Sains Malaysiana 52(8)(2023): 2419-2430
http://doi.org/10.17576/jsm-2023-5208-18

Modelling Wind Speed Data in Pulau Langkawi With Functional Relationship

(Memodelkan Data Kelajuan Angin di Pulau Langkawi dengan Perhubungan Fungsian)

Nur Ain AL-Hameefatul Jamaliyatul[1], Basri Badyalina[1], Nurkhairany Amyra Mokhtar[1,*], Adzhar Rambli[2], Yong Zulina Zubairi [3] & Adilah Abdul Ghapor[4]

[1]*Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Johor Branch, Segamat Campus, 85000 Segamat, Johor, Malaysia.*
[2]*School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.*
[3]*Institute of Advanced Studies, Universiti Malaya, 50603 Kuala Lumpur, Malaysia*
[4]*Department of Decision Science, Faculty of Business and Economics, Universiti Malaya, 50603 Kuala Lumpur, Federal Territory, Malaysia.*

Received: 16 December 2022/Accepted: 1 August 2023

ABSTRACT

Wind speed influenced weather predictions, aerospace operations, and maritime operations, construction projects. This research aims to examine the relationship between Pulau Langkawi wind speed data during the southwest monsoons in 2019 and 2020. To model wind speed data that follows a normal distribution. An error-in-variables model (EIVM) is utilised, which is a linear functional relationship model (LFRM). The QQ-plots will be utilised to investigate the adequacy of the model's fit. The maximum likelihood estimation (MLE) approach is employed to estimate the parameters of the model, while the covariance is calculated using the Fisher Information matrix. As a result, it is found that the estimated values demonstrate consistency and reduced dispersion. Thus, the findings could lead to a better knowledge of wind energy prediction.

Keywords: Linear functional relationship model; maximum likelihood estimation; wind speed

ABSTRAK

Kelajuan angin mempengaruhi ramalan cuaca, operasi aeroangkasa, operasi maritim dan projek pembinaan. Tujuan penyelidikan ini adalah untuk mengkaji hubungan antara data kelajuan angin Pulau Langkawi di Malaysia selama musim barat daya pada tahun 2019 dan 2020. Untuk memodelkan data kelajuan angin yang tertabur secara normal, model ralat dalam pemboleh ubah telah digunakan iaitu model hubungan kefungsian linear. Plot QQ digunakan untuk mengkaji kebolehupayaan penyuaian model terhadap data. Pendekatan anggaran maksimum digunakan untuk menganggar parameter model dan matriks Maklumat Fisher digunakan untuk menghitung kovarians. Keputusan menunjukkan bahawa nilai anggaran adalah konsisten dan kurang terserak. Hasil kajian ini boleh meningkatkan pengetahuan berkenaan ramalan tenaga angin.

Kata kunci: Anggaran kebolehjadian maksimum; kelajuan angin; model hubungan fungsi linear

INTRODUCTION

Over the years, regression methods have continued to be one of the active research areas. In traditional linear regression, it is assumed that there is a linear relationship, and only the variable Y is observed with an error. The traditional regression model considers the explanatory variable X is true, while the dependent variable Y has estimation errors (Mokhtar et al. 2021a).

In the linear functional relationship model (LFRM), both X and Y are linearly connected, identified with error, and can be represented by Equation (1)

$$Y = \alpha + \beta X + \varepsilon \quad (1)$$

where X is the explanatory variable; Y is the dependent variable; α is the value of Y intercept; β is the slope of the linear regression model; and ε is the error term of variable Y. In the traditional linear regression model, X is assumed to be constant and error-free (Ghapor et al. 2015). However, in reality, this may be problematic and cause difficulties in estimation when errors exist in X and Y. Variable errors can be caused by various factors, including sampling error, and observation error. When the measurement error in X is relatively large, the parameter estimation using the traditional linear regression model will result in obvious systematic errors (Chen, Wang & Wang 2018; Mokhtar, Badyalina & Zubairi 2022). Many areas, including econometrics, environmental sciences, engineering, and manufacturing, might experience measurement errors (Buonaccorsi 1996; Doganaksoy & Van Meer 2015).

Adcock introduced the error-in-variables model (EIVM) in 1878 (Arif, Zubairi & Hussin 2019). EIVM varies from traditional linear regression models because, in EIVM, the error terms are considered for the variables. Unlike the traditional linear regression model, EIVM does not distinguish between explanatory (X) and dependent (Y) variables (Hassan, Hussin & Zubairi 2010). There have been several studies by numerous authors on EIVM parameter estimation (Fuller 1987; Lindley 1947; Rosenhead 1963).

EIVM is divided into functional, structural, and ultrastructural relationships (Mokhtar et al. 2021b). This study will concentrate on the functional relationship. The linear functional relationship model (LFRM) can be categorised as either unreplicated or replicated, each with its specific guidelines or recommendations. To estimate the parameters in LFRM, it is important to assume the value of λ is set to one (Arif, Zubairi & Hussin 2022, 2021).

Hanoon et al. (2022) employed three machine learning models, namely Gaussian process regression (GPR), bagged regression trees (BTs) and support vector regression (SVR), to forecast the weekly wind speed (maximum, mean, and minimum) at 14 measurement stations in Malaysia from 2000 and 2019. Numerous studies have investigated wind direction modelling using functional relationships across various years. However, ongoing research is still being conducted on applying functional relationship to model wind speed (Mokhtar, Badyalina & Zubairi 2022; Mokhtar et al. 2021a).

Therefore, we would like to propose a statistical wind speed model using the LFRM with application to wind speed data in Pulau Langkawi, Malaysia. The data was obtained from the Malaysian Meteorological Department throughout the southwest monsoon from the 18th of May to the 15th of September of the year 2019 and year 2020 in Langkawi, it was recorded at a latitude of 6°20' N, and a longitude of 99°44' E. Its highest daily reading is at an altitude of 6.4 meters (Malaysian Meteorology Department 2019). We will investigate and identify the relationship between wind speed data for 2019 and 2020 using a bivariate linear functional relationship model.

