{\nu})]^2} \end{bmatrix} \quad (26)$$

SIMULATION STUDY

To assess the accuracy and the biasness of the parameters of the proposed model, a Monte Carlo simulation study is carried out. The number of simulations is set to be s , meanwhile the values of n , k and ν for the error terms have been generated. In this model, the value of α is circular meanwhile k and ν values are continuous. The number of simulations, s is set to be 5000 for each simulation. The values of X have been generated from the von Mises distribution of $VM(2, 3)$ and the true value of $\alpha = \frac{\pi}{4} = 0.7854$. Without loss of generality, the values of the concentration parameters of the error terms, k and ν are set to be 3, 5, 10 and 15, respectively. For each value of k and ν , the sample size $n = 50, 78, 132, \text{ and } 150$ are considered for the simulation. For this simulation, assume that $k = \nu$.

As mentioned earlier, we simulate the pseudo-replicates of the data. All the data of x and y variables are divided into p -subgroups with m elements such that $p \times m = n$, where p is the maximum divisor of n and $p \leq m$. The data are then arranged in the form of $p \times m$ matrix. Next, the parameter estimation is done for based on the proposed method in the previous section. The steps are repeated for 5000 simulations where the biasness of the proposed parameter estimation for 5000 simulations is measured.

In the pseudo-replicate, for $n = 50$, the subgroup is set at 5 with 10 elements in each subgroup. For $n = 78$, the subgroup is set to be 6 with 13 elements in each group. For $n = 132$, there are 11 subgroups with 12 elements in each group and for $n = 150$, the subgroup is set to be 10 with 15 elements in each sub-groups.

Bias measures are used to investigate the performance of the proposed model. For the bias measures of $\hat{\alpha}$, the mean, circular distance and mean resultant length are used. The mean of circular parameter $\hat{\alpha}$, $\tilde{\alpha}$ is given by $\tilde{\alpha} =$

$$\begin{cases} \tan^{-1} \left(\frac{S}{C} \right) & \text{when } S > 0, C > 0 \\ \tan^{-1} \left(\frac{S}{C} \right) + \pi & \text{when } C < 0 \\ \tan^{-1} \left(\frac{S}{C} \right) + 2\pi & \text{when } S < 0, C > 0 \end{cases}$$

where $C = \sum_{j=1}^s \cos(\hat{\alpha}_j)$ and $S = \sum_{j=1}^s \sin(\hat{\alpha}_j)$. It is analogous to the mean of a sample of linear data (Lee 2010).

The circular distance, d is given by $d = \pi - |\pi[\hat{\alpha} - \alpha]|$. Circular distance between two points is to take the smaller of two arc lengths between the points along the circumference (Jammalamadaka & Sengupta 2001). Meanwhile, the mean resultant length of the parameter estimate of $\hat{\alpha}$ is given by $R =$

$\frac{1}{s} \sqrt{(\sum_{j=1}^s \cos(\hat{\alpha}_j))^2 + (\sum_{j=1}^s \sin(\hat{\alpha}_j))^2}$. The mean resultant is used to measure the concentration of unimodal circular distributions and is used to test if there is a strong preferred direction in measured data (Kutil 2012). If the directions are tightly clustered, then, mean resultant length is close to 1. However, if the directions are widely dispersed, then mean resultant length will be close to 0 (Mardia & Jupp 2000).

To investigate the biasness of $\hat{\kappa}$, the mean of $\hat{\kappa}$, estimated bias and estimated root mean square errors are used. Mean of $\hat{\kappa}$ is given by $\bar{\kappa} = \frac{1}{s} \sum_{j=1}^s \hat{\kappa}_j$. The estimated bias is given by $\bar{\kappa} - k$ and the estimated root mean square error is given by $ERMSE = \sqrt{\frac{1}{s} \sum_{j=1}^s (\hat{\kappa}_j - k)^2}$. $ERMSE$ is used as a standard statistical metric to measure model performance. It measures the difference between the actual values observed from the environment that is being modelled and the values predicted by the proposed model (Chai & Draxler 2014).

For the bias measure of $\hat{\nu}$, the mean, estimated bias and estimated root mean square errors are used. The mean of $\hat{\nu}$ is given by $\bar{\nu} = \frac{1}{s} \sum_{j=1}^s \hat{\nu}_j$. The estimated bias of $\hat{\nu}$ is given by $\bar{\nu} - \nu$ and the estimated root mean square errors is given by $ERMSE = \sqrt{\frac{1}{s} \sum_{j=1}^s (\hat{\nu}_j - \nu)^2}$.

