Available online at
https://mijuitm.com.my

e-ISSN: 2735-0703

Mathematical Sciences and Informatics Journal 7(1) 2026, 1 – 21.

Mathematical Sciences and Informatics Journal

Impact of PM_{10} Concentrations: A Markov Chain Approach to Examining Wind Speed Effects in Petaling Jaya

Shamshimah Samsuddin[1*], Azizul Hakim Nashabudin[1], Muhammad Aizuddin Muhd Anwar[1], Aini Atiqah Md Ramli[1], Arba'iyah Mohd Kamil[1]

[1]*School of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*

ARTICLE INFO

Article history:
Received 6 October 2025
Revised 18 December 2025
Accepted 15 January 2026
Online first
Published 30 April 2026

Keywords:
Wind Speed
Air Pollution
Markov Chain
Simulation
PM_{10} Concentrations

DOI:
10.24191/mij.v7i1.8169

ABSTRACT

Air pollution, particularly PM_{10} concentration, poses significant health and environmental challenges in urban areas such as Petaling Jaya. Wind speed is known to influence pollutant dispersion, however, the dynamic relationship between wind speed and PM_{10} levels over time has not been adequately modelled using probabilistic approaches. This study aims to investigate the influence of wind speed on PM_{10} concentrations in Petaling Jaya from 2013 to 2023. It seeks to model the temporal evolution and dependency between wind speed and PM_{10} concentrations using a Markov Chain framework and to identify the most suitable statistical models for accurate prediction and assessment. A first order Markov Chain model was developed using Transition Probability Matrices (TPMs— mathematical frameworks that capture the likelihood of transitioning between different environmental states over time) constructed through the Count Method. The model's assumptions, which include periodicity, irreducibility, and state classification, were validated through statistical tests. Additionally, a Partial Proportional Odds Model (PPOM), an Ordered Logit Model (OLM), and an Ordered Probit Model (OPM) were applied and compared with the Count Method. Monte Carlo simulations were used to assess the models' performance under varying environmental conditions. The findings reveal that higher wind speeds significantly enhance the dispersion of PM_{10}, whereas lower wind speeds lead to the accumulation of pollutants. Among the models, the OPM best captures the distribution of wind speed, whilst the PPOM demonstrates the highest accuracy in predicting PM_{10} concentrations. These results provide valuable insights into air quality management and environmental policymaking.

[1*] Corresponding author. *E-mail address*: shams611@uitm.edu.my
https://doi.org/10.24191/mij.v7i1.8169

1. INTRODUCTION

Air pollution is a persistent problem in urban areas, especially in rapidly developing cities with high levels of industrial activity and traffic emissions. One of the most concerning air pollutants is Particulate Matter with an aerodynamic diameter of less than 10 micrometres (PM_{10}). Due to its small size, PM_{10} can penetrate deep into the respiratory system and is associated with various adverse health effects. In Malaysia, the Klang Valley frequently experiences elevated PM_{10} concentrations, driven by urbanisation, dense transportation networks, and local meteorological conditions. As a major urban centre within the Klang Valley, Petaling Jaya continues to face significant challenges in maintaining acceptable air quality levels.

1.1 Influence of Wind Speed on PM10 Concentrations

Meteorological conditions, especially wind speed, play a crucial role in determining PM_{10} concentration levels. Higher wind speeds generally enhance the dispersion of airborne particles, leading to lower PM_{10} concentrations. Conversely, low wind speed conditions restrict dispersion and promote the accumulation of pollutants near the ground, resulting in elevated pollution levels. However, the relationship between wind speed and PM_{10} concentrations is not static. PM_{10} levels are influenced not only by current wind conditions but also by previous atmospheric states, indicating the presence of temporal dependence in air pollution behaviour. This dynamic interaction highlights the need for analytical approaches that can effectively capture changes in PM_{10} concentrations over time.

1.2 Probabilistic Modelling of Air Quality Using Markov Chains

To account for temporal dependence in air quality data, probabilistic models such as Markov chains provide a suitable analytical framework. Markov Chain models allow PM_{10} concentration levels to be represented as a sequence of discrete states, in which future conditions depend on the current state. By classifying wind speed and PM_{10} concentrations into predefined categories, transition probabilities can be estimated to describe the likelihood of moving between different pollution states under varying wind conditions. This probabilistic structure provides valuable insights into the persistence, variability, and transition patterns of PM_{10} concentrations. Accordingly, this study applies a Markov Chain approach to examine the influence of wind speed on PM10 concentrations in Petaling Jaya. Transition probabilities are estimated using the Count Method and are compared with alternative statistical models to ensure robustness. In addition, Monte Carlo simulations are employed to explore potential PM_{10} concentration patterns under different wind speed scenarios. The findings of this study are expected to enhance understanding of urban air quality dynamics and to support evidence-based environmental management and policy formulation in Petaling Jaya.

2. LITERATURE REVIEW

This review examines the relationship between wind speed and PM_{10} concentrations in Malaysia with a focus on the Klang Valley, one of the country's most industrialised and polluted regions. Malaysia has a tropical monsoon climate and rising air pollution levels; therefore, it is important to understand how meteorological factors, especially wind speed, influence the dispersion of particulate matter. This study integrates meteorological analysis with stochastic modelling, particularly Markov Chain models, to classify wind speed and PM_{10} levels into discrete states, estimate transition probabilities, and evaluate their correlation for better forecasting and environmental planning (Kar Yong and Awang 2017; Dominick et al. 2012; Zakaria et al. 2019a).

