Jurnal Teknologi

Full Paper

RAINFALL VARIABILITY PREDICTION MODEL IN KUALA TERENGGANU, MALAYSIA USING PRINCIPAL COMPONENT ANALYSIS AND MULTIPLE LINEAR REGRESSION

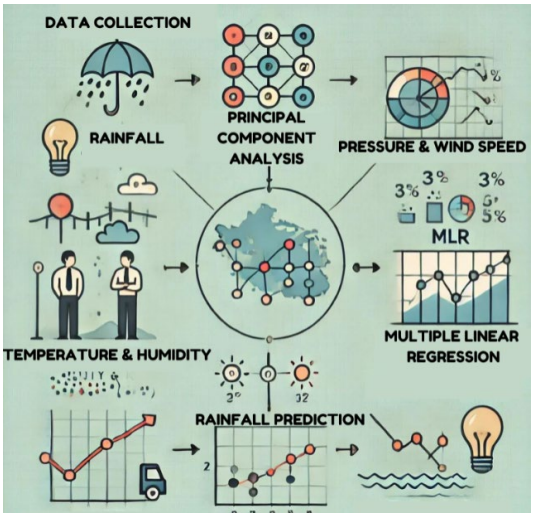
Yinqiu Wang[a], Mohd Saiful Samsudin[a*], Shazlyn Millenana Saharuddin[b], Weijian Chen[a], Demus Matheus Huang[a], Mohd Hafiidz Jaafar[a]

[a]Environmental Technology Division, School of Industrial Technology, Universiti Sains Malaysia, Penang 11800, Malaysia
[b]Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia

Article history
Received
13 August 2024
Received in revised form
28 April 2025
Accepted
16 June 2025
Published Online
24 December 2025

*Corresponding author
saifulsamsudin@usm.my

Graphical abstract



Abstract

This study addresses the impact of climate change on Kuala Terengganu, Malaysia, focusing on rainfall variability prediction. As extreme weather events become more frequent, accurate climate forecasts are essential for effective disaster preparedness. The primary objective is to evaluate the effectiveness of Principal Component Analysis (PCA) and Multiple Linear Regression (MLR) in predicting key climate indicators such as temperature, humidity, and precipitation. Using 2021 climate data, PCA was employed to identify significant variables influencing rainfall, which were then used in an MLR model to predict rainfall variability. The integrated PCA-MLR approach significantly improved prediction accuracy compared to MLR alone, identifying temperature, humidity, and wind speed as critical predictors. The study demonstrates that combining PCA and MLR enhances climate prediction accuracy, aiding better planning and response to climate challenges in Kuala Terengganu. This approach can improve disaster risk management and resilience. Future research should expand datasets and incorporate additional climate variables to refine predictive capabilities.

Keywords: Climate Change, Principal Component Analysis, Multiple Linear Regression, Rainfall Prediction Model

Abstrak

Kajian ini membincangkan kesan perubahan iklim di Kuala Terengganu, Malaysia, memberi tumpuan kepada ramalan kebolehubahan hujan. Apabila peristiwa cuaca ekstrem menjadi lebih kerap, ramalan iklim yang tepat adalah penting untuk kesiapsiagaan bencana yang berkesan. Objektif utama adalah untuk menilai keberkesanan Analisis Komponen Utama (PCA) dan Regresi Linear Berganda (MLR) dalam meramalkan penunjuk iklim utama seperti suhu, kelembapan dan kerpasan. Menggunakan data iklim 2021, PCA telah digunakan untuk mengenal pasti pembolehubah penting yang mempengaruhi hujan, yang kemudiannya digunakan dalam model MLR untuk meramalkan kebolehubahan hujan. Pendekatan PCA-MLR bersepadu meningkatkan ketepatan ramalan dengan ketara berbanding MLR sahaja, mengenal pasti suhu, kelembapan dan kelajuan angin sebagai peramal kritikal. Kajian ini menunjukkan bahawa gabungan PCA dan MLR meningkatkan

ketepatan ramalan iklim, membantu perancangan dan tindak balas yang lebih baik terhadap cabaran iklim di Kuala Terengganu. Pendekatan ini boleh meningkatkan pengurusan dan daya tahan risiko bencana. Penyelidikan masa depan harus mengembangkan set data dan menggabungkan pembolehubah iklim tambahan untuk memperhalusi keupayaan ramalan.

Kata kunci: Perubahan Iklim, Analisis Komponen Utama, Regresi Linear Berganda, Model Ramalan Hujan

1.0 INTRODUCTION

Amidst the persistence of climate change, the greenhouse gas emissions caused by heavy industrialization have promptly promoted the global temperature rise and caused a serious change in various climate parameters [1]. Global human population has been persistently threatened by several climate change phenomena with various magnitudes over the time [2]. According to the 2020 Climate Services Status Report of the World Meteorological Organization (WMO), more than 108 million people in the world have been significantly affected by storms, floods, droughts and forest fires as the results of persistent climate change, with the aforementioned number is predicted to experience a probable 1.5-fold increase by 2030 with a potential loss of approximately USD 20 billion per annum [3]. National Oceanic and Atmospheric Administration (NOAA) National Hurricane Center have also reported the record-breaking occurrence of 30 named tropical storms across Atlantic Ocean in 2020, with 11 of them were confirmed to hit the continental United States and cause severe hurricane rainfall within the affected areas [4]. By December 2020, NOAA National Centers for Environmental Information (NCEI) reported the economic loss caused by those 11 named storms to be approximately USD 41.3 billion [5]. On the other hand, Australia experienced unprecedented bushfires in 2019-2020, which had destroyed 20% of its forests and killed millions of wildlife [6].

Nearly half of the world's population resides within 100 km of the coast [7]. Coastal cities have developed economies, rapid urbanization and dense populations. Rising temperature, frequent heavy rainfall, and rising sea levels significantly contribute to the global warming trend, and human activities continue to influence the structural and morphological characteristics of coastal urban belts [8]. Increasing the resilience of coastal areas to abnormal climate change is thus critical to ensure their long-term sustainability. It is necessary to conduct a comprehensive assessment of the impact of disasters on coastal areas [9].

The potential impact of climate change is enormous and thus cannot be ignored. For instance, the sea temperature change of 0.5 ℃ can trigger strong air-sea interaction [10], which may lead to pronounced climate anomalies in many parts of the world [11]. On the other hand, by the end of 2024, the global mean sea level has significantly risen by at least 260 mm since 1880 by the rate of approximately 1.7 mm/year [12]. Within timespan from 2006 to 2015, the rate of global sea level rise has accelerated to approximately 3.6 mm/year (2.5-fold of the rate in the past 20th century), and the global mean sea level is estimated to reach at least 800 mm (almost 1 m) by the end of 21st century [12], [13].

The climate risk in Malaysia is moderate. Global warming leads to an annual increase in sea level on the peninsula of 1.3-9.4 mm [14]. National Center for Education Statistics (NCES) found that 28% of the coastline (approximately 1,360 km) of the peninsula is threatened by erosion, which may greatly affect the regional socio-economic activities [15]. This foreshadows possible future adverse effects and threats to Malaysia [16]. This will hinder the development and protection of coastal cities [17].

Floods occur in Terengganu almost every year, mainly due to the northeast monsoon. In 2014, massive floods hit large parts of Terengganu, causing 5,550 people to lose their homes. The monthly rainfall during that period exceeded 1,200 mm, which is equivalent to the annual rainfall. The government has invested MYR 132 million (approximately USD 28.8 million, based on an exchange rate of 1 USD = 4.58 MYR as of 11th April 2025) to repair the flood damage [18].

