Applied Mathematical Sciences, Vol. 16, 2022, no. 12, 565 - 572
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2022.917218

Butterworth Wind Direction Statistical Model with Functional Relationship

Nurkhairany Amyra Mokhtar [1], Basri Badyalina [2,*], Kerk Lee Chang [3], Abu Sayed Md. Al Mamun [4] and Yong Zulina Zubairi [5]

[1,2,3] Mathematical Sciences Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA Johor Branch, Segamat Campus 85000 Segamat, Johor, Malaysia

[4]Department of Statistics, University of Rajshahi Rajshahi-6205, Bangladesh.

[5] Centre for Foundation Studies in Science Universiti Malaya, 50603 Kuala Lumpur, Malaysia

[*]Corresponding author

Abstract

Wind direction is vital in monitoring the global climate and weather patterns. The nature of wind direction that is circular makes the data unsuitable to be analysed with regular statistical techniques used for linear data. Therefore, this study analyses wind direction with special statistical techniques for circular data. We model the wind direction data in Butterworth, Penang, Malaysia for southwest monsoon season between June to August in 2016 and 2017 with the functional relationship for circular data through the von Mises distribution. In this study, the parameters are estimated through the maximum likelihood estimation with the least square method. The rotation parameter shows that the value is 6.0965 with low error concentration parameter that is 1.2873. The variance of the parameters is obtained through the Fisher information. This proposed model may be employed in studying the wind direction in Butterworth during southwest monsoon season.

Keywords: Butterworth; parameter estimation; von Mises distribution; Penang; monsoon season

1 Introduction

Modelling of wind direction is essential in applications related to navigation, safety, power transmission and engineering structure design [1, 2]. For example, boaters must know how strong the winds will be, as well as in which direction they will blow [3]. In the army and other fields of research, wind direction is an important factor in weather forecasting and aviation navigation [4]. Besides, wind direction may be modelled to predict wind turbine power and the movement of airborne pollutants [5, 3]. Wind direction is circular and measured in the form of angle.

Circular data is frequently represented as points on the unit circle's diameter that reflect the direction's position. Circular data is measured in the range $[0, 2\pi)$ radians or $[0^\circ, 360^\circ)$ [6]. Different statistical approach is required in treating circular data compared to the approaches used for linear observations by the reason of the geometrical behaviour of the circular observations [7, 8]. Circular observations arise in many fields such as environment science, animal movements (ecology), social science, musicology and physics [9]. It has an intrinsic periodicity that is not present in observations taken on a linear scale. Say, the angle 355° is significantly closer to the angle 5° compared to an angle of 330°, thus, a simple arithmetic mean, for example, might be rather deceptive [10].

We concentrate on wind direction that is angular in character in this research. Our interest is to study the wind direction relationship in Butterworth, located in Penang, Peninsular Malaysia. Penang is known as one of the most well-known destinations in Peninsular Malaysia with tropical climate [11, 12]. The Asian-Australian monsoon has a significant impact on Peninsular Malaysia's climate [13]. Southwesterly winds prevail Peninsular Malaysia from June to August (known as the southwest monsoon). Butterworth is an important business port in Penang as it locates a cargo terminal [14]. Figure 1 illustrates the spot of Butterworth in Peninsular Malaysia.

Figure 1. Location of Butterworth in Peninsular Malaysia.

Outliers would show as anomalies while dealing with data [7]. There are some approaches may be used to detect the outliers when handling the data. Mokhtar et al. (2019) proposed a *covratio* statistics in detecting outliers in circular data [15]. The method introduces *covratio* statistics which involves row deletion approach. Thus, we consider this method in detecting outliers in Butterworth wind direction data that is obtained from Malaysia Meteorological Department.

2 The Functional Relationship Model

In this paper, functional relationship model that is specifically observed for circular data is applied to wind direction data. The specialty of functional relationship model is that it considers the errors in each variable, compared to the ordinary linear regression model which only considers the error term in *y*-variable. The model is $Y = X + \alpha(\text{mod}\, 2\pi)$.