The significance of assessing the relationship of wind speed data which occurred during 2019 and 2020 during the southwest monsoon using LFRM, can assist in studies of potential wind energy and provides enhanced comprehension of the behaviour of wind speed with error terms considered for all variables. We intend to dedicate future research endeavours to exploring the Northeast Monsoon, recognising its importance and the need to broaden our understanding of monsoon dynamics beyond the scope of this current study. Understanding the relationship between variables is important to conclude a statistical analysis.

MATERIALS AND METHODS

MAXIMUM LIKELIHOOD ESTIMATION METHOD (MLE)

The functional relationship is used to model wind speed data. The linear functional relationship model (LFRM) examines the relationships between variables. The maximum likelihood estimation (MLE) approach is preferred for calculating LFRM parameters because the estimators' estimated variance-covariance matrix can be easily generated (Arif, Zubairi & Hussin 2021).

In the LFRM, both X_i and Y_i are linearly connected, identified with error, and can be represented by the equation

$$Y_i = \alpha + \beta X_i \text{ for } i = 1,2,3,...,n \tag{2}$$

with α and β are the intercept value and slope of the model, respectively. In the LFRM, x_i and y_i variables are subject to random errors, δ_i and ε_i for $i = 1,2,3,..., n$ which n is the number of parameter. The error terms δ_i and ε_i are assumed to be mutually independent random variables with normally distributed distributions.

$$x_i = X_i + \delta_i \text{ and } y_i = Y_i + \varepsilon_i \tag{3}$$

$$\delta_i \sim N(0,\sigma_\delta^2) \text{ and } \varepsilon_i \sim N(0,\sigma_\varepsilon^2) \tag{4}$$

According to Arif, Zubairi and Hussin (2020), (n + 4) parameters must be determined in LFRM, which are α, β, σ_ε^2, σ_δ^2, and the incidental parameters $X_1, X_2, ..., X_n$. The log-likelihood function is given by

$$\log L = -n\log(2\pi) - \frac{n}{2}(\log\sigma_\delta^2 + \log\sigma_\varepsilon^2) - \frac{1}{2\sigma_\delta^2}\sum_{i=1}^{n}(x_i - X_i)^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2 \tag{5}$$

According to Ghapor et al. (2014), when $\hat{X}_i = x_i$ and σ_δ^2 approaches zero, the likelihood function will approach infinity. It leads to inconsistencies of the estimators (Fuller 1987). Nevertheless, Ghapor et al. (2015b) have stated that the information on either one of the variances or the ratio of the two variances is needed to overcome the inconsistencies of the estimators. The ratio of error variances is assumed to be $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$, where λ is known. Now, there are (n+3) parameters to be estimated, namely α, β, σ_δ^2, and $X_1, ..., X_n$ (Fuller 1987; Rosenhead 1963). The log-likelihood function is given by

$$\log L = -n\log(2\pi) - \frac{n}{2}\log\lambda - n\log\sigma_\delta^2 - \frac{1}{2\sigma_\delta^2}\left\{\sum_{i=1}^{n}(x_i - X_i)^2 + \frac{1}{\lambda}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2\right\} \tag{6}$$

The LFRM parameters will be determined using the MLE approach.

a) MLE approach for α

The first partial derivative of equation (6) with respect to α is:

$$\frac{\delta}{\delta\alpha}(\log L) = -\frac{1}{2\lambda\sigma_\delta^2}\left\{(2)(-1)\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i)\right\}$$

$$\frac{\delta}{\delta\alpha}(\log L) = \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i)$$

and by setting $\frac{\delta}{\delta X_i}(\log L) = 0$

$$\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i) = 0$$

$$\frac{1}{\sigma_\delta^2}\left(\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\hat{\alpha} - \sum_{i=1}^{n}\hat{\beta}\hat{X}_i\right) = 0$$

and by setting

$$\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\hat{\alpha} - \sum_{i=1}^{n}\hat{\beta}\hat{X}_i = 0$$

$$n\hat{\alpha} = \sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\hat{\beta}\hat{X}_i$$

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\hat{\beta}\hat{X}_i}{n}$$

$$\hat{\alpha} = \overline{y} - \hat{B}\overline{x} \tag{7}$$

b) MLE approach for X_i,

The first partial derivative of equation (6) with respect to X_i is:

$$\frac{\delta}{\delta X_i}(\log L) = -\frac{1}{2\sigma_\delta^2}\left\{\sum_{i=1}^{n}(2)(-1)(x_i - \hat{X}_i) + \frac{1}{\lambda}\sum_{i=1}^{n}(2)(-\hat{\beta})(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i)\right\}$$

$$\frac{\delta}{\delta X_i}(\log L) = \frac{1}{\sigma_\delta^2}\left\{\sum_{i=1}^{n}(x_i - \hat{X}_i) + \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i)\right\}$$

$$\frac{\delta}{\delta X_i}(\log L) = 0$$

and by setting

$$\frac{1}{\sigma_\delta^2}\left\{\sum_{i=1}^{n}(x_i - \hat{X}_i) + \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i)\right\} = 0$$

$$\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\hat{X}_i + \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}y_i - \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}\hat{\alpha} - \frac{\hat{\beta}^2}{\lambda}\sum_{i=1}^{n}\hat{X}_i = 0$$

$$\sum_{i=1}^{n}\hat{X}_i\left(-1 - \frac{\hat{\beta}^2}{\lambda}\right) = -\sum_{i=1}^{n}x_i - \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}y_i + \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}\hat{\alpha}$$

$$\sum_{i=1}^{n}\hat{X}_i = \frac{-\sum_{i=1}^{n}x_i - \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}y_i + \frac{\hat{\beta}}{\lambda}\sum_{i=1}^{n}\hat{\alpha}}{\left(-1 - \frac{\hat{\beta}^2}{\lambda}\right)}$$

$$\sum_{i=1}^{n}\hat{X}_i = \frac{\lambda\sum_{i=1}^{n}x_i + \hat{\beta}\sum_{i=1}^{n}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)}$$

$$\hat{X}_i = \frac{\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)} \tag{8}$$

c) MLE approach for β

The first partial derivative of equation (6) with respect to β is:

$$\frac{\delta}{\delta\beta}(\log L) = -\frac{1}{2\lambda\sigma_\delta^2}\left\{\sum_{i=1}^{n}(2)(-\hat{X}_i)(y_i-\hat{\alpha}-\hat{\beta}\hat{X}_i)\right\}$$

$$\frac{\delta}{\delta\beta}(\log L) = \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i(y_i-\hat{\alpha}-\hat{\beta}\hat{X}_i)$$

and by setting $\frac{\delta}{\delta\beta}(\log L)=0$

$$\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i(y_i-\hat{\alpha}-\hat{\beta}\hat{X}_i)=0$$

$$\sum_{i=1}^{n}\hat{X}_i y_i - \sum_{i=1}^{n}\hat{X}_i\hat{\alpha} - \sum_{i=1}^{n}\hat{\beta}\hat{X}_i^{\,2} = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\hat{X}_i(y_i-\hat{\alpha})}{\sum_{i=1}^{n}\hat{X}_i^{\,2}} \tag{9}$$