RESULTS AND DISCUSSION

Simulation study was carried out to assess the biasness of the parameter estimates of this model proposed in the previous section. Table 2 shows that the circular mean of the parameter estimate $\hat{\alpha}$ in the simulation study are near to the real value of $\hat{\alpha}$ which is 0.7854. The circular distance values are small and near to zero. These suggest that the biasness is very small. Additionally, the mean resultant length values are very near to 1 thus, indicating that the estimation has a good accuracy.

TABLE 2. Biasness of $\hat{\alpha}$

	n	Circular mean	Circular distance	Mean resultant length
$k = 3$ $v = 3$	50	0.7857	0.0003	0.9917
	78	0.7853	0.0001	0.9947
	132	0.7867	0.0013	0.9968
	150	0.7840	0.0014	0.9973
$k = 5$ $v = 5$	50	0.7839	0.0015	0.9955
	78	0.7856	0.0002	0.9971
	132	0.7845	0.0009	0.9983
	150	0.7843	0.0011	0.9984
$k = 10$ $v = 10$	50	0.7857	0.0003	0.9979
	78	0.7855	0.0001	0.9987
	132	0.7857	0.0003	0.9992
	150	0.7858	0.0004	0.9993
$k = 15$ $v = 15$	50	0.7859	0.0005	0.9986
	78	0.7853	0.0001	0.9991
	132	0.7854	0.0000	0.9995
	150	0.7854	0.0000	0.9995

Tables 3 and 4 show that as the value of n increases, the mean of $\hat{\kappa}$ and $\hat{\nu}$ gets closer to real values of k and v , respectively. Looking at the values of estimate bias of $\hat{\kappa}$

and $\hat{\nu}$, they decrease as n increases. The values of ERMSE for both $\hat{\kappa}$ and $\hat{\nu}$ decrease as the value of n increases. These suggest that the parameter estimates are adequate and have small bias.

TABLE 3. Biasness of $\hat{\kappa}$

	n	Mean	Estimate bias	ERMSE
$k = 3$ $v = 3$	50	3.2296	0.2296	0.6871
	78	3.1529	0.1529	0.5156
	132	3.1081	0.1081	0.3764
	150	3.0761	0.0761	0.3427
$k = 5$ $v = 5$	50	5.4938	0.4938	1.2843
	78	5.3093	0.3093	0.9199
	132	5.2703	0.2703	0.7017
	150	5.2258	0.2258	0.6521
$k = 10$ $v = 10$	50	11.1062	1.1062	2.7114
	78	10.7507	0.7507	1.9711
	132	10.6146	0.6146	1.4949
	150	10.4869	0.4869	1.3244
$k = 15$ $v = 15$	50	16.7188	1.7188	4.0835
	78	16.1580	1.1580	3.0059
	132	15.9730	0.9730	2.3130
	150	15.7758	0.7758	2.0390

TABLE 4. Biasness of $\hat{\nu}$

	n	Mean	Estimate bias	ERMSE
$k = 3$ $\nu = 3$	50	3.2255	0.2255	0.6888
	78	3.1307	0.1307	0.5054
	132	3.1033	0.1033	0.3752
	150	3.0790	0.0790	0.3392
$k = 5$ $\nu = 5$	50	5.4991	0.4991	1.2580
	78	5.3105	0.3105	0.9455
	132	5.2758	0.2758	0.7171
	150	5.2304	0.2304	0.6494
$k = 10$ $\nu = 10$	50	11.1437	1.1437	2.6946
	78	10.7690	0.7690	1.9327
	132	10.6334	0.6334	1.4890
	150	10.5040	0.5040	1.3525
$k = 15$ $\nu = 15$	50	16.6189	1.6189	3.9705
	78	16.0663	1.0663	2.8838
	132	15.9451	0.9451	2.2668
	150	15.7851	0.7851	2.0723

From the simulation results, we may say that the proposed model is sufficient to model circular data with very small bias. To highlight, the novelty of this proposed replicated model is we are able to estimate the parameters without having to assume the ratio of the concentration parameter, in contrast to the previous study of un-replicated functional model by Caires and Wyatt (2003).

MODELLING WIND DIRECTION DATA IN MALAYSIA COASTAL STATIONS

Before investigating the bivariate relationship of the wind direction over two consecutive years, we obtained the

summary statistics and distribution of the univariate data. For the Kuala Terengganu coastal station, we consider the wind direction during northeast monsoon season from January to February for two different years, namely 2016 and 2017.