Given that $x_i = X_i + \delta_i$ and $y_i = Y_i + \varepsilon_i$ where i =1, 2,…, n for the parameter of rotation, α. The error terms in this model are distributed independently with the von Mises distribution where $\delta_i \sim VM(0,\kappa)$ and $\varepsilon_i \sim VM(0,\nu)$. The von Mises distribution is prominent in describing circular variables statistically. This distribution is corresponding to the normal distribution of linear data [16]. Its probability density function is given by

$$g(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\ \left(\kappa \cos\left(\theta - \mu\right)\right)$$

where $I_0(\kappa)$ is the modified Bessel function of order zero and the first kind, which is defined by :

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos\theta)\, d\theta$$

for $0 \le x < 2\pi$, $0 \le y < 2\pi$ and $\kappa > 0$ in which μ is the mean direction and κ is the error concentration parameter.

3 Parameter Estimation and Outlier Identification

Maximum likelihood is applied in estimating the parameters. Maximum likelihood estimation is an indispensable tool for many statistical modelling techniques [17]. The log likelihood function of the distribution is given by

$$\log L = -2n \log 2\pi - n \log I_0(\kappa) - n \log I_0(\lambda\kappa)$$
$$+ \kappa \sum_{i=1}^{n} \cos\left(x_i - X_i\right) + \lambda\kappa \sum_{1}^{n} \cos\left(y_i - \alpha - X_i\right)$$

The variable X_i may be estimated by using iteration from

$$\hat{X}_{i1} \approx \hat{X}_{10} + \frac{\sin\left(x_i - \hat{X}_{i0}\right) + \sin\left(y_i - \hat{\alpha} - \hat{X}_{i0}\right)}{\cos\left(x_i - \hat{X}_{i0}\right) + \cos\left(y_i - \hat{\alpha} - \hat{X}_{i0}\right)}$$

The estimation for rotation parameter, α obtained is

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

where $S = \sum_{i=1}^{n} \sin\left(y_i - \hat{X}_i\right)$ and $C = \sum_{i=1}^{n} \cos\left(y_i - \hat{X}_i\right)$

In estimating the concentration parameter, κ, the approximation by Fisher (1993) for equal error concentration case is given as follows:

$$A_1^{-1}(w) = \begin{cases} 2w + w^3 + \dfrac{5}{6}w^3 & \text{when } w < 0.53 \\ -0.4 + 1.39w + \dfrac{0.43}{(1-w)} & \text{when } 0.53 \le w < 0.85 \\ \dfrac{1}{w^3 - 4w^2 + 3w} & \text{when } w \ge 0.85 \end{cases}$$

Thus, we could estimate the error concentration with $\hat{\kappa} = A_1^{-1}(w)$

where $w = \frac{1}{n}\left\{\sum_{i=1}^{n} \cos\left(x_i - \hat{X}_i\right) + \sum_{i=1}^{n} \cos\left(y_i - \hat{\alpha} - \hat{X}_i\right)\right\}$.

Thus, it is

$$\hat{\kappa} = A_1^{-1}\left(\frac{1}{n}\left\{\sum_{i=1}^{n} \cos\left(x_i - \hat{X}_i\right) + \sum_{i=1}^{n} \cos\left(y_i - \alpha - \hat{X}_i\right)\right\}\right)$$

It is worthy to point out that the estimate becomes $\tilde{\kappa} = \frac{\hat{\kappa}}{2}$ since estimating concentration parameter of circular variable requires a correction factor of dividing $\hat{\kappa}$ by 2 [18]. The Fisher Information is applied in obtaining the covariance of the estimated parameters. From Fisher information matrix, F for $\hat{X}_1, \hat{X}_2, ..., \hat{X}_n$, $\tilde{\kappa}$ and $\hat{\alpha}$, the covariance matrix for estimated parameters is

$$\text{covariance}\begin{bmatrix} \tilde{\kappa} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{2n\,[A'(\tilde{\kappa})]} & 0 \\ 0 & \dfrac{2}{\kappa n A(\tilde{\kappa})} \end{bmatrix}.$$

Outliers are data occurrences that make fitting the desired model more challenging [19]. For this model, the existence of outlier is detected by *covratio* statistics. *Covratio* statistics is obtained from the ratio of the determinant of the covariance of the parameters, $|COV|$ and the determinant for the covariance matrix of diminished data, $|COV|_{(-i)}$ where i-th data is deleted one after another. The formula is given by $COVRATIO_{(-i)} = \frac{|COV|}{|COV_{(i)}|}$.

In detecting the existence of outliers, the cut-off point for this functional model is given by the equation $y = 3.7586n^{-0.71}$ [15]. The performance of this cut-off point in detecting outlier may be referred to Mokhtar et al. (2019) [15].