From the previous step in b), we have obtained $\hat{X}_i$. Then, substitute $\hat{X}_i = \frac{\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)}$ into $\hat{\beta}$ to get

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left(\frac{\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)}\right)(y_i-\hat{\alpha})}{\sum_{i=1}^{n}\left(\frac{\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)}\right)^2}$$

$$\hat{\beta}\sum_{i=1}^{n}\left(\frac{\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)}\right)^2 = \sum_{i=1}^{n}\left(\frac{\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})}{\left(\lambda+\hat{\beta}^2\right)}\right)(y_i-\hat{\alpha})$$

$$\frac{\hat{\beta}}{\left(\lambda+\hat{\beta}^2\right)^2}\sum_{i=1}^{n}\left(\lambda x_i + \hat{\beta}\left(y_i-\hat{\alpha}\right)\right)^2$$

$$= \frac{1}{\left(\lambda+\hat{\beta}^2\right)}\sum_{i=1}^{n}\left(\lambda x_i + \hat{\beta}(y_i-\hat{\alpha})\right)\left(y_i-\hat{\alpha}\right)$$

$$\hat{\beta}\sum_{i=1}^{n}\left(\lambda x_i + \hat{\beta}\left(y_i-\hat{\alpha}\right)\right)^2 = \left(\lambda+\hat{\beta}^2\right)\sum_{i=1}^{n}\left(\lambda x_i\left(y_i-\hat{\alpha}\right)\right.$$

$$\left.+\hat{\beta}(y_i-\hat{\alpha})^2\right)$$

$$\hat{\beta}\sum_{i=1}^{n}\left(\left(\lambda x_i\right)^2 + 2\lambda x_i\hat{\beta}\left(y_i-\hat{\alpha}\right) + \left(\hat{\beta}\left(y_i-\hat{\alpha}\right)\right)^2\right)$$

$$= \sum_{i=1}^{n}\begin{pmatrix}\lambda x_i\left(\lambda+\hat{\beta}^2\right)\left(y_i-\hat{\alpha}\right)+\\ \hat{\beta}\left(\lambda+\hat{\beta}^2\right)(y_i-\hat{\alpha})^2\end{pmatrix}$$

$$\hat{\beta}\sum_{i=1}^{n}\left(\lambda x_i\right)^2 + \hat{\beta}\sum_{i=1}^{n}2\lambda x_i\hat{\beta}\left(y_i-\hat{\alpha}\right) + \hat{\beta}\sum_{i=1}^{n}\left(\hat{\beta}\left(y_i-\hat{\alpha}\right)\right)^2$$

$$= \begin{pmatrix}\sum_{i=1}^{n}\lambda x_i\left(\lambda+\hat{\beta}^2\right)\left(y_i-\hat{\alpha}\right)+\\ \sum_{i=1}^{n}\hat{\beta}\left(\lambda+\hat{\beta}^2\right)(y_i-\hat{\alpha})^2\end{pmatrix}$$

$$\lambda^2\hat{\beta}\sum_{i=1}^{n}x_i^{\,2} + 2\lambda\hat{\beta}^2\sum_{i=1}^{n}x_i\left(y_i-\hat{\alpha}\right) + \hat{\beta}^3\sum_{i=1}^{n}\left(y_i-\hat{\alpha}\right)^2$$

$$= \begin{pmatrix}\sum_{i=1}^{n}\lambda x_i\left(\lambda+\hat{\beta}^2\right)\left(y_i-\hat{\alpha}\right)+\\ \sum_{i=1}^{n}\hat{\beta}\left(\lambda+\hat{\beta}^2\right)(y_i-\hat{\alpha})^2\end{pmatrix}$$

$$\hat{\beta}\sum_{i=1}^{n}\lambda^2 x_i^{\,2} + \left(2\lambda\hat{\beta}^2 - \lambda\left(\lambda+\hat{\beta}^2\right)\right)\sum_{i=1}^{n}x_i(y_i-\hat{\alpha})$$

$$+\left(\hat{\beta}^3 - \hat{\beta}\left(\lambda+\hat{\beta}^2\right)\right)\sum_{i=1}^{n}(y_i-\hat{\alpha})^2 = 0$$

$$\hat{\beta}\sum_{i=1}^{n}\lambda^2 x_i^{\,2} + \left(\lambda\hat{\beta}^2-\lambda^2\right)\sum_{i=1}^{n}x_i(y_i-\hat{\alpha}) + \hat{\beta}\lambda\sum_{i=1}^{n}(y_i-\hat{\alpha})^2 = 0$$

Replacing $\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}$,

$$\hat{\beta}\lambda\sum_{i=1}^{n} x_i^2 + \left(\hat{\beta}^2 - \lambda\right)\sum_{i=1}^{n} x_i\left(y_i - \left(\bar{y} - \bar{x}\hat{\beta}\right)\right)$$

$$+\hat{\beta}\sum_{i=1}^{n}\left(y_i - \left(\bar{y} - \bar{x}\hat{\beta}\right)\right)^2 = 0$$

$$\hat{\beta}\lambda\sum_{i=1}^{n} x_i^2 + \hat{\beta}^2\sum_{i=1}^{n} x_i\left(y_i - \bar{y}\right) - \lambda\sum_{i=1}^{n} x_i\left(y_i - \bar{y}\right)$$

$$-\lambda\hat{\beta}\sum_{i=1}^{n} x_i\bar{x} - \hat{\beta}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 - 2\bar{x}\hat{\beta}^2\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = 0$$

$$\hat{\beta}^2\sum_{i=1}^{n} x_i\left(y_i - \bar{y}\right) + \hat{\beta}\left(\lambda\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right.$$

$$\left.-\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right) - \lambda\sum_{i=1}^{n} x_i\left(y_i - \bar{y}\right)^2 = 0$$

$$\hat{\beta} = \frac{S_{yy} - \lambda S_{xx} + \sqrt{\left(S_{yy} - \lambda S_{xx}\right)^2 + 4\lambda S_{xy}^2}}{2S_{xy}}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$,

