Applied Mathematical Sciences, Vol. 15, 2021, no. 9, 471 - 479
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2021.914521

Error-in-Variables Model of Malacca Wind Direction Data with the von Mises Distribution in Southwest Monsoon

Nurkhairany Amyra Mokhtar [1,*], Basri Badyalina [1], Kerk Lee Chang [1], Fatin Farazh Ya'acob [2], Ahmad Faiz Ghazali [1] and Palaniappan Shamala [1]

[1] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, 85000 Segamat, Johor, Malaysia
[*] *Corresponding author*

[2] Faculty of Business and Management, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, 85000 Segamat, Johor, Malaysia

Abstract

Wind direction is an important aspect to be analysed to forecast the weather. In this study, we statistically investigate the relationship of wind direction data in Malacca, Malaysia during southwest monsoon in the year 2019 and 2020. The circular nature of wind direction requires the data to be analysed differently from linear data. We model wind direction data with the von Mises distribution by using an Error-in-Variables model, particularly functional relationship model. The goodness-of-fit of the data is supported by the QQ plots. The parameters of the model are estimated with maximum likelihood estimation and the covariance of the parameters are obtained by using Fisher Information matrix. The results show that the wind direction data in Malacca has a rotation parameter of 0.37348 with the concentration parameter of the error of 1.56960. The information of this study may be used for a better understanding in the prediction of wind energy.

Keywords: wind direction; von Mises distribution; circular data; statistical modelling

1 Introduction

Wind direction is an important aspect in wind energy. The data of wind direction is circular in nature. Other than wind direction, circular data arises in broad fields such as paleomagnetism, wildfire orientation and bioinformatics [1]. In ecology, data on the movement direction of animals is investigated and in medical sciences, protein structure is studied [2]. Circular data are a component in which directions can be described in a two-dimensional space, with the unit circle serving as the corresponding sample space [3]. The measurement is from 0° to 360°, or from 0 radian to 2π radians. Regular statistical analysis that is applied to linear data is not suitable for circular data due the wrapped around nature of a circle [4].

In this paper, our aim is to propose an error-in-variables model (EIVM), specifically the functional relationship model to the circular wind direction in Malacca, Malaysia during southwest monsoon in the years 2019 and 2020. Malacca, situated on the south-western coast of the Peninsular Malaysia (2.29 °N, 102.30 °E), has been listed as a UNESCO World Heritage Site [5]. Peninsular Malaysia is approximately 500 miles from north to south and about 200 miles from east to west. The local climate is tropical, with yearly monsoons of the southwest and northeast [6].

The von Mises distribution is widely employed to study circular data. The von Mises distribution is a circular analogue of the normal distribution [7]. A von Mises distribution, like a normal distribution, is symmetrical and unimodal, with concentration around the mean diminishing to form a bell shape [8]. This distribution is appropriate to model periodic variables [9].

The use of the distribution is frequently discussed in the literature. In 2017, the von Mises distribution was applied to model the sea turtle navigation, wolf movement and brain tumour growth [10]. In engineering, this distribution was applied to model the audio-source directions [11]. The circular variables associated with audio-source translation is directly modelled by the von Mises distribution. The probability distribution function of the Von Mises distribution is given by

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

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

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

where μ is the mean direction and κ is the concentration parameter for $0\le\theta<2\pi$ and $\kappa>0$. The concentration parameter denotes the distribution's spread [12].

2 Error-in-Variables Model

The functional relationship model is a type of EIVM. The EIVM differs from the traditional linear regression model in that the variables are obscured by measuring error [13]. In a traditional regression model, independent variables are assumed to be true, while dependent variables have estimation errors. Variable errors can be influenced by a lot of factors such as sampling error. Errors can be present in the independent variables too [14].

In EIVM, both variables x and y are considered with errors [15]. In fact, measuring errors occur in measurements, and ignoring these errors can have a direct impact on the estimators' desirable criteria [16]. EIV issues usually arise where the aim of the modelling is to gain physical insight into an operation [17]. EIVM is indeed the most statistically appropriate methodology for estimating reactivity ratios because it takes into account the existence of error throughout all variables [18].