https://doi.org/10.24191/mij.v7i1.8169

2.1 Interactions between Wind Speed and PM_{10} Concentrations in Malaysia

Wind speed, usually measured in metres per second or kilometres per hour, describes the movement of air from high pressure to low pressure zones and strongly influences the transport and dispersion of air pollutants. Malaysia's tropical climate and equatorial location produce four seasonal phases: the Northeast Monsoon, the Southwest Monsoon, and two inter-monsoon periods. These phases create fluctuating wind conditions throughout the year. Stronger winds between about 5.2 and 10.3 metres per second occur during the Northeast Monsoon, while gentler winds below 7.7 metres per second dominate the Southwest Monsoon and gusts up to 15.4 metres per second are recorded along the east coast (Nizamani et al., 2018). The Northeast Monsoon from November to March originates in Siberia and brings stronger winds and heavy rainfall to the east coast and Sarawak, making these areas attractive for wind energy development (Najid et al., 2009). Wind plays a key role in energy exchange between the Earth's surface and the atmosphere and is crucial for assessing renewable energy potential (Emeis, 2014; Archer and Jacobson, 2005). Malaysia's highest recorded wind gust of 41.7 metres per second occurred in Kuching, Sarawak in 1992 and the highest mean daily wind speed of 3.8 metres per second was observed in Mersing, Johor (Najid et al., 2009). Terumbu Layang Layang has been identified as one of the most promising sites for wind energy development. Research highlights the effects of monsoon seasons and altitude on wind profiles and their implications for renewable energy and infrastructure resilience (Akorede et al., 2013; Nizamani et al., 2018). Wind speed is commonly measured using cup anemometers which have evolved to provide accurate readings for meteorological and energy applications (Pedersen and Dahlberg, 2024). Recent studies also reveal Malaysia's distinctive upper atmosphere wind pattern. Alias et al. (2024) identified a low-speed sine wave profile with three distinct peaks that contrast sharply with the volatile jet streams elsewhere. While this calmer profile improves aviation safety it poses challenges for large-scale wind energy development but still offers opportunities in selected zones.

Particulate matter with an aerodynamic diameter of less than ten micrometres, known as PM_{10}, is a significant and persistent air quality issue in Malaysia particularly in the Klang Valley (Jamalani et al., 2016). High concentrations are frequently recorded in urban areas such as Shah Alam and Klang (Ahmat et al., 2019) due to rapid urbanisation, industrial activities, and proximity to major ports (Jamalani et al., 2016). The Klang Valley is Malaysia's economic hub with a dense multiethnic population and ranks among Southeast Asia's most polluted regions with PM_{10} contributing to respiratory mortality. During non-haze periods local sources such as motor vehicle emissions and industrial operations dominate while meteorological factors especially during El Niño events further elevate pollution levels (Shaadan et al., 2015). In Petaling Jaya automotive emissions are the main source of air pollution. This is worsened by the city's mixed commercial, residential, and industrial land use where proximity of industrial zones to homes and schools makes continuous air quality monitoring essential. The Air Quality Monitoring Station at Sekolah Kebangsaan Bandar Utama sits within an industrial zone (Kar Yong and Awang, 2017) underscoring the need for efficient monitoring and regulatory measures. Selangor's position as the most populated and economically important state adds to pollution pressures especially from the growing machinery and equipment sector (Zambri and Shabri, 2024).

Wind and weather strongly influence PM_{10} dispersion. Local emissions are amplified by transboundary pollution particularly during the dry season and the Southwest Monsoon when wildfire smoke from Indonesia increases PM_{10} across Malaysia (Sentian et al., 2018). Meteorological factors such as wind speed, wind direction, temperature, and humidity control the dispersion and accumulation of pollutants creating spatial and seasonal variations in air quality (Ahmat et al., 2019; Sentian et al., 2018). Rahim et al. (2023) found that PM_{10} concentrations tend to show negative correlations with meteorological parameters such as wind speed and precipitation but positive correlations with pollutants such as nitrogen dioxide and carbon monoxide. These findings highlight the complex interactions between pollution sources and weather systems and demonstrate the need for localised control measures and regional cooperation.

https://doi.org/10.24191/mij.v7i1.8169

2.2 Application of Markov Chain Models

Having established the complex relationship between meteorological factors and air quality, the application of stochastic modelling techniques becomes essential for quantifying these dynamic interactions. Markov Chain models are widely applied stochastic tools characterised by the estimation of transition probabilities between discrete states, with broad utility in areas such as meteorology, environmental science, and health research. They are beneficial for modelling time-dependent processes due to their fundamental "memoryless" characteristic, which asserts that future states depend only on the current state, making them valuable when working with limited or noisy data (Samsuddin & Ismail, 2017; Shamshad et al., 2005). For instance, they have been used to simulate wind speed states (Aksoy et al., 2004), model daily temperature transitions to analyse heatwaves (Salleh et al., 2020), and forecast the Air Pollution Index (API) (Zakaria et al., 2019b).

The classification of wind speed and direction into discrete states is a key step in simplifying meteorological data for structured modelling, particularly in Markov Chain applications. Fujita and Sasaki (2010) categorised wind speeds into 12 states using the Beaufort scale, allowing fine-grained analysis from "calm" to "storm" conditions. Similarly, Di Giorgio et al. (2020) developed a 13-class wind speed scale, combined with eight cardinal wind direction classes, to form a Central Cardinal Point (CCP) framework, thereby enhancing the generation of meteorologically consistent synthetic time series.

Periodicity in a chain, requiring at least one self-loop to be aperiodic, influences long-term behaviour and is shared across states within the same communicating class (Cannon, 2023; Spedicato et al., 2014). A Markov Chain is irreducible if all states communicate, allowing computation of a stationary distribution and enhancing model utility (Spedicato et al., 2014). States are also classified as transient (visited finitely) or recurrent (visited infinitely), which shapes state behaviour over time (de Souza e Silva & Gail, 2000). Finally, the basis for state space decomposition in Markov models arises from the fact that communication classes and closed classes, such as absorbing states, cannot be exited (Spedicato et al., 2014).

2.3 Statistical Tests

Statistical validation ensures the reliability of Markov Chain applications in environmental modelling. Two primary tests serve this purpose effectively. Dependency tests determine whether current environmental states truly depend on previous conditions, thereby justifying the choice between first- and second-order Markov models. Temporal tests examine the stability of transition probabilities over time by comparing matrices from different periods using Chi-square analysis. When these tests confirm both dependency and temporal stability, researchers can confidently apply a single transition matrix for forecasting and synthetic data generation (Shamshad et al., 2005).

2.4 Estimating Methods

2.4.1 Count Method

The Count Method is a widely used approach in health research for estimating transition probabilities between states over time, particularly when constructing TPMs in Markov Chain models. This method involves tracking and counting observed transitions, such as those across age and gender groups, to estimate probabilities without relying on strong distributional assumptions (Samsuddin & Ismail, 2019a; Jung, 2020). For instance, Samsuddin and Ismail (2019a) applied it effectively to the SOCSO dataset, demonstrating its usefulness in capturing empirical health state transitions. Its simplicity and minimal assumptions make it especially suitable for descriptive longitudinal analyses, where the frequency of transitions is crucial. However, scholars such as Jones (2005) and Jung (2020) have cautioned that the method may be prone to bias or overestimation, particularly in small samples or sparse data situations. Nonetheless, with large datasets and sufficient historical observations, the method can yield consistent estimates and is commonly applied in areas such as environmental studies, health transitions, and credit

https://doi.org/10.24191/mij.v7i1.8169

risk modelling. Whilst the Counting Method offers simplicity and minimal assumptions, more sophisticated approaches provide additional analytical power.