$S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$,

and $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

d) MLE approach for σ_δ^2

The first partial derivative of equation (6) with respect to σ_δ^2 is:

$$\frac{\delta}{\delta\sigma_\delta^2}(\log L) = -\frac{n}{\sigma_\delta^2} + \frac{1}{2\left(\sigma_\delta^2\right)^2}\left\{\sum_{i=1}^{n}(x_i - X_i)^2\right.$$

$$\left.+\frac{1}{\lambda}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2\right\}$$

and by setting $\frac{\delta}{\delta\sigma_\delta^2}(\log L) = 0$

$$-\frac{n}{\sigma_\delta^2} + \frac{1}{2\left(\sigma_\delta^2\right)^2}\left\{\sum_{i=1}^{n}(x_i - X_i)^2 + \frac{1}{\lambda}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2\right\} = 0$$

$$\frac{n}{\sigma_\delta^2} = \frac{1}{2\left(\sigma_\delta^2\right)^2}\left\{\sum_{i=1}^{n}(x_i - X_i)^2 + \frac{1}{\lambda}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2\right\}$$

$$2n\sigma_\delta^2 = \left\{\sum_{i=1}^{n}(x_i - X_i)^2 + \frac{1}{\lambda}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2\right\}$$

$$\hat{\sigma}_\delta^2 = \frac{1}{2n}\left\{\sum_{i=1}^{n}(x_i - X_i)^2 + \frac{1}{\lambda}\sum_{i=1}^{n}(y_i - \alpha - \beta X_i)^2\right\} \qquad (10)$$

VARIANCE–COVARIANCE MATRIX USING FISHER INFORMATION MATRIX

The Fisher Information matrix of parameters $\hat{\alpha}$ and $\hat{\beta}$ is used to calculate the variance and covariance of $\hat{\alpha}$ and $\hat{\beta}$.

The second partial derivative of equation (6) with respect to α is:

$$\frac{\delta^2}{\delta\alpha^2}(\log L) = -\frac{n}{\lambda\sigma_\delta^2}$$

Therefore, $E\left(-\frac{\delta^2}{\delta\alpha^2}(\log L)\right) = \frac{n}{\lambda\sigma_\delta^2}$.

The second partial derivative of equation (6) with respect to β is:

$$\frac{\delta^2}{\delta\beta^2}(\log L) = -\frac{1}{\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^2$$

Therefore, $E\left(-\frac{\delta^2}{\delta\beta^2}(\log L)\right) = \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^2$.

The second partial derivative of equation (6) with respect to α and β is:

$$\frac{\delta^2}{\delta\alpha\ \delta\beta}(\log L) = -\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i$$

Therefore, $E\left(-\frac{\delta^2}{\delta\alpha\ \delta\beta}(\log L)\right) = \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i$.

As a result, the estimated Fisher information matrix, F for $\hat{\alpha}$ and $\hat{\beta}$ is as follows

$$F = \begin{pmatrix} \frac{n}{\lambda\sigma_\delta^2} & \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i \\ \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i & \frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad (11)$$

where $a = \dfrac{n}{\lambda\sigma_\delta^2}$ is a 1×1 matrix, $b = \dfrac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i$ is a $1\times n$ matrix, $c = \dfrac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i$ is a $n\times1$ matrix, and $d = \dfrac{1}{\lambda\sigma^2}\sum_{i=1}^{n}\hat{X}_i^{\,2}$ is a $n\times n$ matrix. *a, b, c* and *d* are the negative expected value of the second partial derivatives for the log-likelihood function.

From the theory of partitioned matrices (Nelder 1977), the inverse of F is

$$F^{-1} = \begin{pmatrix} \left(a - bd^{-1}c\right)^{-1} & -a^{-1}b\left(d - ca^{-1}b\right)^{-1} \\ -d^{-1}c\left(a - bd^{-1}c\right)^{-1} & \left(d - ca^{-1}b\right)^{-1} \end{pmatrix} \tag{12}$$

Thus, the variance and covariance of $\hat{\alpha}$ and $\hat{\beta}$ are

$$V\hat{a}r(\hat{\alpha}) = \left(a - bd^{-1}c\right)^{-1}$$

$$= \left(\frac{n}{\lambda\sigma_\delta^2} - \left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^{\,2}\right)^{-1}\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\right)^{-1}$$

$$= \frac{\left(\lambda + \hat{\beta}^2\right)\hat{\sigma}_2^{\delta}\hat{\beta}}{S_{xy}}\left\{\overline{x}^2\left(1 + \hat{T}\right) + \frac{S_{xy}}{n\hat{\beta}}\right\} \tag{13}$$

$$V\hat{a}r(\hat{\beta}) = \left(d - ca^{-1}b\right)^{-1}$$

$$= \left(\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^{\,2}\right) - \left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\left(\frac{n}{\lambda\sigma_\delta^2}\right)^{-1}\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\right)^{-1}$$

$$= \frac{\left(\lambda + \hat{\beta}^2\right)\hat{\sigma}_2^{\delta}\hat{\beta}}{S_{xy}}\left\{1 + \hat{T}\right\} \tag{14}$$