4 Results

The wind direction data of Butterworth for 2016 and 2017 are checked whether any outlier is present. The values of $COVRATIO_{(-i)}$ are obtained for each data and are checked if they exceed the cut-off point in outlier detection. With the sample size of 92, the cut-off point for this data becomes y = 0.15161. The $COVRATIO_{(-i)}$ values for each data is given in Figure 2. It is shown that no observation has the $COVRATIO_{(-i)}$ that exceeds the cut-off point. Hence, we may say that no outlier is detected and we proceed to fit the data with the functional relationship model.

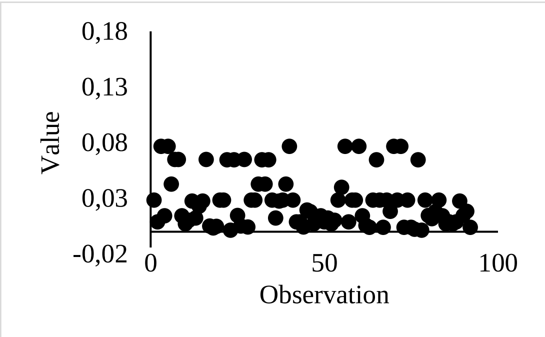

Figure 2. Values of $COVRATIO_{(-i)}$ for each data

The functional relationship is applied in modelling the wind direction in Butterworth during the monsoon season of southwest in both 2016 and 2017. This model is specially developed to fit circular variables statistically. Parameter estimates are obtained and the values are stated in Table 1 below.

Table 1. Parameter estimation of Butterworth wind direction.

Estimation	**Value**
Rotation, $\hat{\alpha}$	6.09653
Variance of $\hat{\alpha}$	0.03133
Concentration, $\tilde{\kappa}$	1.28726
Variance of $\tilde{\kappa}$	0.01869

According to the results in Table 1, we propose a model for Butterworth wind direction for the season of southwest monsoon in 2016 and 2017 with $Y = 6.09653 + X\ (mod\ 2\pi)$ in which the error concentration estimate of 1.28726. The estimate of rotation for the data is 6.09653 which is found to be very close to a whole circle, 2π radians. Checking on the variance for the estimate of rotation, $\hat{\alpha}$ and the estimate of concentration, $\tilde{\kappa}$, given that the values are very small given by 0.03133 and 0.01869, respectively. This is an indicator that the estimates of the parameters are consistent.

5 Conclusion

In conclusion, this paper studies Butterworth wind direction in 2016 and 2017 during southwest monsoon and proposes functional relationship model, for circular variables. The strength of this model is we take into account of the error terms for all variables statistically with appropriate distribution. The *covratio* statistics is considered in identifying the outlier in the data and it validates that no outlier exists in the data set. By using the von Mises distribution, the parameter estimates of the model are obtained through the maximum likelihood method. From this model, the rotation parameter obtained is big, 6.09653 which is very close to 2π radians, that is nearly a complete circle. It is estimated that the error concentration parameter of this data is small which is 1.28726. The model proposed in this study may be used in wind energy analysis for Butterworth, for many purposes.

Acknowledgements. We would like to thank Universiti Teknologi MARA Cawangan Johor Kampus Segamat for supporting this work.

References

[1] Al Yammahi, A., Marpu, P. R., & Ouarda, T. B. Modeling directional distributions of wind data in the United Arab Emirates at different elevations, *Arabian Journal of Geosciences*, **14** (2021), no. 9, 1-12. https://doi.org/10.1007/s12517-021-06864-3

[2] Jammalamadaka, S.R. & Lund U. J. The effect of wind direction on ozone levels: a case study, *Environmental and Ecological Statistics*, **13** (2006), 287–298. https://doi.org/10.1007/s10651-004-0012-7

[3] Sloughter, J. M., Gneiting, T. & Raftery, A.E. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging, *Monthly Weather Review*, **141** (2013), 2107–2119. https://doi.org/10.1175/mwr-d-12-00002.1

[4] Naik, S., Singh, N. K., Patidar, D. S., Panda, D. M & Parida, T. Design of a Wind Vane System with Anemometer Using Pic Microcontroller, *Interna-*

tional Journal of Research in Engineering and Science (IJRES), **8** (2018), 79-83.