Meteorological factors describe the various factors that affect weather conditions at a particular time and place. Climate parameters are important statistical data that describe the long-term climate characteristics of a region, while climate itself is defined as a comprehensive reflection of the average condition and amplitude of changes in meteorological variables from several months to millions of years [19]. Various climate parameters play a key role in climate systems, water resources planning, agriculture, and biological systems [20], [21].

There are two main approaches to climate prediction. One is the dynamic numerical model prediction based on the physical law of partial

differential equations by adding initial conditions and boundary conditions. Dynamic models can simulate nonlinear relationships between phenomena and predict each event and its different effects. So far, a large number of weather and climate prediction systems have been developed by major countries in the world [22]. However, there are two major flaws in pure dynamic model prediction. First, the research and development of numerical models often needs to consume a lot of resources, and the improvement of model performance is very difficult. Secondly, there is a variety of obvious systematic model biases and insufficient spatial resolution, which limit the prediction skills [23], [24]. Moreover, the capability of numerical model prediction also depends on the initial field, that is, the quality of the data assimilation method and the construction of the ensemble prediction scheme. In particular, climate prediction often needs to develop a better-couple data assimilation scheme and fully consider the uncertainties of the observed data and the physical processes of each component model, which makes the construction of the whole prediction system very complicated [25], [26]. The other is a statistical method based on the linear relationship between prediction objects and predictors. For example, Multiple Linear Regression (MLR) [27], Principal Component Analysis (PCA) [28], Singular Value Decomposition (SVD) [29]. Most of these statistical forecasting relationships do not change with time, the calculation is relatively simple, and the model established according to historical data and the current actual state can quickly predict the future. The continuity, data density, and multidimensional dynamics of weather pose significant challenges to weather forecasting.

This study will strive to predict and interpret the atmospheric conditions in various regions. Its main uses include the protection of life and property, agricultural production, public utility planning, and daily life, such as food, clothing, housing and transportation. To more effectively adapt to and respond to the constantly changing environment and climate it is necessary to process climate data and predict the trend of climate parameters. Chemometric techniques have been proven to be a functional tool with simpler and easier-to-interpret results [30]. It can reduce data complexity and understand data better. Multiple linear regression has a strong predictive ability for data complexity, good performance, and is suitable for prediction. Therefore, this study analyzed climate parameters in Kuala Terengganu using chemometrics technology and predicted rainfall using multiple linear regression. The objective of this study is to analyze the climate observation data of Kuala Terengganu Weather Station in 2021, using the principal component analysis, analyze to develop the rain variability prediction model using MLR model.

2.0 METHODOLOGY

Figure 1 shows the sampling site location in Kuala Terengganu, Terengganu, Malaysia, which is located in the eastern part of the peninsula, bordering Kelantan and Pahang [31]. Its coastline is 320 km along the South China Sea. Located between latitude 5 ° 27 'to 5 ° 11'N and longitude 102 ° 57' to 103 ° 13'E. The total area is 13,035 km^2 [18], including eight districts. Most residents live in coastal towns. Kuala Terengganu is the capital city as well as the largest town in the state of Terengganu, located at the mouth of the Terengganu River, with an area of approximately 605 km^3. It faces the South China Sea and has a predominantly northeast monsoon climate with semi-diurnal tides [32]. The sample data in this study were collected from Jabatan Meteorologi Malaysia Station (5° 23'N,103° 06'E) at an elevation of 5.2 m, code name "48618". The sampling site is located at the boundary of the residential area, 250 m to the southeast is the airstrip of the ADMAL Flying Academy, 1,500 m to the northeast is the sea, and 2.5 km to the west is farmland.

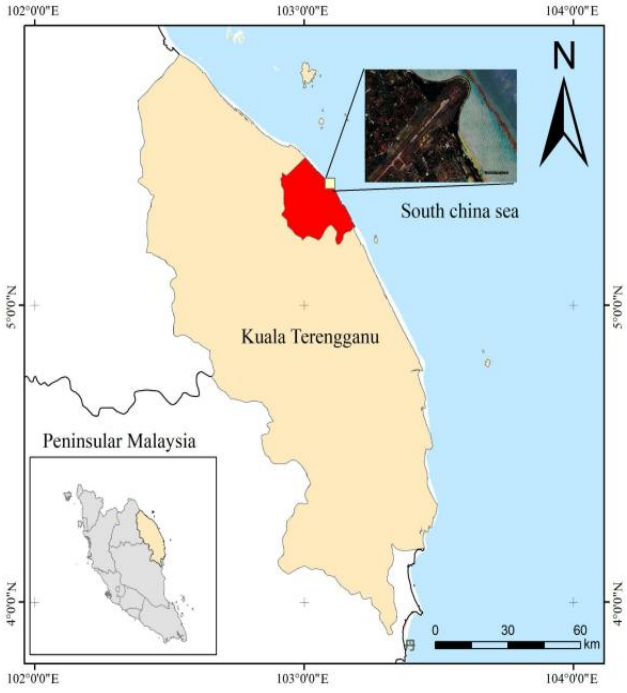


Figure 1 Map of sampling site location

The data source is from 1st January to 31st December 2021. The five climate parameters involved are: 24-hour average mean sea level (MSL) or atmospheric pressure, average temperature, average relative humidity, average wind speed and rainfall from 8 a.m. of the day before to 8 a.m. of the next day.

2.1 Pre-processing Data

Preliminary data manipulation involved gathering and transforming data within the matrix. Values below the detection limit were standardized to half the detection limit. Normality tests, specifically the Shapiro-Wilk test (W test), were conducted to assess the conformity of climate parameter distributions to a normal distribution [33], [34]. Variables exhibiting a normal distribution were subjected to transformation. The data pre-processing followed the methodology outlined in Equation (1) [35], [36]:

$$z_{ij} = \frac{(x_{ij} - \mu)}{\sigma} \quad (1)$$

Here, z_{ij} is the *j*-th value of the standard score of the measured variable *i*, x_{ij} is the *j*-th observation of the variable *i*, μ is the mean value of the variable, and σ is the standard deviation.

The analysis results will be mainly influenced by the maximum amplitude variable [35]. In addition, these transformations help to normalize the dispersion of the distribution and reduce potential classification errors caused by different variable sizes of the group or class [35], [37].

2.2 Box Plot

The box plot, also known as the box-and-whisker plot [38], is a data graph that represents descriptive statistics of a data set. In this study, boxplots of the discriminant parameters were created to assess the different trends of the changes in different parameters (Figure 2). In addition, box plots can help to understand the distribution characteristics and seasonal changes of climate data. The star or asterisks are outliers that signify cases with values more than three times the height of the boxes [39]. The red "+" in the box plot indicates the mean value of the data.

2.3 Principal Component Analysis (PCA)

Principal component analysis has been used as an efficient data downsizing technique proposed by Karl Pearson in 1901. This is a powerful data analysis tool [40] and is usually used to simplify the features of large datasets to a small amount of information rich components. PCA is transformed by converting the observed value of the set of correlation variables into a new variable [41] having a non-linear correlation representing the linear combination of the original variables [39], [42]. PC provides the most meaningful information about the parameters that describe the entire dataset [39].

PCA is used to overcome the redundancy of datasets, obtain valuable information, provide optimal decision-making, describe high contrast to achieve optimal data visualization, reduce complexity, and improve computational efficiency. The purpose is to extract important information from the data and display it in an index compilation [43].

PCA has been widely applied in several fields, such as neuroscience [44], finance [45], facial recognition [46], and environmental monitoring [47]. The principal components (PCs) can define as in Equation (2):

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \ldots + a_{im}x_{mj} \quad (2)$$

Here, z is the component score, *a* is the component loading, x is the measured value of the variable, *i* is the component number, *j* is the sample number and *m* is the total number of variables.