2.4.2 Partial Proportional Odds Model (PPOM)

Partial Proportional Odds Models (PPOM) address the limitations of the Ordered Logit Model (OLM) by allowing coefficients to vary across thresholds when the proportional odds assumption is violated. Mayer and Foster (2015) emphasised that PPOMs capture complexities in ordered data, enhancing insights into underlying relationships, especially in economic analyses. Ara et al. (2014) noted that PPOMs are more parsimonious than the generalised OLM, which allows all coefficients to vary, whilst still preserving the response order, unlike the multinomial logit model. O'Connell and Liu (2011) affirmed that the partial proportional odds approach is widely accepted in ordinal regression due to its straightforwardness and its relation to ordinary logistic regression. Mayer and Foster (2015) employed PPOMs, which enable certain variables to vary across dependent variable categories, to examine how economic shifts impact self-rated health. This approach facilitated a nuanced understanding of how macroeconomic factors influence health outcomes.

2.4.3 Ordered Logit Model (OLM)

The Ordered Logit Model (OLM) is a regression model designed for ordinal response variables, in which the outcome categories exhibit a natural order. This model is ideal for ordinal data, as the proportional odds model is based on the cumulative probabilities of the response variable. This model is part of a larger family of cumulative ordinal models that allow for flexibility in link functions. The logit, probit, and complementary log-log functions are commonly employed (Grilli & Rampichini, 2014). In the context of estimating transition probabilities, OLMs are increasingly valued for their ability to incorporate individual characteristics, resulting in more nuanced and representative results. This model considers the dependent variable as an ordered categorical outcome, with transition probabilities influenced by covariates (Grilli & Rampichini, 2014).

2.4.4 Ordered Probit Model (OPM)

The Ordered Probit Model (OPM) applies to a qualitative dependent variable characterised by categories that possess a natural order or ranking, indicative of the magnitude of an underlying continuous variable or index, as noted in a study by Becker and Kennedy (1992). Research conducted by Johnston et al. (2020) examined the application of flexible probability density functions in ordered response models, where ordered choices are often represented using an OPM, which assumes a normal distribution of errors. In another study by Li et al. (2024), an OPM was utilised to examine the correlation between an ordinal discrete dependent variable and a collection of independent variables. Fountas et al. (2018) emphasised that OPMs are among the techniques used to identify various variables that influence injury severity outcome probabilities. The OPM accurately captures the non-linear correlation between energy use and its predictors, providing a more precise portrayal of the underlying complexity. This approach improves the analysis's robustness by reducing the impact of outliers and data noise, thereby minimising the consequences of measurement errors (Li et al., 2024). When comparing it to the OLM, the OPM was chosen in the study due to its assumption of a normal distribution for the underlying factors affecting electricity consumption, where it offers a clearer understanding of parameters and is less affected by outliers, leading to more robust estimation results, according to Li et al. (2024). The selection of appropriate estimation methods requires careful validation through statistical testing. These tests ensure that the chosen Markov Chain model accurately represents the underlying environmental processes.

https://doi.org/10.24191/mij.v7i1.8169

2.5 Monte Carlo Simulation (MCS)

Monte Carlo Simulation (MCS) serves as a powerful computational tool for handling uncertainty in environmental modelling. Its strength lies in generating probabilistic outcomes through random sampling when analytical solutions prove inadequate. In meteorological applications, MCS has demonstrated value in tropical cyclone modelling, where Schumacher et al. (2010) successfully simulated thousands of storm paths to improve wind speed forecasting accuracy. The method's flexibility allows researchers to incorporate multiple uncertainty sources simultaneously. This makes it especially suitable for validating Markov Chain predictions and generating synthetic environmental datasets that preserve the statistical properties of observed data.

3. METHODOLOGY

Fig. 1 shows the research flow adopted in this study. The process begins with data collection of wind speed and PM_{10} readings from the selected stations. The raw data are then preprocessed to handle missing values and ensure consistency. After cleaning, the variables are classified by grouping wind speed and PM_{10} into five discrete states to prepare them for stochastic modelling. In the next step, a Markov chain model is developed by defining the state space and constructing the transition probability matrix. Transition probabilities are estimated using the Counting Method as well as the PPOM, OPM, and OLM approaches to provide robust estimates. Statistical tests are applied to examine the dependency structure and temporal stationarity of the data to validate the model. Once validated, error measurement metrics such as MAE, MAPE, MSE, and RMSE are calculated to assess model performance. Monte Carlo simulation is then carried out under three different scenarios to project possible outcomes. Finally, the results are interpreted and discussed to draw conclusions on the behaviour of wind speed and PM_{10}, thereby completing the research flow.

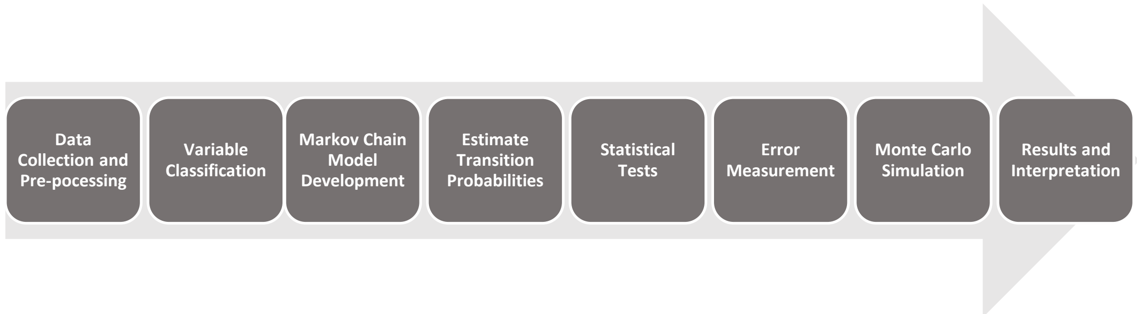

Fig. 1. Research Flow

3.1 Data

The data for this study span 10 years, from 2013 to 2023, and were collected from Petaling Jaya, an industrialised area with high traffic density. Daily observations of wind speed (m/s) and PM_{10} concentration (µg/m³) were primarily sourced from the Malaysia Meteorological Department (MyMET) and the Department of Environment (DOE). Data preprocessing was handled using R software to compute statistical analyses and address missing values, thereby providing consistent and reliable data. A discrete Markov chain was utilised to model states of wind speed and PM_{10} concentration, which were categorised into finite states. Based on the work of Fujita and Sasaki (2010) and Zakaria et al. (2019b), wind speed and PM_{10} concentration have been classified into five categories as shown in Table 1.

https://doi.org/10.24191/mij.v7i1.8169

Table 1. Classification and definition of states

State Number	State Name	
	Wind Speed	PM_{10} Concentrations
1	Gentle Breeze	Good
2	Moderate Breeze	Moderate
3	Fresh Breeze	Unhealthy
4	Strong Breeze	Very Unhealthy
5	Near Gale	Hazardous

This study considers wind speed and PM_{10} concentration as the main variables. The transition probabilities between their states are the outputs of interest. These variables are used in the four estimation methods described in Table 2.