Thus, the variance and covariance of $\hat{\alpha}$ and $\hat{\beta}$ are

$$C\hat{o}v(\hat{\alpha}, \hat{\beta}) = -d^{-1}c\left(a - bd^{-1}c\right)^{-1}$$

$$= -\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^{\,2}\right)^{-1}\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\left(\frac{n}{\lambda\sigma_\delta^2} - \left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\right.$$

$$\left.\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i^{\,2}\right)^{-1}\left(\frac{1}{\lambda\sigma_\delta^2}\sum_{i=1}^{n}\hat{X}_i\right)\right)^{-1}$$

$$= \frac{\left(\lambda + \hat{\beta}^2\right)\hat{\sigma}_2^{\delta}\hat{\beta}\overline{x}}{S_{xy}}\left\{1 + \hat{T}\right\} \tag{15}$$

SIMULATION STUDY FOR LINEAR FUNCTIONAL RELATIONSHIP MODEL (LFRM)

This model is evaluated through some bias measures to investigate the performance of the estimated parameters. The results from the bias measure would indicate the adequacy of the model's parameter estimates. Mean, estimated bias, and mean absolute percentage error (MAPE) are evaluated for the parameter estimates of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_\delta^2$. For simplicity, let $\hat{\theta}_p$ be the estimator of the parameter θ_p. θ_p and $\hat{\theta}_p$ represent the parameter and estimate values of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_\delta^2$, respectively; $(\overline{\theta})$ and $(\overline{\hat{\theta}})$ represent the mean values of the parameter and estimate values of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_\delta^2$.

a) Performance measures for model evaluation

Mean of $\hat{\theta}, \overline{\hat{\theta}} = \dfrac{1}{s}\sum_{p=1}^{s}\hat{\theta}_p$, Estimated bias, $EB = \overline{\hat{\theta}} - \theta$,

Mean absolute percentage error of $\hat{\theta}$, MAPE $(\hat{\theta})$

$$= \frac{1}{s}\sum_{p=1}^{s}\frac{\left|\theta_p - \hat{\theta}_p\right|}{\theta_p}$$

This general formula will be used to obtain bias measures of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_\delta^2$.

In this study, a Monte Carlo simulation is conducted to assess the accuracy and biasness of the estimation parameters in this model. The number of simulations is set to be s=10000, with the value of $\lambda = 1$. X_i has been generated from the normal distribution, and the true value of α is set to be 0.5 and 1, respectively (Ghapor et al. 2014). Next, the true values of β and σ_δ^2 are set to be equal to 1. The sample sizes in the simulation are n = 30, 50,100, 150, 200, 250 and 500.

The detail of the simulation process of LFRM in MATLAB *software* can be described as follows

Step 1: Generate random X_i of size n. Step 2: Generate the random error terms δ_i and ε_i. Step 3: Calculate the value of x_i and y_i. Step 4: Calculate the mean of x_i and y_i. Step 5: Calculate the parameter estimates $\hat{\alpha}$, $\hat{X}_i$, $\hat{\beta}$ and $\hat{\sigma}_\delta^2$. Step 6: Calculate the mean, estimated bias, estimated root mean square error, and mean absolute percentage error of $\hat{\alpha}$, $\hat{X}_i$, $\hat{\beta}$ and $\hat{\sigma}_\delta^2$.

APPLICATION TO REAL DATA

The proposed method's applicability is demonstrated using Pulau Langkawi's wind speed data throughout the southwest monsoon from the 18th of May to the 15th of September for 2019 and 2020. The data were obtained from the Malaysian Meteorological Department (2022). With a sample size (n) of 110, the variable x_i for i = 1,2,3,…, n represents Pulau Langkawi's wind speed data during the southwest monsoon in the year 2019; variable y_i for i = 1,2,3,…, n represents Pulau Langkawi's wind speed data throughout the southwest monsoon in 2020. This paper aims to examine the relationship between wind speed data for two consecutive years and present it as a bivariate functional relationship model for linear data.

The Weibull distribution is a statistical distribution frequently used to model wind speeds (Zaharim et al. 2009). The shape parameter of the Weibull distribution can be fitted using a normal distribution (Amirinia, Kamranzad & Mafi 2017; Kwon 2010). For example, Dookie et al. (2018) identified that normal distribution is suitable for evaluating wind speed in Trinidad and Tobago compared to distributions such as Weibull, Birbaum-Saunders, Exponential, Gamma, Nakagami, and Rayleigh distributions. Graphical comparisons were employed to assess the distributions, while the parameters were estimated using maximum likelihood estimation. It is found that the expected power predicted difference from the actual at Trinidad and Tobago by normal distribution is lower than the Weibull distribution. Therefore, normal distribution will be used to model the relationship between Pulau Langkawi's wind speed data throughout the southwest monsoon in 2019 and 2020.

The normality of the wind speed data is tested using the Kolmogorov-Smirnov test. Kolmogorov-Smirnov test is a well-known and widely used method to test whether the data is normally distributed (Zakaria 2022). The following are the null (H_0) and alternative hypotheses (H_A) used in a Kolmogorov-Smirnov test: H_0: The distribution of the data is normal. H_A: The distribution of the data is not normal.

The Kolmogorov-Smirnov statistic (D) is defined as

$$D = \max_{1 \le i \le n} \left(F(Y_i) - \frac{i-1}{n}, \frac{i}{n} - F(Y_i) \right) \tag{16}$$

where F is the theoretical cumulative distribution. H_0 is rejected if D exceeds the critical value determined from the table obtained by Lo Brano et al. (2011) and Massey (1951). The critical value is derived from the maximum absolute difference between sample and population cumulative distributions for a sample size n (Massey 1951). The equation critical value of D when $\alpha = 0.05$, and the sample size is over 35, is $\frac{1.36}{\sqrt{n}}$ based on Massey (1951) and Hawkins and Kanji (1995). Insert the value of the sample size, 110, into the equation critical value of $D = \frac{1.36}{\sqrt{110}}$. Hence, the critical value is 0.11255.

For Pulau Langkawi's wind speed data throughout the southwest monsoon in 2019 from equation (16), D is 0.103835. On the other hand, for Pulau Langkawi's wind speed data throughout the southwest monsoon in 2020, the D is 0.092321. Since D is below the critical value for both years; hence the H_0 cannot be rejected. This indicates that Pulau Langkawi's wind speed data throughout the southwest monsoon in the year 2019 and 2020 can be assumed to be normally distributed. Therefore, the proposed model in this study is normally distributed and can be used to describe the relationship between wind speed data in 2019 and 2020.