Variational rotation is commonly used to overcome complex interpretation of PCs [48]. When the eigenvalue is greater than 1, a variational rotation of the PC is required to obtain a new set of variables termed as variational factors (VFs). The number of VF generated by maximum variable rotation is equivalent to the number of variables, covering unobservable, hypothetical, and latent variables [49]. The VF coefficient is divided into "strong" (> 0.75), "moderate" (0.50–0.75), and "weak" (0.30–0.49) based on the correlation coefficient [50]. This study used XLSTAT 2019 to calculate PCs.

2.4 Pearson Correlation Coefficient

The Pearson correlation coefficient or *r* value was utilized to identify the statistical linear correlations among the five climate-related parameters in the study location [51]. In general, a sample *r* value between two variables *x* and *y* was mathematically defined as in Equation (3):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (3)$$

Here, r_{xy} represented Pearson correlation coefficient between respective variables x and *y*, *n* was number or size of data points, *i* was index of summation, x_i and y_i represented individual data points of respective variables x and *y*, $\bar{x}$ and $\bar{y}$ were means of respective variables x and *y*.

The correlations between the five climate parameters were visualized using the Pearson correlation map in Blue-Red scales. The *r* values ranged from –1.00 to +1.00, with a stronger correlation (darker color) indicated by the *r* values closer to ±1.00. In contrast, a weaker correlation (lighter color in the correlation maps) was indicated by the *r* values closer to 0.00. The direction of the correlations was denoted by either the positive (red) or negative sign (blue) of the *r* values. Conjointly, a *t*-test with a 95% confidence level (a = 0.05) was conducted to test the statistical significance of the parameter correlations, with *r* values that were displayed in bold format in the correlation map being high-statistically significant ($P \leq 0.05$). In terms of the statistical correlation strength, absolute magnitude of *r* value could be categorized into

"negligible" (0.00−0.09), "weak" (0.10−0.39), "moderate" (0.40−0.69), "strong" (0.70−0.89), and "very strong" (0.90−1.00) [51].

2.5 Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is a multivariate statistical technique [52] that is often applied in specific studies to predict how one variable (called the dependent or target variable) will change based on two or more other variables (called the independent or predictor variables). Regression analysis is divided into univariate and multivariate. Univariate regression studies the linear relationship between two variables and establishes equations [53], [54]. On the other hand, multiple regression analysis includes one independent variable and several independent variables. This method can simultaneously study the impact of multiple independent variables on the independent variables and their interactions to demonstrate the variability of the independent variables. Both analysis methods aim to investigate the relationships, strengths, and interactions between variables under specific control conditions [53], [54]. MLR technology is widely used in environmental research, especially in air pollution research [55]. The Equation (4) is the general expression of the MLR model.

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_ix_i + \varepsilon \quad (4)$$

Here, *y* represented dependent variables, a_i and x_i were the regression coefficient and independent variables, respectively [55], [56]. The statistical analyses were performed using Excel add-ons (XLSTAT2019).

3.0 RESULTS AND DISCUSSION

3.1 Descriptive Statistics

Descriptive statistics were used to summarize the characteristics of a dataset [57], [58]. Through a graphical representation of the box plot, the time-series trend of the five parameters from January to December can be observed as in Figure 2. All the parameter fluctuation ranges can be divided into two parts from November to February and from June to September. Between November and February, rainfall, relative humidity and wind speed increase significantly, while temperatures decrease in reverse. The temperature peaks from June to September, and the relative decrease in humidity and MSL pressure is relatively obvious. According to the climate analysis of the study site, the study site is characterized by two monsoon patterns [59]. From May to September, the southwest monsoon caused rainstorms on the east coast of Peninsular Malaysia, and from November to March, the northeast monsoon was dry. During this period, there is a monsoon interval [60]. The data characteristics of this study are consistent with the climate characteristics of the study site.

3.2 Identification of Rainfall Factors

PCA provides the eigenvector values to determine the number of rotations that the method can run on, thus providing the information on the reliable variables that contribute to the other variable sets. PCs with eigenvalues greater than 1 are generally considered as significant variation factors [61]. According to Table 1, it can be observed that the eigenvalues of F1 and F2 statistically explain 66.111% of the total variance; hence, are considered as significant variation factors.

Table 1 Eigenvalue to determine rotation number

	VF1	VF2	VF3	VF4	F5
Eigenvalue	1.736	1.569	0.875	0.519	0.300
Variability (%)	34.725	31.385	17.498	10.386	6.005
Cumulative (%)	34.725	66.111	83.609	93.995	100.000

Table 1 shows that there are only two rotations in VF, with two D determining the predicted change pattern. Table 2 provides an overview of the predictions and contributions of variables in rainfall research, revealing the importance and reasons for monitoring and controlling variables for future rainfall prediction. VF1 is mainly composed of high positive load relative humidity (0.876) and precipitation (0.779), indicating that the increase of air humidity is the key factor triggering precipitation under tropical climate conditions. When the water vapor content in the air is sufficient, the water vapor may easily undergo condensation once the saturation point has been reached; thus, forming precipitation. VF2 is composed of the positive load of air pressure (0.823), wind speed (0.652) and the negative load of temperature (–0.656). It represents the wind temperature structure dominated by pressure. The state for dry air can be defined as in Equation (5):

$$P = \rho RT \quad (5)$$

Where *P* is the atmosphere pressure, *R* is the gas constant for dry air ($R = 287\ J\ kg^{-1}\ K^{-1}$), ρ is the air density, and *T* is the atmosphere temperature [62]. The equation shows that both a decrease in atmospheric pressure and an increase in air temperature reduce air density. The atmospheric thermodynamic mechanism reflected by VF2 can be used to explain the effects of wind speed and pressure changes on the local climate system.