[5] Shetty, R. P., Sathyabhama, A. & Pai, P. S. Comparison of modeling methods for wind power prediction: a critical study, *Frontiers in Energy*, **14** (2020), no. 2, 347-358. https://doi.org/10.1007/s11708-018-0553-3

[6] Mokhtar, N. A., Zubairi, Y. Z., Hussin, A. G. & Yunus, R. M. On parameter estimation of a replicated linear functional relationship model for circular variables, *MATEMATIKA*, **33** (2017), no. 2, 159–163. https://doi.org/10.11113/matematika.v33.n2.1010

[7] Mokhtar, N. A., Zubairi, Y. Z., Hussin, A. G., Badyalina, B., Ghazali, A. F., Ya'acob, F. F. & Kerk, L. C. Modelling wind direction data of Langkawi Island during Southwest monsoon in 2019 to 2020 using bivariate linear functional relationship model with von Mises distribution, *Journal of Physics: Conference Series*, *IOP Publishing*, **1988** (2021), no. 1, 012097. https://doi.org/10.1088/1742-6596/1988/1/012097

[8] Mahmood, E. A., Rana, S., Midi, H., & Hussin, A. G. Detection of outliers in univariate circular data using robust circular distance, *Journal of Modern Applied Statistical Methods*, **16** (2017), no. 2, 22. https://doi.org/10.22237/jmasm/1509495720

[9] Mastrantonio, G., Lasinio, G. J., Maruotti, A., & Calise, G. Invariance properties and statistical inference for circular data, S*tatistica Sinica*, **29** (2019), no. 1, 67-80. https://doi.org/10.5705/ss.202016.0067

[10] Landler, L., Ruxton, G. D., & Malkemper, E. P. Circular data in biology: advice for effectively implementing statistical procedures, *Behavioral Ecology and Sociobiology*, **72** (2018), no. 8, 1-10. https://doi.org/10.1007/s00265-018-2538-y

[11] Ghaderi, Z., Som, A. P. M. & Henderson J. C. Tourism crises and island destinations: Experiences in Penang, Malaysia, *Tourism Management Perspectives*, **2** (2012), no. 3, 79-84. https://doi.org/10.1016/j.tmp.2012.03.006

[12] Jan, N. A. M., Shabri, A., Hounkpè J. & Badyalina, B. Modelling nonstationary extreme streamflow in Peninsular Malaysia, *International Journal of Water*, **12** (2018), 116-140. https://doi.org/10.1504/ijw.2018.091380

[13] Jamaluddin, A. F., Kamal, M. I. M., Abdullah, M. H., & Marodzi, A. N. Comparison between Satellite-Derived Rainfall and Rain Gauge Observation over Peninsular Malaysia, *Sains Malaysiana,* **51** (2022), no. 1, 67-81. https://doi.org/10.17576/jsm-2022-5101-06

[14] Ming, L. S., Joshi, D., & Choy, T. *Instating a Free Commercial Zone at Penang's North Butterworth Container Terminal*. 2020.

[15] Mokhtar, N. A., Zubairi, Y. Z., Hussin, A. G. & Moslim, N. H. An Outlier Detection Method for Circular Linear Functional Relationship Model Using Covratio Statistics, *Malaysian Journal of Science*, **38** (2019) 46-54. https://doi.org/10.22452/mjs.sp2019no2.5

[16] Moslim, N. H., Mokhtar, N. A., Zubairi, Y. Z., & Hussin, A. G. Understanding the Behaviour of Wind Direction in Malaysia during Monsoon Seasons using Replicated Functional Relationship in von Mises Distribution, *Sains Malaysiana*, **50** (2021), no. 7, 2035-2045. https://doi.org/10.17576/jsm-2021-5007-18

[17] Myung, I. J. Tutorial on maximum likelihood estimation, *Journal of Mmathematical Psychology*, **47** (2003), no. 1, 90-100. https://doi.org/10.1016/s0022-2496(02)00028-7

[18] Caires, S. and Wyatt, L. R. 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** (2003), no. 2, 153-169. https://doi.org/10.1198/1085711031571

[19] Boukerche, A., Zheng, L. & Alfandi, O. Outlier Detection: Methods, Models, and Classification, *ACM Computing Surveys*, **53** (2020), no. 3, 1-37. https://doi.org/10.1145/3381028

Received: October 25, 2022; Published: November 10, 2022