Modelling Wind Speed Data in Pulau Langkawi With Functional Relationship
(Memodelkan Data Kelajuan Angin di Pulau Langkawi dengan Perhubungan Fungsian)

NUR AIN AL-HAMEEFATUL JAMALIYATUL1, BASRI BADYALINA1, NURKHAIRANY AMYRA MOKHTRAR1, ADZHAR RAMBL2, YONG ZULINA ZUBAIRI3 & ADILAH ABDUL GHAPOR4

1Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Johor Branch, Segamat Campus, 85000 Segamat, Johor, Malaysia.
2School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
3Institute of Advanced Studies, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
4Department 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

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)

$$Y = \alpha + \beta X + \epsilon$$ (1)
where \( X \) is the explanatory variable; \( Y \) is the dependent variable; \( \alpha \) is the value of \( Y \) intercept; \( \beta \) is the slope of the linear regression model; and \( \varepsilon \) 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 categorized 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 \( \lambda \) 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 \( X \) and \( Y \) are linearly connected, identified with error, and can be represented by the equation

\[
Y_i = \alpha + \beta X_i \quad \text{for } i = 1, 2, 3, \ldots, n \tag{2}
\]

with \( \alpha \) and \( \beta \) are the intercept value and slope of the model, respectively. In the LFRM, \( X_i \) and \( Y_i \) variables are subject to random errors, \( \delta_i \) and \( \varepsilon_i \) for \( i = 1, 2, 3, \ldots, n \) which \( n \) is the number of parameter. The error terms \( \delta_i \) and \( \varepsilon_i \) are assumed to be mutually independent random variables with normally distributed distributions.

\[
x_i = X_i + \delta_i \quad \text{and} \quad y_i = Y_i + \varepsilon_i \tag{3}
\]

\[
\delta_i \sim N(0, \sigma_{\delta_i}^2) \quad \text{and} \quad \varepsilon_i \sim N(0, \sigma_{\varepsilon_i}^2) \tag{4}
\]
According to Arif, Zubairi and Hussin (2020), \((n + 4)\) parameters must be determined in LFRM, which are \(\alpha, \beta, \sigma^2_\alpha, \sigma^2_\beta\), and the incidental parameters \(X_1, X_2, ..., X_n\). The log-likelihood function is given by

\[
\log L = -n \log(2\pi) - \frac{n}{2} \left( \log \sigma^2_\alpha + \log \sigma^2_\beta \right) - \frac{1}{2\sigma_\alpha^2} \sum_{i=1}^n (x_i - \bar{X})^2 - \frac{1}{2\sigma_\beta^2} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X})^2
\]  

(5)

According to Ghapor et al. (2014), when \(\hat{X}_i = x_i\) and \(\sigma_\beta^2\) 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 \(\lambda = \sigma^2_\alpha / \sigma^2_\beta\), where \(\lambda\) is known. Now, there are \((n + 3)\) parameters to be estimated, namely \(\alpha, \beta, \sigma^2_\alpha, \sigma^2_\beta\), and \(X_1, ..., X_n\) (Fuller 1987; Rosenhead 1963). The log-likelihood function is given by

\[
\log L = -n \log(2\pi) - \frac{n}{2} \log \lambda - n \log \sigma^2_\delta - \frac{1}{2\lambda \sigma^2_\delta} \left\{ \sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X})^2 \right\}
\]  

(6)

The LFRM parameters will be determined using the MLE approach.

a) MLE approach for \(\alpha\)

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

\[
\frac{\delta}{\delta \alpha} (\log L) = -\frac{1}{2\lambda \sigma^2_\delta} \left\{ (2)(-1) \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X}) \right\}
\]

\[
\frac{\delta}{\delta \alpha} (\log L) = \frac{1}{\lambda \sigma^2_\delta} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X})
\]

and by setting \(\delta / \delta \alpha (\log L) = 0\)

\[
\frac{1}{\lambda \sigma^2_\delta} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X}) = 0
\]

\[
\frac{1}{\sigma^2_\delta} \left( \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\alpha} - \sum_{i=1}^n \hat{\beta} \bar{X} \right) = 0
\]

and by setting

\[
\sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\alpha} - \sum_{i=1}^n \hat{\beta} \bar{X} = 0
\]

\[
\frac{\sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\alpha} - \sum_{i=1}^n \hat{\beta} \bar{X}}{\sum_{i=1}^n \bar{X}} = \frac{\sum_{i=1}^n \hat{X}}{\sum_{i=1}^n \bar{X}} = \frac{\sum_{i=1}^n \hat{X}}{\sum_{i=1}^n \bar{X}} \left( \frac{1}{\frac{\hat{\beta}}{\lambda}} - \frac{\hat{\beta}^2}{\lambda^2} \right)
\]

b) MLE approach for \(X_i\)