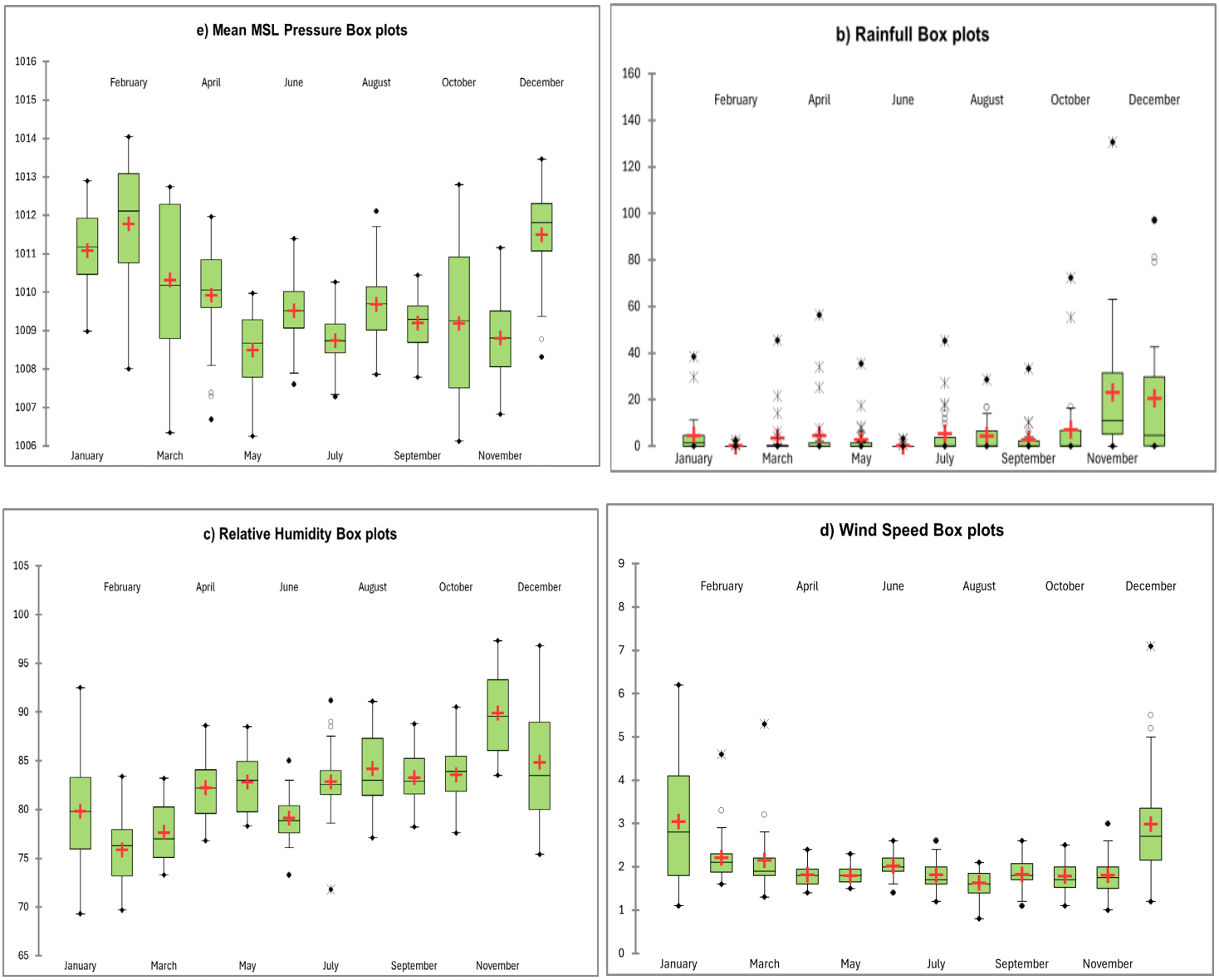


Figure 2 Box plots of five parameters (a) temperature (b) rainfall (c) relative humidity (d) average wind speed (e) Mean MSL pressure

As can be seen from Figure 3, the relationship between rainfall and relative humidity is the closest. When the water vapor content in the air increases, the relative humidity increases. When the humidity rises to a certain level, the water vapor in the air begins to condense and form clouds, which in turn form rain [63]. During rainfall, water on the ground and plants evaporate into the air, increasing the water vapor content in the air and thus increasing the relative humidity [64].

Wind speed, air pressure, and temperature are closely related. The relationship between wind speed and air pressure is indirect and regulated by temperature that influences air density, which in turn affects air pressure and wind speed. As in the ideal gas law, the temperature gradient is directly proportional to the air pressure gradient, which further induces wind dynamics [65], [66]. This is evident as during the monsoons when the convection is unstable, the warm air on the surface rises (becomes less dense), which will leave a low-pressure area behind it that further induce suction of surrounding air into the center, thus, causing significant wind movement. As the cold air in the upper atmosphere sinks into the surface (becomes denser), it will form a high-pressure area, which in turn induces air outflow from the center accompanied by wind movement [65], [66].

Table 2 The variable factor (VFs) load of 5 climate parameters

	VF1	VF2
Mean MSL Pressure	–0.191	0.823
Average Temperature	–0.531	–0.656
Mean Relative Humidity	0.876	–0.222
Mean Wind Speed	–0.077	0.652
Rainfall	0.779	0.157

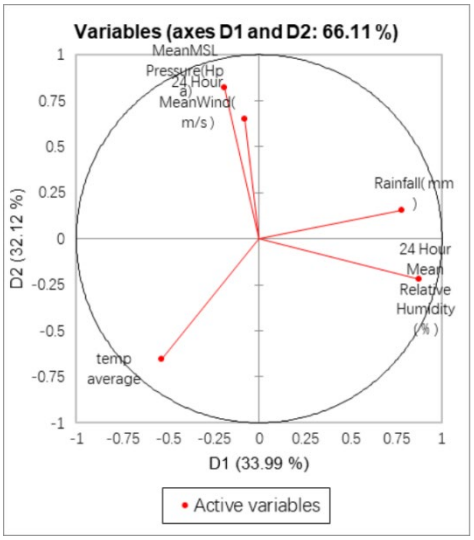


Figure 3 Factor loadings after Varimax rotation

3.3 Correlation Study

The findings on the correlations between the five focused climate parameters in the study location were visualized as in Table 3 and Figure 4. The colors in the matrix in Figure 4 are used to represent the strength of the correlation. The red series indicates the positive correlation, the blue series indicates the negative correlation, and the darker the color indicates the stronger the correlation. Overall, our correlation study revealed that most of the correlations between the focused climate parameters are high-statistically significant ($P \leq 0.05$), except the correlation between rainfall and mean MSL pressure which could be caused by high variabilities between these two parameters.

According to Table 3 and Figure 4, rainfall and average temperature show a light blue color, showing a weak negative correlation ($r = –0.293$) may suggest evaporative cooling that reflects heat transfer from the surrounding air into the raindrops when the rainfall occurs in the atmosphere with unsaturated air; thus, lowering the surrounding air temperature [67]. Rainfall and humidity are orange, and there is a moderate positive correlation between them ($r = 0.460$) is explained by the increase in water vapor content in the air during high-humid conditions that may increase the probability of rainfall [68], [69]. Precipitation and average wind speed are yellow, showing a weak positive correlation ($r = 0.139$) indicates that higher wind speed may encourage water evaporation into the atmosphere, which induces boundary layer destabilization and deep convection associated with moisture convergence; thus, promoting precipitation [70], [71]. Overall, these correlation findings are well-aligned to the PCA results.

Whereas the correlations among other four climate parameters are statistically evident, although their correlation magnitude may not be necessarily high due to their significant variability caused by the monsoonal variation at the study location. Mean wind speed and mean MSL pressure are shown in yellow with a weak positive correlation ($r = 0.246$). Pressure gradient is one of the main factors driving wind speed [72]. When the pressure gap between the two places is large, the air will flow from the high pressure area to the low pressure area, forming a wind. The greater the pressure gradient, the stronger the wind speed. This is consistent with the findings of [73]. The mean wind speed is negatively correlated with mean temperature ($r = –0.187$) and mean relative humidity ($r = –0.193$), both of which are shown in green. In the tropics, when the daytime temperature is higher, the air is heated up, which may cause the hot air to rise and cause the local air flow to be smooth, resulting in lower wind speed; And possibly due to the effect of humidity on air density, the air is more stable at high humidity, resulting in lower wind speed. This is similar to the findings produced by [74]. In addition, the current analysis shows that mean relative humidity and mean MSL pressure ($r = –0.280$) and mean temperature ($r = –0.321$) are both shown in light blue, showing a weak negative correlation [74], [75]. According to the Equation (6), (7), and (8) [76]:

$$RH = 100\frac{e}{e_s} \quad (6)$$

$$e = C\exp\left[\frac{A\,t_d}{B + t_d}\right] \quad (7)$$

$$e_s = C\exp\left[\frac{A\,t}{B + t}\right] \quad (8)$$

Where RH is relative humidity, e is the actual water vapor pressure, e_s is the saturation (or equilibrium) water vapor pressure, t and t_d are initial air temperature and dew point temperature (temperature at which air, with certain level of water vapor, must be cooled to become completely saturated) respectively, and A, B and C are the adjusted coefficients of 17.625, 243.04 ℃ and 610.94 Pa respectively.