Table 2. Variables, definitions, and roles in the study

Variable	Unit	Definition	Role in the Model
Wind Speed	m/s	Average daily wind speed measured at the monitoring station	Independent variable; categorised into five states for Markov Chain analysis
PM_{10} Concentration	μg/m³	Average daily concentration of PM_{10} (particulate matter <10 microns) measured at the monitoring station	Dependent variable; categorised into five states for Markov Chain analysis
State Number	–	Categorical representation of combined wind speed and PM_{10} conditions (States 1–5)	Defines the state space for the Markov Chain model
Transition Probability (P_{ij})	–	Probability of moving from state i at time t to state j at time $t+1$	Parameter of the Transition Probability Matrix (TPM)

3.2 Markov Chain Model

This study employed a discrete first-order Markov Chain to model transitions for wind speed and PM_{10} concentration independently. The transition probability matrix (TPM) represents the probability of transitioning from one state to another and assumes that future states depend only on the current state, not on past states. The TPM is represented as:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1k} \\ P_{21} & P_{22} & \cdots & P_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ P_{k1} & P_{k2} & \cdots & P_{kk} \end{bmatrix} \tag{1}$$

where:

(i) p_{ij}: Transition probability from state i to state j

(ii) $\sum_{j=1}^{k} p_{ij} = 1$: The sum of each row equals 1

Eqs. (2) and (3) presents the corresponding TPMs in square matrix format, where the rows represent the current states and the columns represent the possible future states.

TPM for wind speed:

$$P_{WS} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} \end{bmatrix} \tag{2}$$

https://doi.org/10.24191/mij.v7i1.8169

TPM for PM_{10}:

$$P_{PM} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} \end{bmatrix} \tag{3}$$

Based on Tagliaferri et al. (2016), the first-order Markov Chain follows the property:

$$P\left(X_{t_n} = s_j \middle| \begin{array}{c} X_{t_0} = s_{i_0}, X_{t_1} \\ = s_{i_1}, \cdots, X_{t_{m-1}} \\ = s_{i_{m-1}}, X_{t_m} = s_{i_m} \end{array}\right) = P\left(X_{t_n} = s_j \middle|, X_{t,m} = s_i\right) = P_{ij} \tag{4}$$

where:

(i) X_t: Stochastic process for wind speed and PM_{10} concentrations

(ii) S**:** State space

The memoryless property enables the construction of a TPM that controls the probability of transitions between different states. To ensure the suitability of the Markov Chain, key properties in Table 3 were analysed.

Table 3. Properties of the Markov Chain

Property	Description
Periodicity	Aperiodic: $d(x)$ = 1, Periodic: $d(x)$ > 1
Irreducibility	For any pair $x, y, P^n(x, y) > 0$
State Classification	Recurrent: $H(x, x)$ = 1, Transient: $H(x, x)$ < 1
Communicating Classes	If for all $x, y, P^n(x, y) > 0$ and $x, y, P^m(y, x) > 0$

3.3 Estimating Transition Probabilities

To estimate the transition probabilities in the Markov Chain model, four different approaches were employed and compared. These methods ranged from simple counting techniques to sophisticated ordered regression models, each offering unique advantages for modelling ordinal categorical data.

3.3.1 Counting Method

The Counting Method is a statistical technique that does not depend on assumptions, making it robust for environmental data analysis. (Birim et al., 2023). This study employed this method to construct the TPM by discretising continuous wind speed and PM_{10} concentrations into categorical states. State transitions were counted, and transition probabilities were estimated using:

$$\hat{P}_{ij} = \frac{N_{ij}}{\sum_{i,j} N_{ij}} \tag{5}$$

where:

(i) N_{ij}: Number of transitions from state *i* to state *j*

(ii) *i* = *j*: 1, 2, 3, 4, and 5

https://doi.org/10.24191/mij.v7i1.8169

3.3.2 Partial Proportional Odds Model (PPOM)

The PPOM is used to model transitions whilst relaxing the proportional odds assumption. This flexibility enables the modelling of transitions between wind speed and PM_{10} concentration categories, as well as situations in which the proportional odds assumption is only partially applicable. The PPOM can be expressed as follows:

$$P(Y_i > j) = \frac{\exp(\alpha_j + X_i\beta_j)}{1 + [\exp(\alpha_j + X_i\beta_j)]},\ \ \mathrm{j} = 1,2,\ldots,\mathrm{M} - 1 \tag{6}$$

where:

(i) $P(Y_i > j)$: The probability that the wind speed or PM_{10} concentration for the *i-th* observation exceeds the *j*-th threshold

(ii) Y_i: The ordinal dependent variable representing the wind speed or PM_{10} concentration for the *i*-th observation

(iii) j: The threshold (transition point) between categories of wind speed or PM_{10} concentration

(iv) X_i: The vector of predictors (e.g., wind speed, PM_{10}) for the *i*-th observation

(v) β_j: The category-specific coefficients capturing the effect of predictors on wind speed or PM_{10} concentration transitions for the *j*-th category

(vi) α_j: Threshold parameter specific to the *j*-th category transition for wind speed and PM_{10} concentrations

(vii) M: The total number of categories for the wind speed and PM_{10} variables

3.3.3 Ordered Probit Model (OPM)

The OPM assumes a latent continuous variable that underlies the ordinal categories of wind speed and PM_{10} concentration. The probability of observing a specific category *j* for wind speed or PM_{10} is calculated as follows:

$$P(Y_i = j) = \Phi(\tau_j - X_i\beta) - \Phi(\tau_{j-1} - X_i\beta) \tag{7}$$

where:

(i) Y_i: Ordinal categories of wind speed or PM_{10} concentrations

(ii) j: Categories of the ordinal variable (wind speed and PM_{10})

(iii) X_i: Vector of predictors such as wind speed and PM_{10}

(iv) Φ: Cumulative distribution function (CDF) of the standard normal distribution

(v) β: Coefficients indicating how each predictor affects the underlying wind speed or PM_{10} variable

(vi) τ_0: $-\infty$ and $\tau_M = \infty$ for boundary conditions

(vii) τ_j: Threshold parameters that define the cut-off between categories *j* and *j+1*

https://doi.org/10.24191/mij.v7i1.8169

3.3.4 Ordered Logit Model (OLM)

The OLM assesses the ordered relationship between wind speed and PM_{10} concentration. The OLM has a random disturbance term, ε_i that normally distributed. The vector of β parameters is estimated using the Maximum Likelihood method. The probability of each outcome $P(Y > j)$ is estimated as:

$$P(Y > j) = \frac{\exp\left(X_i\beta - K_j\right)}{1 + \left[\exp\left(X_i\beta - K_j\right)\right]}, j = 1, 2, \cdots, M - 1 \tag{8}$$

which implies:

$$P(Y_i = 1) = 1 - \frac{\exp\left(X_i\beta - K_j\right)}{1 + \left[\exp\left(X_i\beta - K_j\right)\right]}, \tag{9}$$

$$P(Y_i = j) = \frac{\exp\left(X_i\beta - K_j\right)}{1 + \left[\exp\left(X_i\beta - K_j\right)\right]} - \frac{\exp\left(X_i\beta - K_j\right)}{1 + \left[\exp\left(X_i\beta - K_j\right)\right]}, j = 2, \cdots, M - 1 \tag{10}$$

$$P(Y_i = M) = \frac{\exp\left(X_i\beta - K_{M-1}\right)}{1 + \left[\exp(X_i\beta - K_{M-1})\right]} \tag{11}$$

where:

(i) $P(Y \le j)$: Probability that the dependent variable Y falls into category j or below

(ii) K_j: Threshold parameter for the j-*th* category

(iii) β: Coefficients for the independent variables *X*

The performance of these four estimation methods was evaluated using statistical tests and error metrics to determine the most suitable approach for modelling wind speed and PM_{10} concentration transitions in the study area.

3.4 Statistical Tests

3.4.1 Dependency Test

The dependency test assesses whether the successive transitions in the data are independent or follow a Markov Chain structure. The null hypothesis assumes independence of transitions, whilst rejection of the null hypothesis supports the validity of the Markov property. The dependency statistic (α) is calculated using:

$$\alpha = 2\sum_{i,j} n_{ij} \ln\frac{p_{ij}}{p_j} \tag{12}$$

where:

(i) n_{ij}: Frequency of transitions from state i to state *j*

(ii) p_{ij}: Transition probability from state i to *j*

(iii) p_j: Marginal probability of state j, calculated as:

https://doi.org/10.24191/mij.v7i1.8169

$$p_j = \frac{\sum_i n_{ij}}{\sum_{i,j} n_{ij}} \tag{13}$$

3.4.2 Temporal Test

Temporal stationarity ensures that a single TPM is sufficient to describe the entire process, thereby simplifying long-term analysis and enabling robust forecasting. The temporal stationarity test uses the test statistic X^2, which is mathematically defined as:

$$X^2 = 2\sum_{r=1}^{\mathrm{T}}\sum_{i,j} n_{ij}(t)\ln\frac{p_{ij}(t)}{p_{ij}} \tag{14}$$

where:

(i) T: The number of time intervals (e.g., years or seasons)

(ii) $n_{ij}(t)$: The frequency of transitions from state i to state j during time interval t.

(iii) $p_{ij}(t)$: Transition probabilities from state i to state j for the t-*th* time interval

(iv) p_{ij}: Overall transition probabilities, averaged across all time intervals

3.5 Error Measurement

This study computes error metrics by flattening multi-dimensional TPMs into vectors, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). This matrix flattening simplifies comparisons between actual (y_i) and predicted $\widehat{(y_i)}$ transition probabilities (Snyman, Fox, & Bryant, 2023). By converting TPMs into a manageable vector format, this approach ensures accurate, efficient, and scalable error evaluations in Markov Chain modelling for wind speed and PM_{10} concentration analysis.

3.5.1 Mean Absolute Error (MAE)

MAE helps measure the average magnitude of errors between predicted and actual values, thereby providing an easily interpretable measure of accuracy. Lower MAE indicates better model performance in estimation.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{15}$$

where:

(i) n: Number of observations

(ii) y_i: Actual value, derived from the vector of observed TPM using the Count Method

(iii) $\hat{y}_i$: Predicted value, derived from the vector of TPM fitted models

3.5.2 Mean Absolute Percentage Error (MAPE)

MAPE helps express errors as a percentage of actual values, making it useful for assessing relative accuracy across different scales. MAPE is sensitive to small reference values, which can result in large percentage errors even if the absolute errors are small.

https://doi.org/10.24191/mij.v7i1.8169

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \tag{16}$$

where:

(i) n: Number of observations

(ii) y_i: Actual value, derived from the vector of observed TPM using the Count Method

(iii) $\hat{y}_i$: Predicted value, derived from the vector of TPM fitted models

3.5.3 Mean Square Error (MSE)

MSE calculates the average squared differences between predicted and actual values, thereby giving more weight to larger errors and emphasising model performance in high-error cases. Smaller MSE values indicate better model performance. It is also an appropriate tool for assessing the impact of outliers in prediction.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{17}$$

where,

(i) n: Number of data points

(ii) y_i: Actual value, derived from the vector of observed TPM using the Count Method

(iii) $\hat{y}_i$: Predicted value, derived from the vector of TPM fitted models

3.5.4 Root Mean Square Error (RMSE)

RMSE provides an error measurement in the same unit as the original data, making it easier to interpret relative to actual values. A lower RMSE implies a better prediction of how wind speed influences PM_{10} concentrations.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} = \sqrt{\text{MSE}} \tag{18}$$

3.6 Monte Carlo Simulation (MCS)

Monte Carlo Simulation (MCS) was applied to model stochastic behaviour in wind speed and PM_{10} concentration. For this study, the researchers assumed the initial state to be State 1 for wind speed (low wind speed) and State 5 for PM_{10} concentrations (high concentrations), reflecting the observed relationship between these variables. The TPM was used in each scenario to simulate day-to-day changes in PM_{10} levels.