Figures 2 and 3 show the rose diagrams of wind direction data in Kuala Terengganu in 2016 and 2017, respectively. The rose diagram is a histogram presented in a circle, almost identical to the pie chart for linear data. However, each sector represents the frequency or number of observations which falls in the range of angles (Hassan et al. 2009).

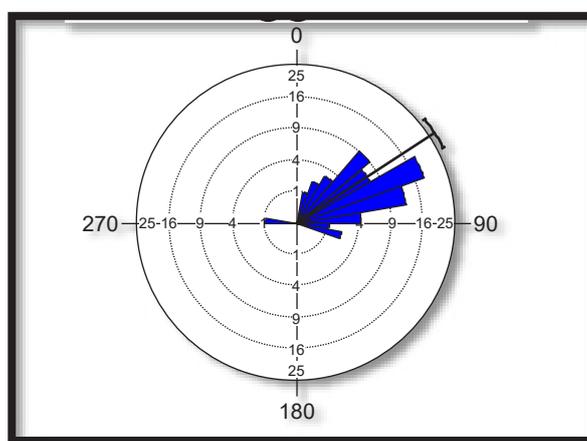


FIGURE 2. Rose diagram of wind direction data of Kuala Terengganu in northeast monsoon 2016

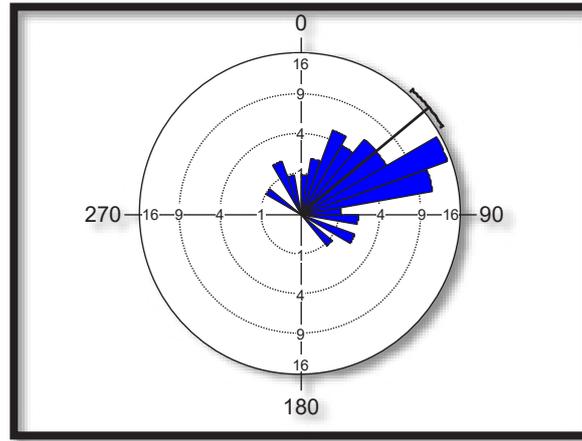


FIGURE 3. Rose diagram of wind direction data of Kuala Terengganu in northeast monsoon 2017

Table 5 shows the summary statistics for wind direction data of Kuala Terengganu in 2016 and 2017. The wind direction has bigger error concentration parameter as

compared to the 2017. The wind direction data for 2016 and 2017 both have the median of 60° or in radian is 1.0472 rad with mean resultant length are 0.922 and 0.853, respectively.

TABLE 5. Summary statistics of wind direction data of Kuala Terengganu

Year	2016	2017
Data type	Angles	Angles
Mean vector	56.779°	50.066°
Length of mean vector	0.922	0.853
Median	60°	60°
Concentration	6.691	3.722

As mentioned earlier, our aim is to investigate the relationship of the wind direction. In this case, we let x be the wind direction of Kuala Terengganu in 2016 and y be the wind direction in 2017. The estimate of the rotation parameter, $\hat{\alpha}$ is 6.1618 radian which is close to zero. This means that the relationship of the wind direction of Kuala Terengganu in 2016 and 2017 are almost the same. Meanwhile, the value of estimation of the error

concentration of wind direction data of Kuala Terengganu in 2016 and 2017 are quite high, which are 6.0341 and 4.4766, respectively. High concentration indicates that the random error is concentrated and less dispersed. This is not surprising as the univariate analysis suggests that the concentration is higher in 2016 as compared to 2017. Small variance of parameter estimates indicate that the parameter estimates are very close to their means.

The concentration parameter is important as the values influence the mathematical formulation of the model. By applying the proposed model with the von Mises

distribution, the relationship between the variables x and y can be described by $\hat{Y} = 6.1618 + \hat{X} \pmod{2\pi}$ where the values of the parameter estimates, and their variance are given in Table 6.

TABLE 6. The parameter estimates of the replicated LFRM for wind direction data in Kuala Terengganu in 2016 and 2017 during northeast monsoon

$\hat{\alpha}$ (variance)	$\hat{\kappa}$ (variance)	$\hat{\nu}$ (variance)
6.1618 (0.00726)	6.0341 (1.0857)	4.4766 (0.7470)

We also investigate if the bivariate relationship can fit well during the other monsoon namely the southwest monsoon season, which is from August to September. Here we consider the Alor Setar coastal station where the relationship of the wind direction for two consecutive years are investigated. In this case, variable x is the wind

direction data in 2016 and the variable y is the wind direction data in 2017. Similar to the earlier analysis, we plot the distribution using rose diagrams and obtained the summary statistics of the wind direction for each year. Figures 4 and 5 show the rose diagrams of wind direction data of Alor Setar in 2016 and 2017, respectively.