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

\[
\frac{\delta}{\delta X_i} (\log L) = -\frac{1}{2\lambda \sigma^2_\delta} \sum_{i=1}^n (2)(-1)(x_i - \hat{X}) + \frac{1}{\lambda \sigma^2_\delta} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X})
\]

\[
\frac{\delta}{\delta X_i} (\log L) = \frac{1}{\lambda \sigma^2_\delta} \sum_{i=1}^n (x_i - \hat{X}) + \frac{\hat{\beta}}{\lambda} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} \bar{X})
\]

and by setting \(\delta / \delta X_i (\log L) = 0\)

\[
\sum_{i=1}^n x_i - \sum_{i=1}^n \hat{X} + \frac{\hat{\beta}}{\lambda} \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\alpha} - \sum_{i=1}^n \hat{\beta} \bar{X} = 0
\]

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

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

\[ \hat{X}_i = \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{(\lambda + \beta^2)} \]  

(8)

c) MLE approach for \( \beta \)

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

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

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

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

\[ \frac{1}{\lambda \sigma^2} \sum_{i=1}^{n} \hat{X}_i(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i) = 0 \]

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

\[ \hat{\beta} = \frac{\sum_{i=1}^{n} \hat{X}_i (y_i - \hat{\alpha})}{\sum_{i=1}^{n} \hat{X}_i^2} \]  

(9)

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

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

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

\[ \hat{\beta} \sum_{i=1}^{n} \left( \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{(\lambda + \beta^2)} \right)^2 = \sum_{i=1}^{n} \left( \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{(\lambda + \beta^2)} \right)(y_i - \hat{\alpha}) \]
Replacing \( \hat{\alpha} = \bar{y} - \bar{x} \hat{\beta} \),

\[
\hat{\beta} \sum_{i=1}^{n} x_i^2 + (\hat{\beta}^2 - \lambda) \sum_{i=1}^{n} x_i (y_i - (\bar{y} - \bar{x} \hat{\beta})) + \hat{\beta} \sum_{i=1}^{n} (y_i - (\bar{y} - \bar{x} \hat{\beta})) = 0
\]

\[
\hat{\beta} \sum_{i=1}^{n} x_i^2 + \hat{\beta} \sum_{i=1}^{n} x_i (y_i - (\bar{y} - \bar{x} \hat{\beta})) - \lambda \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \lambda \sum_{i=1}^{n} x_i (y_i - \bar{y}) = 0
\]

\[
-\lambda \hat{\beta} \sum_{i=1}^{n} x_i (y_i - \bar{y}) - \lambda \sum_{i=1}^{n} x_i (y_i - \bar{y}) - 2 \hat{\beta} \hat{\beta} \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0
\]

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

\[
- \lambda \sum_{i=1}^{n} x_i (y_i - \bar{y})^2 - \lambda \sum_{i=1}^{n} x_i (y_i - \bar{y})^2 = 0
\]

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

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

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

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

d) MLE approach for \( \sigma^2 \)

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

\[
\frac{\delta^2}{\delta \sigma^2} (\log L) = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4} \left\{ \sum_{i=1}^{n} (x_i - X_i)^2 \right\}
\]

\[
+ \frac{1}{\lambda} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2
\]

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

\[
- \frac{n}{\sigma^2} + \frac{1}{2\sigma^4} \left\{ \sum_{i=1}^{n} (x_i - X_i)^2 + \frac{1}{\lambda} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2 \right\} = 0
\]

\[
n \sigma^2 = \frac{1}{2\lambda} \left\{ \sum_{i=1}^{n} (x_i - X_i)^2 + \frac{1}{\lambda} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2 \right\}
\]

\[
2\lambda \sigma^2 = \left\{ \sum_{i=1}^{n} (x_i - X_i)^2 + \frac{1}{\lambda} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2 \right\}
\]

\[
\sigma^2 = \frac{1}{2n} \left\{ \sum_{i=1}^{n} (x_i - X_i)^2 + \frac{1}{\lambda} \sum_{i=1}^{n} (y_i - \alpha - \beta X_i)^2 \right\}
\]

\[
(10)
\]

**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 \( \alpha \) is:

\[
\frac{\delta^2}{\delta \alpha^2} (\log L) = -\frac{n}{\lambda \sigma^2}
\]

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

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

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

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

The second partial derivative of equation (6) with respect to \( \alpha \) and \( \beta \) is:

\[
\frac{\delta^2}{\delta \alpha \delta \beta} (\log L) = -\frac{1}{\lambda \sigma^2} \sum_{i=1}^{n} \hat{X}_i \hat{Y}_i
\]