It can be seen from the equation that the saturated water vapor pressure increases exponentially with the temperature, while the actual water vapor pressure does not necessarily synchronize well; hence, the relative humidity tends to decrease as the temperature rises. When the air is in a low pressure system, the air rises, and the

temperature decreases with the increase of height, which makes the water vapor in the air more easy to cool and condense to form clouds and precipitation; thus, increasing the relative humidity. This is a key reason why areas of low pressure are often accompanied by cloudy or rainy weather [77].

Lastly, average temperature is identified to be moderate-negatively correlated to mean MSL pressure (r = –0.410), which is similar to the findings of [78]. This correlation could be attributed to the direct contribution of high temperatures in the tropics to the formation of cyclones as low atmospheric pressure systems [79].

Table 3 Pearson correlation coefficient matrix of five focused climate parameters in the study location

Variables	Mean MSL Pressure	Average Temperature	Mean Relative Humidity	Mean Wind Speed	Rainfall
Mean MSL Pressure	**1**				
Average Temperature	**–0.410**	**1**			
Mean Relative Humidity	**–0.280**	**–0.321**	**1**		
Mean Wind Speed	**0.246**	**–0.187**	**–0.193**	**1**	
Rainfall	–0.039	**–0.293**	**0.46**	**0.139**	**1**

**The bolded correlation coefficient values are highly significant ($P \leq 0.05$).*

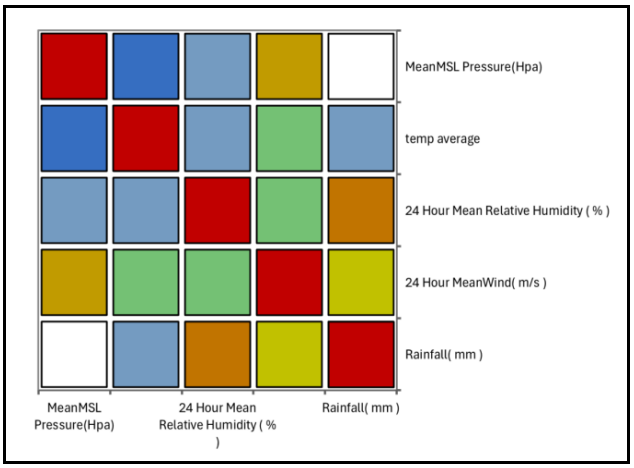


Figure 4 Pearson correlation map of the five focused climate parameters in the study location

The findings on Pearson correlations between rainfall and other four focused climate parameters are well-aligned with the bar chart of standardized coefficients for rainfall linear regression model in Figure 5. The height of the histogram represents the strength of each variable's contribution to the precipitation prediction model, while the positive and negative directions of the column represent the direction of statistical influence (positive or negative correlation).

The results reveal the mean relative humidity as the largest contributor to the rainfall occurrence in the study location, which is followed by mean wind speed and air temperature. There is no statistically significant influence of mean MSL pressure on the rainfall occurrence at 95% confidence interval level, which may further confirm the high variabilities and negligible statistical correlation between these two parameters at the study location.

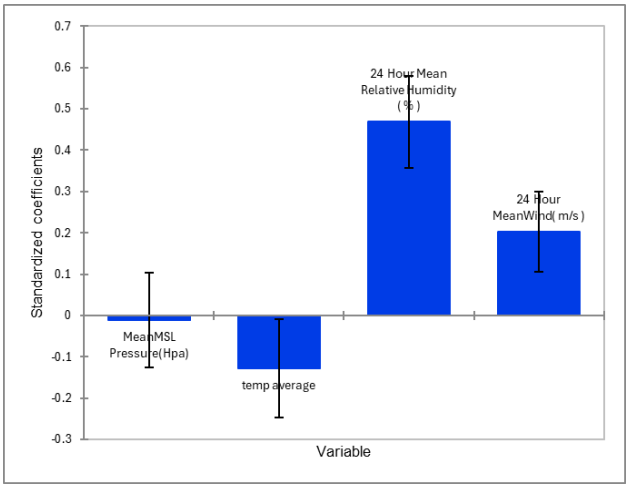


Figure 5 Bar chart of standardized coefficients for the four focused climate parameters (mean MSL pressure, average temperature, mean relative humidity, and mean wind speed) for rainfall linear regression model

3.4 Prediction of the Rainfall Data using MLR

An MLR model was used to finely depict the dynamic performance of variables. The core mechanism of this model is linear least square fitting with the attempts of identifying the best-fitting line for the set of data points via minimization of the sum of squared differences or errors between the real data values and the estimated values by the model [80].

Coefficient of determination (R^2) was used to evaluate the performance of the MLR model. The R^2 value only provides information about how well it performs on external data [81]. Root mean square error (RMSE) measures residual error, which gives an estimate of the average difference between observed and simulated values for climate. If the R^2 value is 1 and the RMSE value is minimal, a better model should be executed [82]. The results showed that the statistical R^2 of goodness of fit was 0.293 and the RMSE was 13.581 (Table 4), which indicate relatively low R^2 larger RMSE value. The numerical results further imply that the current model has a certain predictive capability, although it might not be highly accurate as there is a noticeable deviation between the predicted value and the actual observed value caused by high data variability or noise.

Table 4 Summary of regression of the rainfall's variable produced by this study

Goodness of fit statistics	
Observations	334.000
Sum of weights	334.000
DF	329.000
R^2	0.293
Adjusted R^2	0.284
MSE	184.431
RMSE	13.581

Our established rainfall variability prediction model (hence named as Wang's Rainfall Model) produced the equation for predicting the rainfall intensity (mm) as in the Equation (9):

$$Predicted\ rainfall\ (mm) = 36.947025196464 - 0.105363766442354 \times P - 1.81024357154312 \times T + 1.43988809750382 \times H + 3.87787550643935 \times W \quad (9)$$

Here, *P* represents MSL Pressure (hPa), *T* is the 24-hour average temperature (℃), *H* is the 24-hour average humidity (%), and *W* is the 24-hour average wind speed (m/s).

3.5 Validation of Wang's rainfall variability prediction model

Figure 6 displays the significant gaps or errors between the current forecast accuracy of our rainfall MLR model and the actual rainfall. From the perspective of practical application, although the model can be used as a preliminary estimation tool of precipitation trend, its prediction accuracy is not enough to support high-sensitivity real-time forecasting or disaster warning. Hence, larger sample size along with hyperparameter tuning might be considered to ensure higher accuracy and precision in rainfall variability prediction. As the sample size increases, the standard error will be more minimized, which contributes to higher precision of rainfall estimation by the regression model with minimized influence of outliers [83]. This further results in lower variance for coefficient estimates and higher R^2 value, which indicates more explained variability by the model and better model fit. Furthermore, hyperparameter tuning is also essential to ensure best suitability and optimal performance of the model in predicting the rainfall variability with high accuracy and precision [84]. Nevertheless, our rainfall MLR model is shown to exhibit roughly similar rainfall trend shape as in the actual one, which may serve as a feasible baseline rainfall variability prediction model.

It is also important to consider the importance of data quality over data quantity. Noisy data, erroneous data, or irrelevant data may negatively affect the model. Systematic errors or biases in the data may induce unfeasible statistical modeling and poor data generalization [85].