There are three types of EIVM, namely functional, structural and ultrastructural EIVM. The statistical properties of the variables in a functional relationship model are fixed. Meanwhile, the variables in the structural relationship model are random, while the variables in the ultrastructural relationship model are a combination of the functional and structural relationship models [19]. Some examples of the application of EIVM were the description of the total nitrogen content of sandalwood in the forest farm in Hainan Province China and the research in measuring the level of carbon dioxide that polluted an urban environment [20].

3 Parameter Estimation

In this study, we model the wind direction data of Malacca by using a functional relationship model for circular data [25]. The model is given by

$$Y = \alpha + X \ (mod \ \ 2\pi) \tag{3}$$

where the X and Y variables are considered with the random errors δ_i and ε_i, respectively, where $X_i = x_i + \delta_i$ and $Y_i = y_i + \varepsilon_i$.

The errors are distributed with von Mises distribution of $\delta_i \sim VM(0, \kappa)$ and $\varepsilon_i \sim VM(0, \kappa)$ [21]. The parameter estimation is derived by using the method of maximum likelihood. The log-likelihood function of the distribution is given by

$$\begin{aligned} log\, L\,(\alpha, \kappa, X; x, y) = & -2n\, log\, 2\,\pi - 2n\, log\, I_0\,(\kappa) \\ & + \kappa \sum_{i=1}^{n} cos(x_i - X_i) + \kappa \sum_{i=1}^{n} cos(y_i - \alpha - X_i) \end{aligned} \tag{4}$$

for $0 \leq x < 2\pi$, $0 \leq y < 2\pi$ and $\kappa > 0$ where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero, which can be defined by :

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa\, cos\theta} d\theta \tag{5}$$

where κ is the concentration parameter of the measurement error.
The estimation of the variable X_i is obtained iteratively given by

$$\hat{X}_{i1} \approx \hat{X}_{10} + \frac{sin\left(x_i - \hat{X}_{i0}\right) + sin\left(y_i - \hat{\alpha} - \hat{X}_{i0}\right)}{cos\left(x_i - \hat{X}_{i0}\right) + cos\left(y_i - \hat{\alpha} - \hat{X}_{i0}\right)} \tag{6}$$

and the estimation of the rotation parameter is given by

$$\hat{\alpha} = \begin{cases} tan^{-1}\left\{\frac{S}{C}\right\} & \text{when } S > 0, C > 0 \\ tan^{-1}\left\{\frac{S}{C}\right\} + \pi & \text{when } C < 0 \\ tan^{-1}\left\{\frac{S}{C}\right\} + 2\pi & \text{when } S < 0, C > 0 \end{cases} \tag{7}$$

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

The Fisher approximation is applied to estimate the concentration parameter κ for the case of equal error concentration [22]. The approximation is given by:

$$A_1^{-1}(w) = \begin{cases} 2w + w^3 + \frac{5}{6}w^3 & \text{when } w < 0.53 \\ -0.4 + 1.39w + \frac{0.43}{(1-w)} & \text{when } 0.53 \le w < 0.85 \\ \frac{1}{w^3 - 4w^2 + 3w} & \text{when } w \ge 0.85 \end{cases} \tag{8}$$

Thus, the estimation becomes

$$\hat{\kappa} = A_1^{-1}(w) \text{ where } w = \frac{1}{n}\{\textstyle\sum_{i=1}^{n} cos(x_i - \hat{X}_i) + \sum_{i=1}^{n} cos(y_i - \hat{\alpha} - \hat{X}_i)\}. \tag{9}$$

Thus, it is

$$\hat{\kappa} = A_1^{-1}\left(\frac{1}{n}\left\{\sum_{i=1}^{n} cos(x_i - \hat{X}_i) + \sum_{i=1}^{n} cos(y_i - \alpha - \hat{X}_i)\right\}\right) \tag{10}$$

For circular data, the estimation of a concentration parameter is to be corrected by dividing it by 2 [21]. Hence, the estimate becomes $\tilde{\kappa} = \frac{\hat{\kappa}}{2}$.

4 Outlier Identification

Outliers are data points that deviate dramatically from the rest of the data set [23]. Spotting outliers yields substantial actionable information in a wide range of applications, such as fraud detection [24]. In this section, *covratio* method is studied to detect the presence of outlier for circular data [21]. The *covratio* statistics is constructed from the Fisher Information matrix of the parameter estimates. It is defined by

$$COVRATIO_{(-i)} = \frac{|COV|}{|COV_{(-i)}|} \tag{11}$$

where $|COV|$ is the determinant of the covariance matrix of the parameter estimates is given by

$$|COV| = \frac{1}{n^2\tilde{\kappa}[A'_1(\tilde{\kappa})]^2} \tag{12}$$

and $|COV_{(-i)}|$ is the determinant of the covariance matrix for the reduced data set by excluding the *i*-th row.