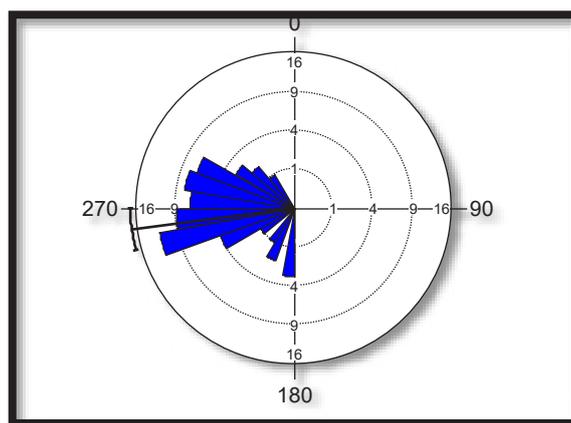


FIGURE 4. Rose diagram of wind direction data of Alor Setar in southwest monsoon 2016

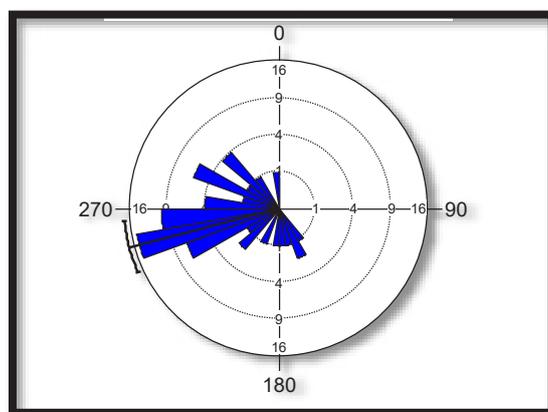


FIGURE 5. Rose diagram of wind direction data of Alor Setar in southwest monsoon 2017

TABLE 7. Summary statistics of wind direction data of Alor Setar

Year	2016	2017
Data type	Angles	Angles
Mean vector	262.606°	255.429°
Length of mean vector	0.87	0.785
Median	260°	250°
Concentration	4.159	2.696

From Table 7, it can be seen that the mean vector and the length of mean vector are almost similar at both years. However, 2016 has a higher error concentration as compared to 2017. Then, we fit the data into the proposed replicated linear functional relationship model. The values of the parameter estimation and their variance are given in Table 8. The estimate of the rotation parameter, $\hat{\alpha}$ is 6.15384 radian which is close to zero too. This means that the relationship of the wind direction of Alor Setar

in 2016 and 2017 are almost the similar. The value of estimation of the error concentration of x and y are 4.68439 and 2.69337, respectively. It shows that the random error of the wind direction data of Alor Setar in 2016 is more concentrated to be compared to the year 2017 and is more consistent with the earlier univariate analysis. In order to investigate the relationship between the two years, we found that the data fit replicated linear functional model and the model is given by $\hat{Y} = 6.15384 + \hat{X} \pmod{2\pi}$.

TABLE 8. The parameter estimates of the replicated LFRM for wind direction data in Alor Star in 2016 and 2017 during southwest monsoon

$\hat{\alpha}$ (variance)	$\hat{\kappa}$ (variance)	$\hat{\nu}$ (variance)
6.15384 (0.01191)	4.68439 (0.62100)	2.69337 (0.29868)

CONCLUSION

To conclude, in this paper, we have proposed a replicated functional relationship model to investigate the relationship of wind direction two coastal stations in Malaysia for two different monsoon seasons, in Kuala Terengganu station during northeast monsoon and in Alor Setar during southwest monsoon. With the von Mises distribution, we derived the maximum likelihood estimation of the model with pseudo-replicated group of the replicated form of the functional relationship. The novelty of this approach is that assumption on the ratio

of concentration parameters is no longer necessary. Also, we derive the covariance matrix of the parameters based on Fisher Information. From the Monte Carlo simulation study, small bias measures were obtained, suggesting the viability of the model. The empirical evidence is used in the evaluation of the proposed model.

ACKNOWLEDGEMENTS

We would like to thank Universiti Teknologi MARA, National Defence University of Malaysia and Universiti Malaya (Grant GPF006H-2018) for supporting this work,

and Dr Siti Fatimah Hassan from Universiti Malaya for sharing us her wind direction data.