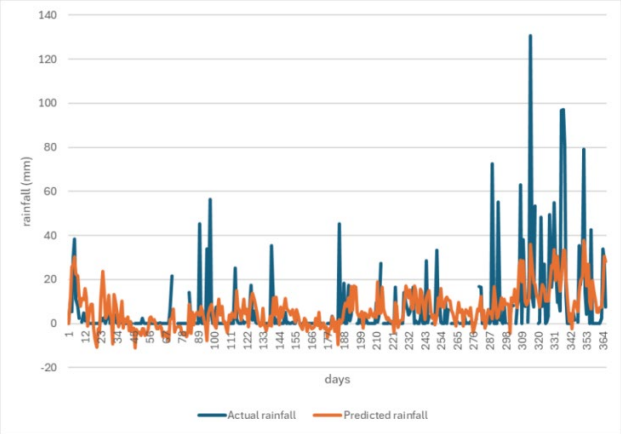


Figure 6 Line chart of actual rainfall compared to MLR model

4.0 CONCLUSION

The preliminary results of this study show that the PCA discerns average daily temperature, relative humidity, and means wind speed as the key meteorological parameters that play important roles in regulating the rainfall dynamics in Kuala Terengganu. Additionally, despite the low predictive capability of our current rainfall MLR model due to high data variability or noise, it is capable to capture the general rainfall trend shape similar to the actual one.

Overall, the application of MLR-based PCA has feasibly facilitated rainfall prediction by our current study. Combination of MLR and PCA can provide good performance metrics and improve efficiency by eliminating collinearity problems and reducing the number of predictor variables. This model is thus proved to be a useful tool for meteorological agencies to make more effective climate predictions in Malaysia as it may help in reducing the cost of instruments or tools used in sampling activities and analysis. Future research studies or any rainfall monitoring scheme plan should consider the incorporation of large historical data size with high-quality in establishing rainfall variability prediction model along with hyperparameter tuning. The utilization of both PCA and MLR can further develop better sampling strategies to help local and government authorities more effectively respond to climate disasters. This study strongly recommends using statistical data and trend analysis in data analysis, which not only provides more valuable information but also significantly reduces sampling costs and time. Therefore, these approaches are crucial for the accuracy and effectiveness of future rainfall prediction.

Acknowledgement

This work was supported by the Universiti Sains Malaysia, Research University Transdisciplinary (RUTrans) Grant Scheme (Grant Number: R502-KR-RUT002-00000001110-K134) and was supported in full or in part with Kurita Asia Research Grant (24Pmy142) provided by Kurita Water and Environment Foundation.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Mansouri, A., B. Aminnejad, and H. Ahmadi. 2018. Investigating the Effect of Climate Change on Inflow Runoff into the Karun-4 Dam Based on IPCC's Fourth and Fifth Report. *JWSS-Isfahan University of Technology*. 22(2): 345–359.

[2] Melillo, J. M., T. T. Richmond, and G. Yohe. 2014. Climate Change Impacts in the United States. In *Third National Climate Assessment*. 52: 150–174.

[3] WMO. 2020. *2020 State of Climate Services: Risk Information and Early Warning Systems*. WMO-No. 1252. https://library.wmo.int/idurl/4/57191.

[4] NOAA. 2021. Record-Breaking Atlantic Hurricane Season Draws to an End, Dec. 2020. https://www.noaa.gov/media-release/record-breaking-atlantic-hurricane-season-draws-to-end.

[5] NOAA National Centers for Environmental Information (NCEI). 2025. *U.S. Billion-Dollar Weather and Climate Disasters*. https://www.ncei.noaa.gov/access/billions/.

[6] Komesaroff, P., and I. Kerridge. 2020. *A Continent Aflame: Ethical Lessons from the Australian Bushfire Disaster*. Springer.

[7] Steven, A., *et al.* 2020. *Coastal Development: Resilience, Restoration and Infrastructure Requirements*. Washington, DC: World Resources Institute.

[8] Bagheri, M., *et al.* 2021. "Application of Multi-Criteria Decision-Making Model and Expert Choice Software for Coastal City Vulnerability Evaluation." *Urban Science*. 5(4): 84.

[9] Gallina, V., S. Torresan, A. Zabeo, A. Critto, T. Glade, and A. Marcomini. 2020. A Multi-Risk Methodology for the Assessment of Climate Change Impacts in Coastal Zones. *Sustainability*. 12(9): 3697.

[10] Trenberth, K. E., and D. P. Stepaniak. 2001. Indices of El Niño Evolution. *Journal of Climate*. 14(8): 1697–1701.

[11] McPhaden, M. J., S. E. Zebiak, and M. H. Glantz. 2006. ENSO as an Integrating Concept in Earth Science. *Science*. 314(5806): 1740–1745.

[12] NOAA Climate.gov. 2025. Climate Change: Global Sea Level. https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level.

[13] Wang, F., Y. Shen, J. Geng, and Q. Chen. 2024. Global Mean Sea Level Change Projections up to 2100 Using a Weighted Singular Spectrum Analysis. *Journal of Marine Science and Engineering*. 12(12): e2124. https://doi.org/10.3390/jmse12122124.

[14] Nicholls, R. J., and A. Cazenave. 2010. Sea-Level Rise and Its Impact on Coastal Zones. *Science*. 328(5985): 1517–1520.

[15] Bagheri, M., A. H. Nik, S. A. Shahid, M. F. Ramli, and H. Z. Abidin. 2021. Impacts of Future Sea-Level Rise under Global Warming Assessed from Tide Gauge Records: A Case Study of the East Coast Economic Region of Peninsular Malaysia. *Land (Basel)*. 10(12): 1382.

[16] Awang, N. A., and M. R. A. Hamid. 2013. Sea Level Rise in Malaysia. *Sea Level Rise Adaptation Measures. Hydrolink*. 2: 47–49.

[17] Rezania, S., N. Z. A. Rahim, A. A. Z. Abidin, H. Dahalan, and M. A. Mohd. 2020. Technical Aspects of Biofuel Production from Different Sources in Malaysia—A Review. *Processes*. 8(8): 993.

[18] Othman, W. Z., N. N. A. Tukimat, A. H. Azman, S. N. Rahmat, and B. Winarta. 2022. Analysis of Climate Variability and Trends in the Context of Climate Changes: Case Study in Terengganu. *International Journal of Integrated Engineering*. 14(9): 88–97.

[19] Guo, R. 2021. Cross-Border Environmental Pollution and Human Health. In *Cross–Border Resource Management*, 4th ed. 291–337.

[20] Van Eekelen, M. W., P. Droogers, R. P. Hoogeveen, R. Gupta, and S. Savenije. 2015. A Novel Approach to Estimate Direct and Indirect Water Withdrawals from Satellite Measurements: A Case Study from the Incomati Basin. *Agriculture, Ecosystems & Environment*. 200: 126–142.

[21] Dhakal, A. S., and K. Sullivan. 2014. Shallow Groundwater Response to Rainfall on a Forested Headwater Catchment in Northern Coastal California: Implications of Topography, Rainfall, and Throughfall Intensities on Peak Pressure Head Generation. *Hydrological Processes*. 28(3): 446–463.

[22] He, J., J. Wu, and J. Luo. 2020. Introduction to Climate Forecast System Version 1.0 of Nanjing University of Information Science and Technology. *Transactions of Atmospheric Sciences*. 43(1): 128–143.

[23] Mu, M., W. Duan, D. Chen, and W. Yu. 2015. Target Observations for Improving Initialization of High-Impact Ocean-Atmospheric Environmental Events Forecasting. *National Science Review*. 2(2): 226–236.

[24] Palmer, T. N. 2000. Predicting Uncertainty in Forecasts of Weather and Climate. *Reports on Progress in Physics*. 63(2): 71.

[25] Luo, J.-J., *et al.* 2016. Current Status of Intraseasonal–Seasonal-to-Interannual Prediction of the Indo-Pacific Climate. In *Indo-Pacific Climate Variability and Predictability*. 63–107. Singapore: World Scientific.

[26] Duan, W., and F. Zhou. 2013. Non-Linear Forcing Singular Vector of a Two-Dimensional Quasi-Geostrophic Model. *Tellus A: Dynamic Meteorology and Oceanography*. 65(1): 18452.