The cut-off equation for outlier detection in bivariate functional relationship model for circular data with $y = 3.7586n^{-0.71}$. Since the sample size used for this study is n =139, thus the cut-off point becomes $y = 3.7586n^{-0.71} = 0.113106$. This cut-off equation is used to detect the outlier for the wind direction data in this study.

5 Results

The preliminary univariate analysis of Malacca wind direction data during the southwest monsoon is described graphically with the rose diagrams for the years 2019 and 2020. Figures 1 and 2 are the rose diagrams of the wind direction data of Malacca in 2019 and 2020, respectively.

From the rose diagrams, we can see that the patterns of the data for both years are different. Hence, in this study, we investigate the relationship of the data for both years and describe it in the form of EIVM, specifically with the bivariate functional relationship model for circular wind direction data.

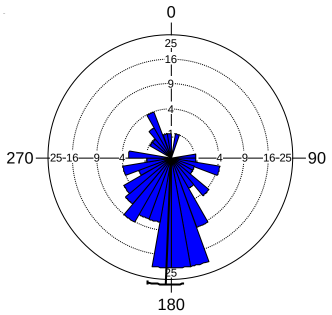

Figure 1. The rose diagram of Malacca wind direction data in 2019.

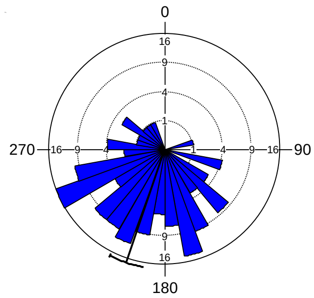

Figure 2. The rose diagram of Malacca wind direction data in 2020.

The Q-Q plots of the von Mises distribution for Malacca wind direction data during southwest monsoon of both years 2019 and 2020 are shown in Figures 3 and 4, respectively. The QQ-plot is a graphical method for determining the suitability of a statistical approach for the data at hand [4].

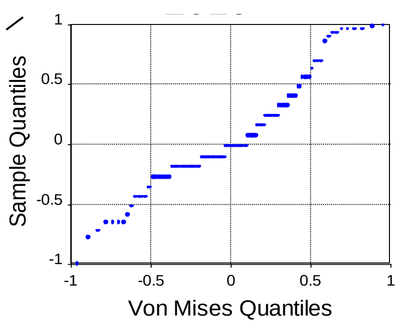

Figure 3. Von Mises Q-Q plot of wind direction data during southwest monsoon in 2019

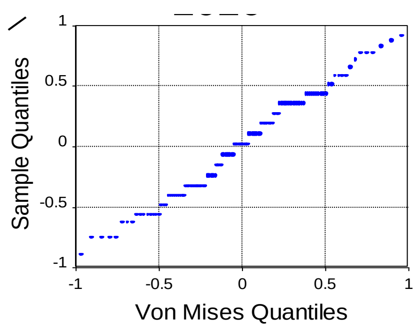

Figure 4. Von Mises Q-Q plot of wind direction data during southwest monsoon in 2020.

We try to figure out if any outlier presents in the data. The values of $COVRATIO_{(-i)}$ for the wind direction data are described in Figure 5 and none of the values exceeds the cut-off equation of y = 0.113106. Therefore, no observation is categorised as the outlier. Hence, the wind direction data during southwest monsoon of Malacca is modelled without having to eliminate any of the data.