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

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

\[
F = \begin{pmatrix}
\frac{n}{\lambda \sigma^2} & \frac{1}{\lambda \sigma^2} \sum_{i=1}^{n} \hat{X}_i \\
\frac{1}{\lambda \sigma^2} \sum_{i=1}^{n} \hat{X}_i & \frac{1}{\lambda \sigma^2} \sum_{i=1}^{n} \hat{X}_i^2
\end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\]

(11)
where $a = \frac{n}{\lambda \sigma^2_\alpha}$ is a $1 \times 1$ matrix, $b = \frac{1}{\lambda \sigma^2_\beta}$ is a $1 \times n$ matrix, $c = \frac{1}{\lambda \sigma^2_\gamma}$ is a $n \times 1$ matrix, and $d = \frac{1}{\lambda \sigma^2_\delta}$ is a $n \times n$ matrix. $a$, $b$, $c$ and $d$ are the partial derivatives for the log-likelihood function.

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

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

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

$$\text{Var}(\hat{\alpha}) = (a - bd^{-1}c)^{-1}$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -d^{-1}c(a - bd^{-1}c)^{-1}$$

$$\text{Var}(\hat{\beta}) = (d - ca^{-1}b)^{-1}$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -d^{-1}c(a - bd^{-1}c)^{-1}$$

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

$$\text{Var}(\hat{\alpha}) = (a - bd^{-1}c)^{-1}$$

$$\text{Var}(\hat{\beta}) = (d - ca^{-1}b)^{-1}$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -d^{-1}c(a - bd^{-1}c)^{-1}$$

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

$$\text{Var}(\hat{\alpha}) = (a - bd^{-1}c)^{-1}$$

$$\text{Var}(\hat{\beta}) = (d - ca^{-1}b)^{-1}$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -d^{-1}c(a - bd^{-1}c)^{-1}$$

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

$$\text{Var}(\hat{\alpha}) = (a - bd^{-1}c)^{-1}$$

$$\text{Var}(\hat{\beta}) = (d - ca^{-1}b)^{-1}$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -d^{-1}c(a - bd^{-1}c)^{-1}$$

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 $\hat{\sigma}^2_\delta$. For simplicity, let $\hat{\theta}$ be the estimator of the parameter $\theta$. $\widetilde{\theta}$ and $\hat{\theta}$ represent the parameter and estimate values of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}^2_\delta$, respectively; $(\bar{\theta})$ and $(\hat{\theta})$ represent the mean values of the parameter and estimate values of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}^2_\delta$.

a) Performance measures for model evaluation

Mean of $\hat{\theta}$, $\tilde{\theta} = \frac{1}{s}\sum_{i=1}^{s}\hat{\theta}_i$. Estimated bias, $EB = \tilde{\theta} - \theta$.

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

$$= \frac{1}{s}\sum_{i=1}^{s}\left|\frac{\hat{\theta}_i - \theta}{\theta}\right|$$

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

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 $\lambda = 1$. $X$ has been generated from the normal distribution, and the true value of $\alpha$ is set to be 0.5 and 1, respectively (Ghapor et al. 2014). Next, the true values of $\beta$ and $\sigma^2_\delta$ 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 $X$ of size $n$. Step 2: Generate the random error terms $\delta_i$ and $\epsilon_i$. 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 $\hat{\alpha}, \hat{X}, \hat{\beta}$ and $\hat{\sigma}^2_\delta$. Step 6: Calculate the mean, estimated bias, estimated root mean square error, and mean absolute percentage error of $\hat{\alpha}, \hat{X}, \hat{\beta}$ and $\hat{\sigma}^2_\delta$.

**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 $i = 1, 2, 3, \ldots$, represents Pulau Langkawi’s wind speed data during the southwest monsoon in the year 2019; variable $y_i$ for $i = 1, 2, 3, \ldots$, 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

$$D = \max_{i=1}^{n} \left( F(Y_i) - \frac{i-0.5}{n} \right)$$

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.
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 $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_\epsilon^2$. 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 $\lambda$ 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 $\lambda$ and $n$.