The scenarios examined include:

(i) Scenario 1: Low Wind Speed and Higher PM_{10} Concentrations

(ii) Scenario 2: High Wind Speed and Lower PM_{10} Concentrations

(iii) Scenario 3: Extreme Events with Simulated Wind Speed and Simulated PM_{10}

https://doi.org/10.24191/mij.v7i1.8169

4. RESULTS AND DISCUSSION

4.1 Data Analysis

Based on Table 4, the dataset consists of 4,017 valid observations for wind speed and PM_{10} concentration, with no missing values. The wind speed ranges from –1.1 m/s to 19.2 m/s, whilst PM_{10} concentration varies between 0 and 369.67 μg/m³. The mean and median wind speed are closely aligned (7.392 m/s and 7.3 m/s, respectively), indicating a relatively symmetrical distribution. In contrast, the PM_{10} concentration has a median of 35.25 μg/m³ and a mean of 39.43 μg/m³, suggesting potential right skewness. The negative wind speed value reflects recorded flows in the opposite direction to the defined reference direction, rather than an absence of wind, whereas the positive values indicate flows in the reference direction. These statistics provide an initial understanding of the data characteristics and variability.

Table 4. Summary of descriptive statistics

	Wind Speed	PM_{10} Concentrations
Valid Data	4017	4017
Missing data	0	0
(Min, Max)	(-1.1, 19.2)	(0, 369.67)
Mean	7.392	39.43
Median	7.3	35.25

4.2 Estimating Transition Probabilities

4.2.1 Counting Method

The Counting Method was applied to define transition states for wind speed and PM_{10} concentrations, categorised into five levels based on Table 1. The TPMs were derived from transition count matrices, normalising the occurrences into probabilities. Results from Eq. (19) indicate high persistence in wind speed conditions, with a strong tendency to remain in the same state over time. PM_{10} concentrations also show stability in lower pollution levels, whilst higher pollution states exhibit greater variability. These findings highlight the gradual changes in wind speed and the dynamic nature of air quality transitions. The Markov Chain analysis in Table 5 confirms that both wind speed and PM_{10} concentration exhibit dependent, stationary behaviour with irreducible and aperiodic properties, ensuring that all states are recurrent and belong to a single communicating class. These findings validate the suitability of the Markov Chain framework for modelling wind speed transitions and their impact on air quality dynamics.

$$
\begin{gathered}
P_{WS} = \\
\begin{bmatrix}
0.350163 & 0.472313 & 0.141694 & 0.029316 & 0.006515 \\
0.111355 & 0.605436 & 0.222271 & 0.051732 & 0.009206 \\
0.126728 & 0.551843 & 0.238479 & 0.077189 & 0.005760 \\
0.145455 & 0.518182 & 0.268182 & 0.054545 & 0.013636 \\
0.117647 & 0.470588 & 0.235294 & 0.147059 & 0.029412
\end{bmatrix} \\
P_{PM} = \\
\begin{bmatrix}
0.937221 & 0.061589 & 0.000595 & 0.000289 & 0.000289 \\
0.355786 & 0.594128 & 0.043178 & 0.005181 & 0.001727 \\
0.039474 & 0.315789 & 0.565789 & 0.039474 & 0.039474 \\
0.055556 & 0.166667 & 0.277778 & 0.388889 & 0.111111 \\
0.125000 & 0.125000 & 0.125000 & 0.125000 & 0.125000
\end{bmatrix}
\end{gathered}
\tag{19}
$$

Table 5. Markov Chain Analysis

Property	Wind Speed	PM_{10} Concentration
Dependency Test	Dependent	Dependent
Temporal Test	Stationary	Stationary
Aperiodicity	Aperiodic	Aperiodic
Irreducibility	Irreducible	Irreducible
State Classification	All Recurrent	All Recurrent
Communicating Class	Single Class	Single Class

https://doi.org/10.24191/mij.v7i1.8169

4.2.2 Partial Proportional Odds Model (PPOM)

The PPOM-based TPM for wind speed (P_{WS}) and PM_{10} concentration (PPM) demonstrates structured probabilistic shifts between states, capturing variability in transitions. The model in Eq. (20) refines traditional ordinal regression by allowing non-proportional odds in specific categories, thereby enhancing predictive accuracy in understanding environmental fluctuations.

$$P_{WS} = \begin{bmatrix} 0.221341 & 0.588254 & 0.157596 & 0.035421 & 0.005318 \\ 0.159412 & 0.579169 & 0.204141 & 0.049688 & 0.007591 \\ 0.117043 & 0.546803 & 0.256181 & 0.069148 & 0.010825 \\ 0.084800 & 0.495101 & 0.309497 & 0.095188 & 0.015415 \\ 0.060827 & 0.430235 & 0.357900 & 0.129132 & 0.021907 \end{bmatrix}$$
$$P_{PM} = \begin{bmatrix} 0.932532 & 0.065233 & 0.002091 & 0.000126 & 0.000017 \\ 0.384089 & 0.568618 & 0.044127 & 0.002781 & 0.000385 \\ 0.027366 & 0.448768 & 0.458103 & 0.057308 & 0.008455 \\ 0.001268 & 0.038124 & 0.351205 & 0.450448 & 0.158956 \\ 0.000057 & 0.001789 & 0.026259 & 0.164610 & 0.807285 \end{bmatrix} \quad (20)$$

4.2.3 Ordered Probit Model (OPM)

The TPM derived from the OPM in Eq. (21) indicates structured transitions for wind speed (P_{WS}) and PM_{10} concentration (P_{PM}), capturing probabilistic movements across different states. The model effectively handles ordinal data whilst assuming a normal distribution for the underlying latent variable, thereby offering a refined approach to understanding environmental variations.

$$P_{WS} = \begin{bmatrix} 0.211724 & 0.584354 & 0.166374 & 0033465 & 0.004083 \\ 0.159455 & 0.576670 & 0.207236 & 0.049474 & 0.007166 \\ 0.116444 & 0.551845 & 0.249047 & 0.070521 & 0.012142 \\ 0.082386 & 0.512041 & 0.288772 & 0.096927 & 0.019875 \\ 0.056432 & 0.460597 & 0.323074 & 0.128461 & 0.031436 \end{bmatrix}$$
$$P_{PM} = \begin{bmatrix} 0.927639 & 0.071772 & 0.000585 & 0.000004 & 0.00000007 \\ 0.439357 & 0.509432 & 0.049137 & 0.001950 & 0.000124 \\ 0.038897 & 0.469956 & 0.386537 & 0.084633 & 0.019977 \\ 0.000370 & 0.055679 & 0.305127 & 0.310010 & 0.328814 \\ 0.000000 & 0.000687 & 0.023942 & 0.096806 & 0.878565 \end{bmatrix} \quad (21)$$