REFERENCES

- Albani, A. & Ibrahim, M.Z. 2013. Preliminary development of prototype of Savonius wind turbine for application in low wind speed in Kuala Terengganu, Malaysia. *International Journal of Scientific & Technology Research* 2(3): 102-108.
- Caires, S. & Wyatt, L.R. 2003. A linear functional relationship model for circular data with an application to the assessment of ocean wave measurement. *Journal of Agricultural, Biological and Environmental Statistics* 8(2): 153-169.
- Çevik, H.H., Çunkaş, M. & Polat, K. 2019. A new multistage short-term wind power forecast model using decomposition and artificial intelligence methods. *Physica A: Statistical Mechanics and its Applications* 534: 122177.
- Chai, T. & Draxler, R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific Model Development Discussions* 7(1): 1525-1534.
- Chalabi, A., Lokman, H.M., Suffian, I.M., Karamali, K., Karthigeyan, V. & Masita, M. 2006. Monitoring shoreline change using Ikonos image and aerial photographs: A case study of Kuala Terengganu area, Malaysia. In *ISPRS Commission VII Mid-Term Symposium "Remotes Sensing: from Pixels to Processes"*. Enschede, the Netherlands. pp. 8-11.
- Cumming, G. 2006. Understanding replication: confidence intervals, p-values and what's likely to happen next time. In *International Conference on Teaching Statistics (ICOTS-7)*.
- Eli, A., Shaffie, M. & Zin, W.Z.W. 2012. Preliminary study on bayesian extreme rainfall analysis: A case study of Alor Setar, Kedah, Malaysia. *Sains Malaysiana* 41(11): 1403-1410.
- Galea, M. & de Castro, M. 2017. Robust inference in a linear functional model with replications using the t distribution. *Journal of Multivariate Analysis* 160: 134-145.
- Hassan, S.F., Hussin, A.G. & Zubairi, Y.Z. 2009. Analysis of Malaysian wind direction data using ORIANA. *Modern Applied Science* 3(3): 115-119.
- Hussin, A.G., Fieller, N. & Stillman, E. 2005. Pseudo-replicates in the linear circular functional relationship model. *Journal of Applied Sciences* 5(1): 138-143.
- Ibrahim, A., Juahir, H., Toriman, M.E., Mustapha, A., Azid, A. & Isiyaka, H.A. 2015. Assessment of surface water quality using multivariate statistical techniques in the Terengganu River Basin. *Malaysian Journal of Analytical Sciences* 19(2): 338-348.
- Jammalamadaka, S.R. & Sengupta, A. 2001. *Topics in Circular Statistics*. Singapore: World Scientific Publishing.
- Kamisan, N.A.B., Hussin, A.G., Zubairi, Y.Z. & Hassan, S.F. 2011. Distribution of wind direction recorded at maximum wind speed: A case study of Malaysian wind data for 2005. *International Journal of Physical Sciences* 6(7): 1840-1850.
- Kamisan, N.A.B., Hussin, A.G. & Zubairi, Y.Z. 2010. Finding the best circular distribution for southwesterly monsoon wind direction in Malaysia. *Sains Malaysiana* 39(3): 387-393.
- Khatib, T., Mohamed, A. & Sopian, K. 2012. Optimization of a PV/wind micro-grid for rural housing electrification using a hybrid iterative/genetic algorithm: Case study of Kuala Terengganu, Malaysia. *Energy and Buildings* 47: 321-331.
- Kutil, R. 2012. Biased and unbiased estimation of the circular mean resultant length and its variance. *Statistics* 46(4): 549-561.
- Lee, A. 2010. Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(4): 477-486.
- Mardia, K.V. & Jupp, P.E. 2000. *Directional Statistics*. New York: John Wiley & Sons.
- Sanusi, N., Zaharim, A., Mata, S. & Sopian, K. 2017. A Weibull and finite mixture of the von Mises distribution for wind analysis in Mersing, Malaysia. *International Journal of Green Energy* 14(12): 1057-1062.
- Nor Hafizah Moslim
Institute of Advanced Studies
Universiti Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia
- Nurkhairany Amyra Mokhtar
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Cawangan Johor, Kampus Segamat
85000 Segamat, Johor Darul Takzim
Malaysia
- Yong Zulina Zubairi*
Centre for Foundation Studies in Science
Universiti Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia
- Abdul Gapor Hussin
Faculty of Defence Sciences and Technology
National Defence University of Malaysia
Kem Sungai Besi, 57000 Kuala Lumpur, Federal Territory
Malaysia

*Corresponding email; email: yzulina@um.edu.my

Received: 23 June 2020

Accepted: 19 November 2020