[27] Mekanik, F., M. A. Imteaz, S. Gato-Trinidad, and A. Elmahdi. 2013. Multiple Regression and Artificial Neural Network for Long-Term Rainfall Forecasting Using Large Scale Climate Modes. *Journal of Hydrology (Amsterdam)*. 503: 11–21.

[28] Moradkhani, H., and M. Meier. 2010. Long-Lead Water Supply Forecast Using Large-Scale Climate Predictors and Independent Component Analysis. *Journal of Hydrologic Engineering*. 15(10): 744–762.

[29] Fattorini, M., and C. Brandini. 2020. "Observation Strategies Based on Singular Value Decomposition for Ocean Analysis and Forecast. *Water (Basel)*. 12(12): 3445.

[30] Mazlum, N., and A. Ö. S. Mazlum. 1999. Interpretation of Water Quality Data by Principal Components Analysis. *Turkish Journal of Engineering and Environmental Sciences*. 23(1): 19–26.

[31] Bagheri, M., *et al.* 2021. Land-Use Suitability Assessment Using Delphi and Analytical Hierarchy Process (D-AHP) Hybrid Model for Coastal City Management: Kuala Terengganu, Peninsular Malaysia. *ISPRS International Journal of Geo-Information*. 10(9). https://doi.org/10.3390/ijgi10090621.

[32] Yusof, K. M. K. K., S. S. Ismail, A. Azid, M. S. A. Sani, N. M. Isa, and M. Z. M. Zawawi. 2020. Variability on Particulate Matter and Meteorology Dataset During the Hazy Period in Eastern Region of Peninsular Malaysia. *Data in Brief*. 29: 105210.
[33] Shafii, N. Z., *et al.* 2019. Application of Chemometrics Techniques to Solve Environmental Issues in Malaysia. *Heliyon*. 5(10).
[34] Juahir, H., A. Retnam, M. A. Zali, and M. F. Hashim. 2011. A Comparison Between Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) for River Class Prediction at Klang River, Malaysia. In *Contemporary Environmental Quality Management in Malaysia and Selected Countries*. Serdang: Universiti Putra Malaysia Press.
[35] Gazzaz, N. M., M. K. Yusoff, M. F. Ramli, A. Z. Aris, and H. Juahir. 2012. Characterization of Spatial Patterns in River Water Quality Using Chemometric Pattern Recognition Techniques. *Marine Pollution Bulletin*. 64(4): 688–698.
[36] Kowalkowski, T., R. Zbytniewski, J. Szpejna, and B. Buszewski. 2006. Application of Chemometrics in River Water Classification. *Water Research*. 40(4): 744–752.
[37] Simeonov, V., J. W. Einax, I. Stanimirova, and J. Kraft. 2002. Environmetric Modeling and Interpretation of River Water Monitoring Data. *Analytical and Bioanalytical Chemistry*. 374: 898–905.
[38] Spitzer, M., J. Wildenhain, J. Rappsilber, and M. Tyers. 2014. BoxPlotR: A Web Tool for Generation of Box Plots. *Nature Methods*. 11(2): 121–122.
[39] Samsudin, M. S., A. Azid, S. I. Khalit, M. S. A. Sani, and F. Lananan. 2019. Comparison of Prediction Model Using Spatial Discriminant Analysis for Marine Water Quality Index in Mangrove Estuarine Zones. *Marine Pollution Bulletin*. 141: 472–481.
[40] Pearson, K. 1901. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space."*The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 2(11): 559–572. https://doi.org/10.1080/14786440109462720.
Jolliffe, I. T. 2002. *Principal Component Analysis for Special Types of Data*. Springer.
[41] Adeen, N., M. Abdulazeez, and D. Zeebaree. 2020. Systematic Review of Unsupervised Genomic Clustering Algorithms Techniques for High Dimensional Datasets. *Technological Reports of Kansai University*. 62(3): 355–374.
[42] Hasan, B. M. S., and A. M. Abdulazeez. 2021. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*. 2(1): 20–30.
[43] Hasan, B. M. S., and A. M. Abdulazeez. 2021. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*. 2(1): 20–30.
[44] Zhuang, X., Z. Yang, and D. Cordes. 2020. A Technical Review of Canonical Correlation Analysis for Neuroscience Applications. *Human Brain Mapping*. 41(13): 3807–3833.
[45] Carreras Simó, M., and G. Coenders. 2020. Principal Component Analysis of Financial Statements: A Compositional Approach = Análisis en Componentes Principales de los Estados Financieros: Un Enfoque Composicional. *Revista de Métodos Cuantitativos para la Economía y la Empresa*. 29: 18–37.
[46] Wang, M., L. Song, K. Sun, and Z. Jia. 2020. F-2D-QPCA: A Quaternion Principal Component Analysis Method for Color Face Recognition. *IEEE Access*. 8: 217437–217446.
[47] Karpuzcu, M. E., D. Fairbairn, W. A. Arnold, B. L. Barber, E. Kaufenberg, W. C. Koskinen, P. J. Novak, P. J. Rice, and D. L. Swackhamer. 2014. Identifying Sources of Emerging Organic Contaminants in a Mixed Use Watershed Using Principal Components Analysis. *Environmental Science: Processes & Impacts*. 16(10): 2390–2399. https://doi.org/10.1039/C4EM00324A.
[48] Kim, J.-O., and C. W. Mueller. 1978. *Introduction to Factor Analysis: What It Is and How to Do It*. No. 13. Sage.
[49] Vega, M., R. Pardo, E. Barrado, and L. Debán. 1998. Assessment of Seasonal and Polluting Effects on the Quality of River Water by Exploratory Data Analysis. *Water Research*. 32(12): 3581–3592.
[50] Liu, C.-W., K.-H. Lin, and Y.-M. Kuo. 2003. Application of Factor Analysis in the Assessment of Groundwater Quality in a Blackfoot Disease Area in Taiwan. *Science of the Total Environment*. 313(1–3): 77–89.
[51] Schober, P., C. Boer, and L. A. Schwarte. 2018. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*. 126(5): 1763–1768.
[52] Samsudin, M. S., A. Azid, S. I. Khalit, A. S. M. Saudi, and M. A. Zaudi. 2017. River Water Quality Assessment Using APCS-MLR and Statistical Process Control in Johor River Basin, Malaysia. *International Journal of Advanced and Applied Sciences*. 4(8): 84–97.
[53] Tabachnick, B. G., and L. S. Fidell. 1983. *Using Multivariate Statistics*. New York: HarperCollins Publishers.
[54] Büyüköztürk, Ş. 2002. *Sosyal Bilimler için Veri Analizi El Kitabı*. Ankara: Pegem Yayıncılık.
[55] Yusof, K. M. K. K., A. Azid, M. S. A. Sani, M. S. Samsudin, S. N. S. Muhammad Amin, N. L. A. Rani, and M. A. Jamalani. 2019. The Evaluation on Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) Models over Particulate Matter (PM10) Variability during Haze and Non-Haze Episodes: A Decade Case Study. *Malaysian Journal of Fundamental and Applied Sciences*. 15(2): 164–172. https://doi.org/10.11113/mjfas.v15n2.1004.
[56] Özdemir, U., and S. Taner. 2014. Impacts of Meteorological Factors on PM10: Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) Approaches. *Environmental Forensics*. 15(4): 329–336. https://doi.org/10.1080/15275922.2014.950133.
[57] Johnson, R. A., and G. K. Bhattacharyya. 2019. *Statistics: Principles and Methods*. 8th ed. Hoboken, NJ: John Wiley & Sons.
[58] Juahir, H., M. Z. Yusoff, S. M. Zain, A. Z. Aris, M. K. Yusoff, M. F. Ramli, and N. A. Mohd. 2011. Spatial Water Quality Assessment of Langat River Basin (Malaysia) Using Environmetric Techniques. *Environmental Monitoring and Assessment* 173: 625–641. https://doi.org/10.1007/s10661-010-1411-x.
[59] Suhaila, J., S. M. Deni, W. Z. Wan Zin, and A. A. Jemain. 2010. Spatial Patterns and Trends of Daily Rainfall Regime in Peninsular Malaysia during the Southwest and Northeast Monsoons: 1975–2004. *Meteorology and Atmospheric Physics*. 110: 1–18. https://doi.org/10.1007/s00703-010-0092-6.
[60] Fadzil, M., M. A. Mohd Akhir, and Y. J. Chuen. 2011. Seasonal Variation of Water Characteristics during Inter-Monsoon along the East Coast of Johor. *Journal of Sustainability Science and Management*. 6(2): 206–214.
[61] Karlis, D., G. Saporta, and A. Spinakis. 2003. A Simple Rule for the Selection of Principal Components. *Communications in Statistics—Theory and Methods*. 32(3): 643–666. https://doi.org/10.1081/STA-120018556
[62] Holton, J. R. 1992. *An Introduction to Dynamic Meteorology*. International Geophysics Series, Vol. 48. San Diego, CA: Academic Press.
[63] Umoh, A. A., A. O. Akpan, and B. B. Jacob. 2013. Rainfall and Relative Humidity Occurrence Patterns in Uyo Metropolis, Akwa Ibom State, South-South Nigeria. *IOSR Journal of Engineering*. http://www.iosrjen.org.
[64] Mawonike, R., and G. Mandonga. 2017. The Effect of Temperature and Relative Humidity on Rainfall in Gokwe Region, Zimbabwe: A Factorial Design Perspective. *International Journal of Multidisciplinary Academic Research*. 5(2). http://www.multidisciplinaryjournals.com.
[65] Spiridonov, V., and M. Ćurić. 2021. Atmospheric Pressure and Wind. In *Fundamentals of Meteorology*. 87–114. Cham: Springer. https://doi.org/10.1007/978-3-030-52655-9_5.