4.2.4 Ordered Logit Model (OLM)

The OLM's TPM results present similar structured transitions for wind speed and PM_{10} concentration, demonstrating a well-defined probabilistic framework. By assuming a logistic distribution for the latent variable, the OLM in Eq. (22) provides an alternative ordinal modelling approach, ensuring interpretability and robustness in predicting categorical environmental data.

$$P_{WS} = \begin{bmatrix} 0.213445 & 0.588255 & 0.157563 & 0035418 & 0.005319 \\ 0.159411 & 0.579172 & 0.204130 & 0.049693 & 0.007594 \\ 0.117021 & 0.546781 & 0.256199 & 0.069169 & 0.010831 \\ 0.084766 & 0.495033 & 0.309542 & 0.095234 & 0.015426 \\ 0.060790 & 0.430115 & 0.357956 & 0.129212 & 0.021928 \end{bmatrix}$$
$$P_{PM} = \begin{bmatrix} 0.932531 & 0.065234 & 0.002092 & 0.000126 & 0.000017 \\ 0.384100 & 0.568602 & 0.044134 & 0.002778 & 0.000385 \\ 0.027369 & 0.448759 & 0.458167 & 0.057243 & 0.008461 \\ 0.001268 & 0.038125 & 0.351448 & 0.450119 & 0.159039 \\ 0.000057 & 0.001790 & 0.026289 & 0.164494 & 0.807370 \end{bmatrix} \quad (22)$$

https://doi.org/10.24191/mij.v7i1.8169

4.3 Error Measurement

According to Tables 6 and 7, the OPM provided the best predictions for wind speed, achieving the lowest MSE, RMSE, and MAE values, thereby indicating its strong performance in capturing transition patterns. For PM_{10} concentrations, the PPOM demonstrated the best predictive accuracy, yielding the lowest MAPE of 76.56%, making it the most reliable model among those tested.

Table 6. Model Accuracy Test for Wind Speed

Model	MSE	RMSE	MAE	MAPE
PPOM	0.00257	0.05067	0.03324	24.08 %
OPM	0.00217	0.04658	0.02922	25.31 %
OLM	0.00256	0.05068	0.03325	24.08 %

Table 7. Model Accuracy Test for PM_{10} Concentrations

Model	MSE	RMSE	MAE	MAPE
PPOM	0.02727	0.16513	0.08403	76.56%
OPM	0.03657	0.19123	0.10381	81.05%
OLM	0.02727	0.16515	0.08403	76.58%

4.4 Monte Carlo Simulation (MCS)

4.4.1 Scenario 1: Low Wind Speed and Higher PM10 Concentrations

According to Fig. 2, low wind speeds contribute to higher PM_{10} concentrations as pollutants accumulate due to limited dispersion. The analysis using the TPM and MCS reveals frequent spikes in pollution levels when wind speeds remain low, often reaching hazardous states. However, occasional drops in PM_{10} indicate the influence of other factors such as varying emissions, industrial activities, and atmospheric conditions. These findings underscore the importance of considering multiple variables when addressing air quality issues.

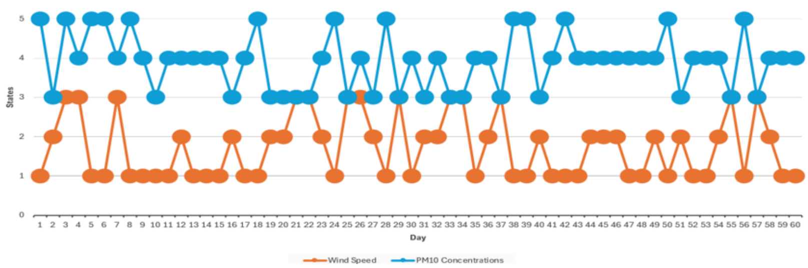

Fig. 2. Scenario 1

4.4.2 Scenario 2: High Wind Speed and Lower PM_{10} Concentrations

High wind speeds generally improve air quality by dispersing pollutants, keeping PM_{10} levels low, as shown in Fig. 3. When wind speeds are strong, air quality mostly remains in the “Good” or “Moderate” categories, reinforcing the negative correlation between wind speed and pollution. However, occasional spikes in PM_{10} suggest that external pollution sources still contribute to short-term fluctuations. Overall, strong winds play a crucial role in maintaining cleaner air by preventing the accumulation of pollutants.

https://doi.org/10.24191/mij.v7i1.8169

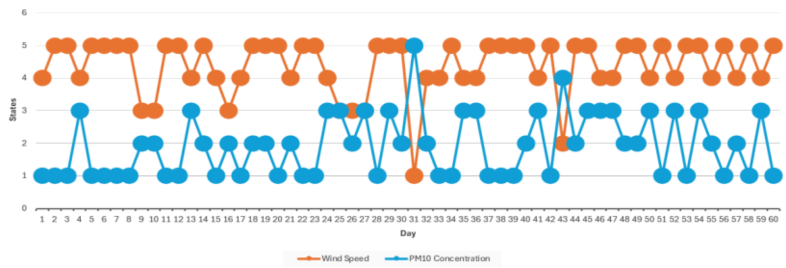

Fig. 3. Scenario 2

4.4.3 Scenario 3: Extreme Events with Simulated Wind Speed and PM_{10} Concentrations

According to Fig. 4, fluctuations in wind speed lead to dynamic changes in PM_{10} levels, with moderate winds causing variations influenced by external factors such as emissions and industrialisation. A sharp increase in wind speed significantly reduces PM_{10} concentrations, demonstrating the strong dispersal effect of higher winds. The findings confirm a negative correlation between wind speed and pollution levels, emphasising that strong winds help maintain cleaner air by preventing pollutant accumulation in populated areas.