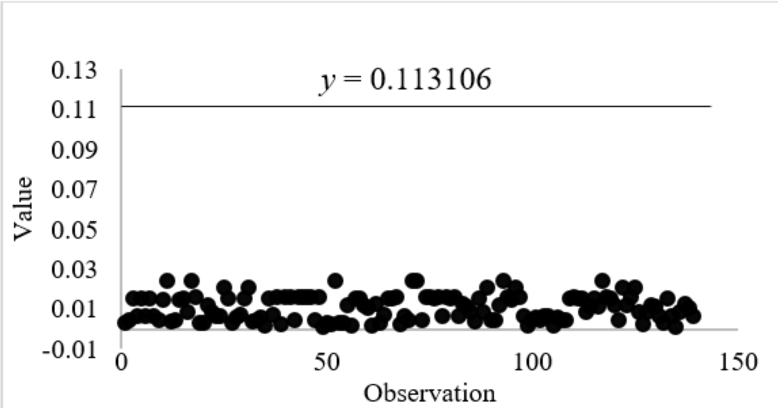

Figure 5. Outlier detection of Malacca wind direction data using *covratio* statistics.

The data is then fitted to the functional relationship model that is discussed in Sections 3 and 4. Table 1 shows the values of parameter estimate for the data.

Table 1. Parameter estimates of Malacca wind direction data.

Detail	Value
Rotation parameter, $\hat{\boldsymbol{\alpha}}$	0.37348
Variance of $\hat{\boldsymbol{\alpha}}$	0.01496
Concentration parameter, $\tilde{\boldsymbol{\kappa}}$	1.56960
Variance of $\tilde{\boldsymbol{\kappa}}$	0.01538

From Table 1, the model proposed for wind direction data of Malacca during southwest monsoon season in 2019 and 2020 is $Y = 0.37348 + X\ (mod\ 2\pi)$ with a small error concentration parameter of 1.56960. The parameter estimates show that the rotation parameter is 0.37348 which is very near to 0 radians. The variance of $\hat{\alpha}$ and $\tilde{\kappa}$ are small given by 0.01496 and 0.01538, respectively, which indicate the consistency of the estimated parameters. Also, the variances show that the values are less dispersed.

6 Conclusion

As a conclusion, this study proposes a statistical model of wind direction data in Malacca, Malaysia during southwest monsoon season in 2019 and 2020 by using an error-in-variables model, specifically the bivariate functional relationship model. In this model, both of the variables x and y are considered with the presence of error terms. The *covratio* statistics is adopted in determining if there is any outlier exists in the data. It shows that there is no outlier presents in the data. The parameters of this data are estimated by using the maximum likelihood method of the von Mises distribution. The results show that the rotation parameter is very near to 0 radian and the value of the concentration parameter of the error terms is small and less concentrated. This model may be applied to forecast the wind direction of Malacca during the southwest monsoon and to aid in the management of outdoor activities while keeping safety and weather in account.

Acknowledgements. The authors gratefully acknowledge the financial support by Universiti Teknologi MARA Cawangan Johor under grant Bestari 600-UiTMCJ (PJIA. 5/2) for supporting this study.

References

[1] C. Ley, & T. Verdebout, *Modern directional statistics*, CRC Press, 2017. https://doi.org/10.1201/9781315119472

[2] J. Cremers, & I. Klugkist, One direction? A tutorial for circular data analysis using R with examples in cognitive psychology, *Frontiers in Psychology*, **9** (2018), 2040. https://doi.org/10.3389/fpsyg.2018.02040

[3] C. E. Rodríguez, G. Núñez-Antonio & G. Escarela, A Bayesian mixture model for clustering circular data, *Computational Statistics & Data Analysis*, **143** (2020), 106842. https://doi.org/10.1016/j.csda.2019.106842

[4] N. A. Mokhtar, Y. Z. Zubairi, A. G. Hussin & N. H. Moslim, Parameter Estimation and Covariance Matrix of the Linear Functional Relationship Model for Circular Variables with Unequal Error Concentrations, *ASM Science Journal*, **13** (2020), 1-17. https://doi.org/10.32802/asmscj.2020.sm26(1.17)

[5] G. Manteghi, T. Mostofa, & M. P. B. M. Noor, A Field Investigation on the Impact of the Wider Water Body on-Air, Surface Temperature and Physiological Equivalent Temperature at Malacca Town, *International Journal of Environmental Science and Development*, **11** (2020), 286-291. https://doi.org/10.18178/ijesd.2020.11.6.1264

[6] N. A. M. Jan, A. Shabri, J. Hounkpè & B. Badyalina, B. Modelling non-stationary extreme streamflow in Peninsular Malaysia, *International Journal of Water*, **12** (2018), 116-140. https://doi.org/10.1504/ijw.2018.10012408