QQ-plots for wind speed data from both years are constructed to show the data's goodness-of-fit to the normal distribution. QQ-plots illustrate the data distribution. The points will fall on a reference line if the two data sets are from the normal distribution. The QQ-plots for wind speed data in the year 2019 and year 2020 are displayed in Figures 1 and 2, respectively.

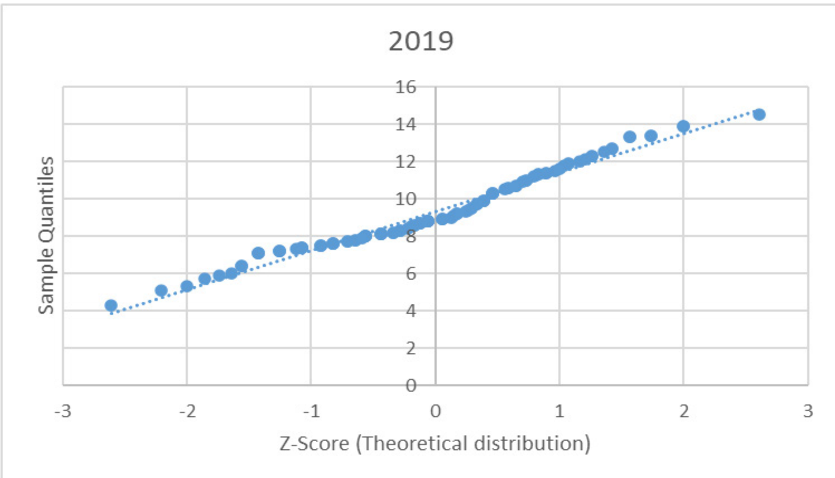

FIGURE 1. QQ-plots for wind speed data in 2019

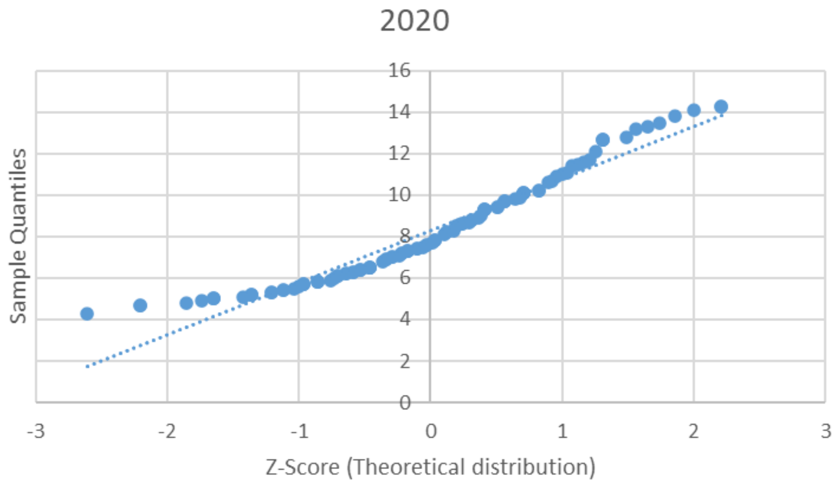

FIGURE 2. QQ-plots for wind speed data in 2020

The Kolmogorov-Smirnov test and the QQ-plots support that the wind speed data in the year 2019 and year 2020 will be treated with normal distribution. The steps in applying the wind speed data in Pulau Langkawi to LFRM are as follows: Step 1: Insert data Pulau Langkawi's wind speed data throughout the southwest monsoon in the year 2019 as x_i, in the year 2020 as y_i, and let $\lambda = 1$. Step 2: Calculate the mean of x_i and y_i. Step 3: Fit the data by using LFRM from Equation (6). Step 4: Calculate the parameter estimates $\hat{\alpha}$, $\hat{X}_i$, $\hat{\beta}$ and $\hat{\sigma}_\delta^2$. Step 5: Calculate the Var ($\hat{\alpha}$), Var ($\hat{\beta}$) and Cov ($\hat{\alpha}$, $\hat{\beta}$).

SIMULATION RESULTS AND DISCUSSION

From Tables 1 and 2, the mean of $\hat{\alpha}$ becomes closer to the real value of $\hat{\alpha}$ as we increase the value of n. As n increases, the estimate converges to the true value because the EB ($\hat{\alpha}$) approaches zero. MAPE ($\hat{\alpha}$) shows a modest decrease as n increases. Therefore, the estimation seems adequate for most values of λ and n.

From Tables 3 and 4, the mean of $\hat{\beta}$ becomes closer to the real value of $\hat{\beta}$ as we increase the value of n. The EB ($\hat{\beta}$) shows a modest decrease as n increases. The value of MAPE ($\hat{\beta}$) decreases. Therefore, the estimation seems adequate for most values of λ and n.

TABLE 1. Performance measurement for $\hat{\alpha}$, when $\alpha = 1$, $x_i = 1$ and $y_i = 1$

n	Mean ($\bar{\hat{\alpha}}$)	EB ($\hat{\alpha}$)	MAPE ($\hat{\alpha}$)
30	0.9907	-0.0093	0.2362
50	1.0017	0.0017	0.1657
100	1.0018	0.0018	0.1158
150	1.0011	0.0011	0.0933
200	1.0004	0.0004	0.0808
250	1.0003	0.0003	0.0719
500	1.0001	0.0001	0.0503

TABLE 2. Performance measurement for $\hat{\alpha}$, when α = 0.5, x_i = 1 and y_i = 1

n	Mean ($\bar{\hat{\alpha}}$)	EB ($\hat{\alpha}$)	MAPE ($\hat{\alpha}$)
30	0.5030	0.0030	0.4706
50	0.5034	0.0034	0.3342
100	0.5016	0.0016	0.2320
150	0.5011	0.0011	0.1867
200	0.5012	0.0012	0.1640
250	0.5007	0.0007	0.1448
500	0.5004	0.0004	0.1007