[66] Wooten, R. D. 2011. Statistical Analysis of the Relationship between Wind Speed, Pressure and Temperature. *Journal of Applied Sciences*. 11(15): 2712–2722. https://doi.org/10.3923/jas.2011.2712.2722.
[67] Rosenfeld, D., U. Lohmann, G. B. Raga, C. D. O'Dowd, M. Kulmala, S. Fuzzi, A. Reissell, and M. O. Andreae. 2014. Global Observations of Aerosol–Cloud–Precipitation–Climate Interactions. *Reviews of Geophysics*. 52(4): 750–808. https://doi.org/10.1002/2013RG000441.
[68] Bretherton, C. S., M. E. Peters, and L. E. Back. 2004. Relationships between Water Vapor Path and Precipitation over the Tropical Oceans. *Journal of Climate*. 17(7): 1517–1528. https://doi.org/10.1175/1520-0442(2004)017.
[69] Zhu, L., X. Chen, and L. Bai. 2020. Relative Roles of Low-Level Wind Speed and Moisture in the Diurnal Cycle of Rainfall over a Tropical Island under Monsoonal Flows. *Geophysical Research Letters*. 47(8): e2020GL087467. https://doi.org/10.1029/2020GL087467.
[70] Folkins, I., T. Mitovski, and J. R. Pierce. 2014. A Simple Way to Improve the Diurnal Cycle in Convective Rainfall over Land in Climate Models. *Journal of Geophysical Research: Atmospheres*. 119(5): 2113–2130. https://doi.org/10.1002/2013JD020149.
[71] Back, L. E., and C. S. Bretherton. 2005. The Relationship between Wind Speed and Precipitation in the Pacific ITCZ. *Journal of Climate*. 18(20): 4317–4328. https://doi.org/10.1175/JCLI3519.1.
[72] Wu, J., J. Zha, D. Zhao, and Q. Yang. 2018. Changes in Terrestrial Near-Surface Wind Speed and Their Possible Causes: An Overview. *Climate Dynamics*. 51(5): 2039–2078. https://doi.org/10.1007/s00382-017-3977-6.
[73] Gore, R. D., and B. W. Gawali. 2023. Analysis of Weather Parameters Using Machine Learning. In *Proceedings of the International Conference on Artificial Intelligence and Smart Systems*. 569–589. Atlantis Press. https://doi.org/10.2991/978-94-6463-196-8_44.
[74] Siloko, I. U., and O. O. Uddin. 2023. A Statistical Study of Wind Speed and Its Connectivity with Relative Humidity and Temperature in Ughelli, Delta State, Nigeria. *Science World Journal*. 18(3): 404–413. https://doi.org/10.4314/swj.v18i3.13,
[75] Emekwuru, N., and O. Ejohwomu. 2023. Temperature, Humidity and Air Pollution Relationships during a Period of Rainy and Dry Seasons in Lagos, West Africa. *Climate*. 11(5): 113. https://doi.org/10.3390/cli11050113.
[76] Lawrence, Mark G. 2005. The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bulletin of the American Meteorological Society*. 86(2): 225–234. https://doi.org/10.1175/BAMS-86-2-225.
[77] Speer, Michael S., Paul Wiles, and Acacia Pepler. 2009. Low Pressure Systems off the New South Wales Coast and Associated Hazardous Weather: Establishment of a Database. *Australian Meteorological and Oceanographic Journal*. 58(1): 29–39.
[78] Mohr, Matthias. 2004. Problems with the Mean Sea Level Pressure Field over the Western United States. *Monthly Weather Review*. 132(8): 1952–1965. https://doi.org/10.1175/1520-0493(2004)132<1952:PWTMSL>2.0.CO;2.
[79] Randel, William J., and Eric J. Jensen. 2013. Physical Processes in the Tropical Tropopause Layer and Their Roles in a Changing Climate. *Nature Geoscience*. 6(3): 169–176. https://doi.org/10.1038/ngeo1733.
[80] Henry, Robert C., Clifford W. Lewis, Philip K. Hopke, and Henry J. Williamson. 1984. Review of Receptor Model Fundamentals. *Atmospheric Environment*. 18(8): 1507–1515. https://doi.org/10.1016/0004-6981(84)90375-5.
[81] Azid, Azman, *et al.* 2018. Air Quality Modelling Using Chemometric Techniques. *Journal of Fundamental and Applied Sciences*. 9(2S): 443–459. https://doi.org/10.4314/jfas.v9i2s.30.
[82] Pai, Tzu-Yuan, *et al.* 2010. Predicting Hourly Ozone Concentration in Dali Area of Taichung County Based on Multiple Linear Regression Method. *International Journal of Applied Science and Engineering*. 7(2): 127–132.
[83] Diez, David M., Christopher D. Barr, and Mine Çetinkaya-Rundel. 2012. *OpenIntro Statistics*. Boston, MA: OpenIntro.
[84] L. P., and S. S. D. 2023. A Novel Model for Rainfall Prediction Using Hybrid Stochastic-Based Bayesian Optimization Algorithm. *Environmental Science and Pollution Research*. 30: 92555–92567. https://doi.org/10.1007/s11356-023-28734-z.
[85] Dockès, Jérémie, Gaël Varoquaux, and Jean-Baptiste Poline. 2021. Preventing Dataset Shift from Breaking Machine-Learning Biomarkers. *GigaScience*. 10(9): giab055. https://doi.org/10.1093/gigascience/giab055.