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

<table>
<thead>
<tr>
<th>$n$</th>
<th>Mean ($\tilde{\alpha}$)</th>
<th>EB ($\hat{\alpha}$)</th>
<th>MAPE ($\hat{\alpha}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.9907</td>
<td>-0.0093</td>
<td>0.2362</td>
</tr>
<tr>
<td>50</td>
<td>1.0017</td>
<td>0.0017</td>
<td>0.1657</td>
</tr>
<tr>
<td>100</td>
<td>1.0018</td>
<td>0.0018</td>
<td>0.1158</td>
</tr>
<tr>
<td>150</td>
<td>1.0011</td>
<td>0.0011</td>
<td>0.0933</td>
</tr>
<tr>
<td>200</td>
<td>1.0004</td>
<td>0.0004</td>
<td>0.0808</td>
</tr>
<tr>
<td>250</td>
<td>1.0003</td>
<td>0.0003</td>
<td>0.0719</td>
</tr>
<tr>
<td>500</td>
<td>1.0001</td>
<td>0.0001</td>
<td>0.0503</td>
</tr>
</tbody>
</table>
### TABLE 2. Performance measurement for $\hat{\alpha}$, when $\alpha = 0.5$, $x_i = 1$ and $y_i = 1$

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

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

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

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

<table>
<thead>
<tr>
<th>$n$</th>
<th>Mean ($\hat{\beta}$)</th>
<th>EB ($\hat{\beta}$)</th>
<th>MAPE ($\hat{\beta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.1062</td>
<td>0.1062</td>
<td>0.3453</td>
</tr>
<tr>
<td>50</td>
<td>1.0418</td>
<td>0.0418</td>
<td>0.2180</td>
</tr>
<tr>
<td>100</td>
<td>1.0146</td>
<td>0.0146</td>
<td>0.1444</td>
</tr>
<tr>
<td>150</td>
<td>1.0116</td>
<td>0.0116</td>
<td>0.1176</td>
</tr>
<tr>
<td>200</td>
<td>1.0076</td>
<td>0.0076</td>
<td>0.0998</td>
</tr>
<tr>
<td>250</td>
<td>1.0078</td>
<td>0.0078</td>
<td>0.0883</td>
</tr>
<tr>
<td>500</td>
<td>1.0029</td>
<td>0.0029</td>
<td>0.0623</td>
</tr>
</tbody>
</table>
From Tables 5 and 6, the mean of $\hat{\sigma}^2_\delta$ becomes closer to the real value of $\delta^2_{\hat{\sigma}}$ as we increase the value of $n$. The EB ($\hat{\sigma}^2_{\delta}$) shows a modest decrease as $n$ increases. MAPE of $\hat{\sigma}^2_{\delta}$ decline when $n$ increases. Therefore, the estimation seems adequate for most values of $\lambda$ 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 $\hat{\sigma}^2_\delta$ is 0.0699. It is worthwhile to note that the variance values of $\hat{\alpha}$ and $\hat{\beta}$ are small, where $\hat{\text{Var}}(\hat{\alpha}) = 0.0328$ and $\hat{\text{Var}}(\hat{\beta}) = 0.0004$, 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 $\hat{\sigma}^2_\delta$, when $\alpha = 1$, $x_i = 1$ and $y_i = 1$

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

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

<table>
<thead>
<tr>
<th>$n$</th>
<th>Mean ($\hat{\sigma}^2_\delta$)</th>
<th>EB ($\hat{\sigma}^2_\delta$)</th>
<th>MAPE ($\hat{\sigma}^2_\delta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.9774</td>
<td>-0.0226</td>
<td>0.2067</td>
</tr>
<tr>
<td>50</td>
<td>0.9909</td>
<td>-0.0091</td>
<td>0.1611</td>
</tr>
<tr>
<td>100</td>
<td>0.9936</td>
<td>-0.0064</td>
<td>0.1139</td>
</tr>
<tr>
<td>150</td>
<td>0.9950</td>
<td>-0.0050</td>
<td>0.0917</td>
</tr>
<tr>
<td>200</td>
<td>0.9995</td>
<td>-0.0005</td>
<td>0.0802</td>
</tr>
<tr>
<td>250</td>
<td>0.9981</td>
<td>-0.0019</td>
<td>0.0704</td>
</tr>
<tr>
<td>500</td>
<td>0.9989</td>
<td>-0.0011</td>
<td>0.0500</td>
</tr>
</tbody>
</table>
## TABLE 7. Parameter estimates of Pulau Langkawi, Kedah wind speed data

<table>
<thead>
<tr>
<th>Detail</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>-3.1005</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>1.2221</td>
</tr>
<tr>
<td>$\sigma^2_\beta$</td>
<td>0.0699</td>
</tr>
<tr>
<td>$\text{Var}(\hat{\alpha})$</td>
<td>0.0328</td>
</tr>
<tr>
<td>$\text{Var}(\hat{\beta})$</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\text{Cov}(\hat{\alpha}, \hat{\beta})$</td>
<td>0</td>
</tr>
</tbody>
</table>

## 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


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