TABLE 3. Performance measurement for $\hat{\beta}$, when α =1, x_i =1 and y_i = 1

n	Mean ($\bar{\hat{\beta}}$)	EB ($\hat{\beta}$)	MAPE ($\hat{\beta}$)
30	1.0511	0.0511	0.3535
50	1.0339	0.0339	0.2165
100	1.0175	0.0175	0.1432
150	1.0131	0.0131	0.1165
200	1.0069	0.0069	0.0999
250	1.0077	0.0077	0.0889
500	1.0029	0.0029	0.0625

TABLE 4. Performance measurement for $\hat{\beta}$, when α = 0.5, x_i = 1 and y_i = 1

n	Mean ($\bar{\hat{\beta}}$)	EB ($\hat{\beta}$)	MAPE ($\hat{\beta}$)
30	1.1062	0.1062	0.3453
50	1.0418	0.0418	0.2180
100	1.0146	0.0146	0.1444
150	1.0116	0.0116	0.1176
200	1.0076	0.0076	0.0998
250	1.0078	0.0078	0.0883
500	1.0029	0.0029	0.0623

From Tables 5 and 6, the mean of $\hat{\sigma}_\delta^2$ becomes closer to the real value of $\hat{\sigma}_\delta^2$ as we increase the value of n. The EB ($\hat{\sigma}_\delta^2$) shows a modest decrease as n increases. MAPE of $\hat{\sigma}_\delta^2$ decline when n increases. Therefore, the estimation seems adequate for most values of λ and n.

Table 7 shows the parameter estimates for wind speed data collected in Pulau Langkawi, Kedah, during the 2019 and 2020 southwest monsoons when fitted with a functional relationship model for linear variables. The result shows that the y-intercept $\hat{\alpha}$ is -3.1005, the slope parameter $\hat{\beta}$ is 1.2221 and $\hat{\sigma}_\delta^2$ is 0.0699. It is worthwhile to note that the variance values of $\hat{\alpha}$ and $\hat{\beta}$ are small, where $V\hat{a}r\left(\hat{\alpha}\right)=0.03280$ and $V\hat{a}r\left(\hat{\beta}\right)=0.0004$, thus, implying the values are close to the mean. The model for Pulau Langkawi's wind speed data throughout the southwest monsoon in 2019 and 2020 is $Y=-3.1005+1.2221X$.

TABLE 5. Performance measurement for $\hat{\sigma}_\delta^2$, when α =1, x_i = 1 and y_i = 1

n	Mean ($\bar{\hat{\sigma}}_\delta^2$)	EB ($\hat{\sigma}_\delta^2$)	MAPE ($\hat{\sigma}_\delta^2$)
30	0.9801	-0.0199	0.2086
50	0.9888	-0.0112	0.1591
100	0.9950	-0.0050	0.1125
150	0.9955	-0.0045	0.0923
200	0.9977	-0.0023	0.0808
250	0.9968	-0.0032	0.0713
500	0.9999	-0.0001	0.0502

TABLE 6. Performance measurement for $\hat{\sigma}_\delta^2$, when α = 0.5, x_i = 1 and y_i = 1

n	Mean ($\bar{\hat{\sigma}}_\delta^2$)	EB ($\hat{\sigma}_\delta^2$)	MAPE ($\hat{\sigma}_\delta^2$)
30	0.9774	-0.0226	0.2067
50	0.9909	-0.0091	0.1611
100	0.9936	-0.0064	0.1139
150	0.9950	-0.0050	0.0917
200	0.9995	-0.0005	0.0802
250	0.9981	-0.0019	0.0704
500	0.9989	-0.0011	0.0500

TABLE 7. Parameter estimates of Pulau Langkawi, Kedah wind speed data

Detail	Value
$\hat{\alpha}$	-3.1005
$\hat{\beta}$	1.2221
σ_{δ}^{2}	0.0699
$V\hat{a}r(\hat{\alpha})$	0.0328
$V\hat{a}r(\hat{\beta})$	0.0004
Cov($\hat{\alpha}, \hat{\beta}$)	0

CONCLUSIONS

This paper employs the linear functional relationship model (LFRM) to explore the association between wind speed on Pulau Langkawi during the southwest monsoon season in 2019 and 2020. The functional relationship model assumes the presence of unobservable errors. The maximum likelihood method (MLE) is utilised to estimate all the parameters, and the covariance matrix of the estimated parameters is derived from the Fisher Information matrix. The parameter estimation demonstrates favourable accuracy and consistency based on the Monte Carlo simulation findings. As the sample size increases, the mean of predicted estimations converges closer to the true values, and the estimated bias tends to approach zero. Moreover, the mean absolute percentage error decreases with increasing sample size. Furthermore, the estimated parameters exhibit low variance, indicating consistent and less dispersed values. These results highlight the model's practicality and support its applicability in practical scenarios. The model obtained explained the relationship between Pulau Langkawi's wind speed data throughout the southwest monsoon in the year 2019 and year 2020 which is $Y = -3.1005 + 1.2221X$. The application of this model in practical settings could assist in the management of outdoor activities by considering weather conditions and safety aspects. By understanding the relationship between wind speeds across different years and describing it through a bivariate functional relationship model, future research can extend the model's use to various locations. It will enable a comprehensive approach to decision-making and planning, incorporating the influence of wind speeds on outdoor activities.

ACKNOWLEDGEMENTS

The authors express their gratitude to the Ministry of Higher Education, Malaysia for supporting this study through the Fundamental Research Grant Scheme FRGS/1/2021/STG06/UITM/02/5.

REFERENCES

Amirinia, G., Kamranzad, B. & Mafi, S. 2017. Wind and wave energy potential in southern Caspian Sea using uncertainty analysis. *Energy* 120: 332-345. https://doi.org/10.1016/j.energy.2016.11.088

Arif, A.M., Zubairi, Y.Z. & Hussin, A.G. 2022. Outlier detection in balanced replicated linear functional relationship model. *Sains Malaysiana* 51(2): 599-607. https://doi.org/10.17576/jsm-2022-5102-23

Arif, A.M., Zubairi, Y.Z. & Hussin, A.G. 2021. Maximum likelihood estimation of replicated linear functional relationship model. *Title of Journal* 10(1): 301-308.