[7] K. V. Mardia, G. Hughes, C. C. Taylor, & H. Singh, A multivariate von Mises distribution with applications to bioinformatics, *Canadian Journal of Statistics*, **36** (2008), 99-109. https://doi.org/10.1002/cjs.5550360110

[8] G. D. Ruxton, Testing for departure from uniformity and estimating mean direction for circular data, *Biology Letters*, **13** (2017), 2016075620160756. https://doi.org/10.1098/rsbl.2016.0756

[9] S. Takamichi, Y. Saito, N. Takamune, D. Kitamura, & H. Saruwatari, Phase reconstruction from amplitude spectrograms based on directional-statistics deep neural network, *Signal Processing*, **169** (2020), 1-12. https://doi.org/10.1016/j.sigpro.2019.107368

[10] T. Hillen, K. J. Painter, A. C. Swan & A. D. Murtha, Moments of von Mises and Fisher distributions and applications, *Mathematical Biosciences & Engineering*, **14** (2017), 673-694. https://doi.org/10.3934/mbe.2017038

[11] Y. Ban, X. Alameda-Pineda, C. Evers & R. Horaud, Tracking multiple audio sources with the von Mises distribution and variational EM, *IEEE Signal Processing Letters*, **26** (2019), 798-802. https://doi.org/10.1109/lsp.2019.2908376

[12] M. D. A. Ferreira, F. R. M. Rodrigues, M. A. Luersen, P. C. Borges, R. Nanda & M. R. D. Almeida, Von Mises stresses on Mushroom-loop archwires for incisor retraction: a numerical study, *Dental Press Journal of Orthodontics*, **25** (2020), 44-50. https://doi.org/10.1590/2177-6709.25.4.044-050.oar

[13] A. G. Hussin, Numerical Comparisons for Various Estimators of Slope Parameters for Unreplicated Linear Functional Model, *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, **20** (2004), 19-30.

[14] Z. Chen, X. Wang & H. Wang, Preliminary research on total nitrogen content prediction of sandalwood using the error-in-variable models based on digital image processing, *PloS one*, **13** (2018), e0202649. https://doi.org/10.1371/journal.pone.0202649

[15] N. A. Mokhtar, Y. Z. Zubairi, A. G. Hussin, & R. M. Yunus, On parameter estimation of a replicated linear functional relationship model for circular variables, *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, **33** (2017), 159-163. https://doi.org/10.11113/matematika.v33.n2.1010

[16] A. S. M. A. Mamun, A. G. Hussin, Y. Z. Zubairi, A. H. M. Imon & S. Rana, Small-Sample confidence interval for the slope of linear structural relationship model, *Electronic Journal of Applied Statistical Analysis*, **10** (2017), 374-383.

[17] T. Söderström, *Errors-in-variables methods in system identification*, Springer, 2018.

[18] A. J. Scott & A. Penlidis, Computational Package for Copolymerization Reactivity Ratio Estimation: Improved Access to the Error-in-Variables-Model, *Processes*, **6** (2018), 8. https://doi.org/10.3390/pr6010008

[19] S. F. Hassan, A. G. Hussin & Y. Z. Zubairi, Estimation of Functional Relationship Model for Circular Variables and Its Application in Measurement Problem, *Chiang Mai Journal of Science*, **37** (2010), 195-205.

[20] Z. Chen, X. Wang & H. Wang, Preliminary research on total nitrogen content prediction of sandalwood using the error-in-variable models based on digital image processing. *PloS one*, **13** (2018), e0202649. https://doi.org/10.1371/journal.pone.0202649

[21] N. A. Mokhtar, Y. Z. Zubairi, A. G. Hussin & N. H. Moslim, 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

[22] N. I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, 1993. https://doi.org/10.1017/cbo9780511564345

[23] J. Chen, S. Sathe, C. Aggarwal & D. Turaga, Outlier detection with autoencoder ensembles, *in Proceedings of the 2017 SIAM International Conference on Data Mining-Society for Industrial and Applied Mathematics,* pp. 90-98, June, 2017. https://doi.org/10.1137/1.9781611974973.11

[24] H. Wang, M. J. Bah & M. Hammad, Progress in outlier detection techniques: A survey, *IEEE Access,* **7** (2019), 107964-108000. https://doi.org/10.1109/access.2019.2932769

Received: May 27, 2021; Published: June 27, 2021