Arif, A.M., Zubairi, Y.Z. & Hussin, A.G. 2020. Parameter estimation in replicated linear functional relationship model in the presence of outliers. *Malaysian Journal of Fundamental and Applied Sciences* 16(2): 158-160. https://doi.org/10.11113/mjfas.v16n2.1633

Arif, A.M., Zubairi, Y.G. & Hussin, A.G. 2019. On robust estimation for slope in linear functional relationship model. *Sains Malaysiana* 48(1): 237-242. https://doi.org/10.17576/jsm-2019-4801-27

Buonaccorsi, J.P. 1996. A modified estimating equation approach to correcting for measurement error in regression. *Biometrika* 83(2): 433-440. https://doi.org/10.1093/biomet/83.2.433

Chen, Z., Wang, X. & Wang, H. 2018. Preliminary research on total nitrogen content prediction of sandalwood using the error-in-variable models based on digital image processing. *PLoS ONE* 13(8): 1-22. https://doi.org/10.1371/journal.pone.0202649

Doganaksoy, N. & van Meer, H. 2015. An application of the linear errors-in-variables assessment an application of the linear errors-in- variables model in semiconductor device performance assessment. *Quality Engineering* 27(4): 500-511. https://doi.org/10.1080/08982112.2015.1068330

Dookie, I., Rocke, S., Singh, A. & Ramlal, C. 2018. Evaluating wind speed probability distribution models with a novel goodness of fit metric: A Trinidad and Tobago case study. *International Journal of Energy and Environmental Engineering* 9(3): 323-339. https://doi.org/10.1007/s40095-018-0271-y

Fuller, W. 1987. *Measurement Error Models*. New York: John Wiley.

Ghapor, A.A., Zubairi, Y.Z., Mamun, A.S.M.A. & Imon, A.H.M.R. 2015a. A robust method to estimate the slope parameter in linear functional relationship model. *Recent Advances in Renewable Energy Sources.* pp. 42-49.

Ghapor, A.A., Zubairi, Y.Z., Mamun, A.S.M.A. & Imon, A.H.M.R. 2015b. A robust nonparametric slope estimation in linear functional relationship model. *Pakistan Journal of Statistics* 48(1): 339-350. https://doi.org/10.17576/jsm-2019-4801-27

Ghapor, A.A., Zubairi, Y.Z., Mamun, A.S.M.A. & Imon, A.H.M.R. 2014. On detecting outlier in simple linear functional relationship model using COVRATIO statistic. *Pakistan Journal of Statistics* 30(1): 129-142.

Hanoon, M.S., Ahmed, A.N., Kumar, P., Razzaq, A., Zaini, N., Huang, Y.F., Sherif, M., Sefelnasr, A., Chau, K.W. & El-Shafie, A. 2022. Wind speed prediction over Malaysia using various machine learning models: Potential renewable energy source. *Engineering Applications of Computational Fluid Mechanics* 16(1): 1673-1689. https://doi.org/10.1080/19942060.2022.2103588

Hassan, S.F., Hussin, A.G. & Zubairi, Y.Z. 2010. Estimation of functional relationship model for circular variables and its application in measurement problems. *Chiang Mai Journal of Science* 37(2): 195-205.

Hawkins, D.I. & Kanji, G.K. 1995. 100 statistical tests. *Journal of Marketing Research* 32(1): 112. https://doi.org/10.2307/3152117

Kwon, S.D. 2010. Uncertainty analysis of wind energy potential assessment. *Applied Energy* 87(3): 856-865. https://doi.org/10.1016/j.apenergy.2009.08.038

Lindley, D.V. 1947. Regression lines and the linear functional relationship. *Supplement to the Journal of the Royal Statistical Society* 9(2): 218-244.

Lo Brano, V., Orioli, A., Ciulla, G. & Culotta, S. 2011. Quality of wind speed fitting distributions for the urban area of Palermo, Italy. *Renewable Energy* 36(3): 1026-1039. https://doi.org/10.1016/j.renene.2010.09.009

Malaysian Meteorological Department. 2022. Northeast Monsoon Report Kuala Lumpur Monsoon Activity May 2022. *MET Malaysia*. pp. 1-2. https://www.met.gov.my/home

Malaysian Meteorology Department. 2019. *Annual Report of Malaysian Meteorological Department 2019*. Malaysian Meteorological Deparment. www.met.gov.my

Massey, F.J. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46(253): 68. https://doi.org/10.2307/2280095

Mokhtar, N.A., Badyalina, B. & Zubairi, Y.Z. 2022. Functional model of wind direction data in Kuching, Sarawak, Malaysia. *Applied Mathematical Sciences* 16(7): 349-357.

Mokhtar, N.A., Badyalina, B., Chang, K.L., Ya'acob, F.F., Ghazali, A.F. & Shamala, P. 2021a. Error-in-variables model of Malacca wind direction data with the von Mises distribution in southwest monsoon. *Applied Mathematical Sciences* 15(9): 471-479. https://doi.org/10.12988/ams.2021.914521

Mokhtar, N.A., Zubairi, Y.Z., Hussin, A.G., Badyalina, B., Ghazali, A.F., Ya'Acob, F.F., Shamala, P. & Kerk, L.C. 2021b. 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* 1988(1): 012097. https://doi.org/10.1088/1742-6596/1988/1/012097

Nelder, J. 1977. Theory and application of the linear model. *Journal of the Royal Statistical Society. Series A (General)* 140(3): 384-385.

Rosenhead, J. 1963. The advanced theory of statistics: Volume 2: Inference and relationship. *J Oper. Res. Soc.* 14: 97-98. https://doi.org/10.2307/3006935

Zaharim, A., Najid, S.K., Razali, A.M. & Sopian, K. 2009. Analysing Malaysian wind speed data using statistical distribution. *Proceedings of the 4th IASME / WSEAS International Conference on Energy & Environment (EE'09)*. pp. 363-370. http://www.wseas.us/e-library/conferences/2009/cambridge/EE/EE55.pdf

Zakaria, M. 2022. The limitation of widely used data normality tests in clinical research. *Auditory and Vestibular Research* 31(1): 1-3. https://doi.org/10.18502/avr.v31i1.8127

*Corresponding author; email: nurkhairany@uitm.edu.my