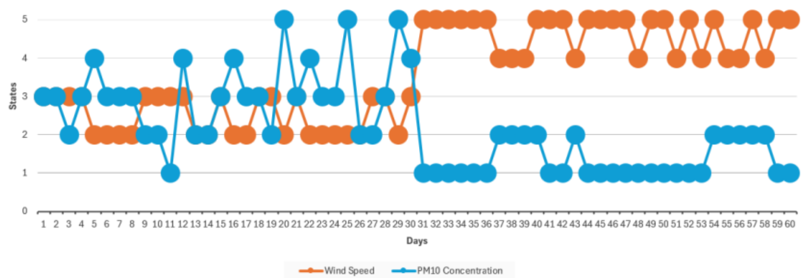

Fig. 4. Scenario 3

4.5 Discussion

This study provides evidence on the changing relationship between wind speed and PM_{10} concentrations by combining Markov Chain modelling with several estimation approaches. The high persistence observed in wind speed states, as shown by the transition probability matrix using the Count Method, suggests that atmospheric conditions remain stable over short periods. This stability can allow pollution to build up during low wind periods, showing that meteorological conditions are an important factor in how air pollution develops over time. The comparison of the three modelling approaches shows more than just differences in performance. All three methods, namely PPOM, OPM, and OLM, can describe ordered environmental data, but their predictive ability depends on the statistical nature of the variable. OPM produced better predictions for wind speed, which fits situations where the data follow a more normal pattern and are less influenced by extreme values. PPOM performed better for PM_{10} concentrations, which tend to show irregular peaks. These results suggest that model choice should follow the statistical characteristics of each environmental variable rather than applying one model to all cases.

The Monte Carlo simulation scenarios add further understanding of how these transition dynamics behave under real conditions. The scenario of high pollution during low wind speeds reflects common urban air quality events were stagnant conditions trap pollutants. The scenario of high wind speeds reducing PM_{10} levels shows the role of natural dispersion in improving air quality. By simulating these conditions, the study moves beyond description and provides estimates of the likelihood of such events.

https://doi.org/10.24191/mij.v7i1.8169

There are some limitations. The Markov Chain approach assumes that the future state depends only on the current state, which may not capture other influences such as temperature or emissions. The dataset also represents a single region and may not reflect wider seasonal or yearly patterns. While the error measures used (MAE, MAPE, MSE, RMSE) show model accuracy, they cannot fully correct for any structural issues in the models. Future research could address these points by including more variables or expanding the study to other regions.

Despite these constraints, the findings have clear implications. The study shows the value of combining Markov Chain models with flexible ordinal regression methods to better understand environmental processes. In practical terms, the results can support early warning systems and targeted actions. Knowing the likelihood of pollution episodes under different wind conditions can help authorities plan traffic or industrial controls and give timely public alerts. This integrated approach provides a strong basis for environmental policy and public health efforts aimed at reducing air quality risks.

5 CONCLUSION

This eleven-year longitudinal study represents the first comprehensive Markov Chain analysis of wind-PM_{10} relationships in the Malaysian urban context. This study has developed probability matrices for wind speed and PM_{10} concentrations using the Count Method, capturing transitions from 2013 to 2023. A normalisation process was implemented to ensure alignment with Markov model assumptions. The first-order Markov property was validated through dependency tests, which confirmed that transitions depend only on the current state, whilst temporal tests reinforced the validity of a time-homogeneous Markov model. The OPM performed best in predicting wind speed, whilst the PPOM excelled in forecasting PM_{10} concentrations. Monte Carlo simulations revealed the significant impact of wind speed on PM_{10} dispersion, demonstrating that lower wind speeds lead to higher pollution levels. To mitigate air pollution in Petaling Jaya, several recommendations emerge from this research. Firstly, enhancing air quality monitoring with advanced instruments and real-time analytics is recommended. Secondly, urban planning strategies such as green infrastructure and optimised layouts can aid pollutant dispersion. Additionally, advanced modelling techniques, including Monte Carlo simulations and Markov chains, should be utilised for accurate forecasting. Furthermore, integrating additional factors such as humidity, temperature, and industrial activity can further refine predictive models. These findings have direct implications for Malaysia's National Ambient Air Quality Standards and support evidence-based policymaking for urban air quality management. Overall, a comprehensive approach combining research, monitoring, policy measures, and public engagement is essential for effective air quality management.

6. ACKNOWLEDGEMENT/FUNDING

The authors gratefully acknowledge the support of Universiti Teknologi MARA (UiTM), Shah Alam, Selangor. This research received no external funding. The authors also greatly appreciate the reviewers for their constructive comments.

7. CONFLICT OF INTEREST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

https://doi.org/10.24191/mij.v7i1.8169

8. AUTHORS' CONTRIBUTIONS

Shamshimah Samsuddin, Azizul Hakim Nashabudin, and Muhammad Aizuddin Muhd Anwar carried out the research, and wrote and revised the article. Aini Atiqah Md Ramli conceptualised the central research idea and provided the theoretical framework. Shamshimah Samsuddin designed the research and supervised research progress. Arba'iyah Mohd Kamil anchored the review, revisions, and approved the article submission.

REFERENCES

Ahmat, H., Musa, S., Nazamid, N., & Zaharin, N. A. (2019). Classification of high and low level of PM10 concentrations in Klang and Shah Alam, Malaysia. *Malaysian Journal of Computing, 4*(2), https://doi.org/10.24191/mjoc.v4i2.6688

Akorede, M. F., Mohd Rashid, M. I., Sulaiman, M. H., Mohamed, N. B., & Ab Ghani, S. B. (2013). Appraising the viability of wind energy conversion system in the Peninsular Malaysia. Energy Conversion and Management, 76, 801–810. https://doi.org/10.1016/j.enconman.2013.08.018

Aksoy, H., Toprak, Z. F., Aytek, A., & Ünal, N. E. (2004). Stochastic generation of hourly mean wind speed data. *Renewable Energy, 29*(14), 2111–2131. https://doi.org/10.1016/j.renene.2004.03.011

Alias, M. S., Rafie, A. S. M., Hamid, M. F. A., Gires, E., & Aris, K. D. M. (2024). Wind profiles in Peninsular Malaysia: A comprehensive upper air analysis. *Sains Malaysiana, 53*(3), 487–499. https://doi.org/10.17576/jsm-2024-5303-02

Ara, R., Kearns, B., Vanhout, B. A., & Brazier, J. E. (2014). Predicting preference-based utility values using partial proportional odds models. BMC Research Notes, 7(1). https://doi.org/10.1186/1756-0500-7-438

Archer, C. L., & Jacobson, M. Z. (2005). Evaluation of global wind power. Journal of Geophysical Research D: Atmospheres, 110(12), 1–20. https://doi.org/10.1029/2004JD005462

Becker, W. E., & Kennedy, P. E. (1992). A Graphical Exposition of The Ordered Probit. In *Econometric Theory* (Vol. 8, Issue 1). https://doi.org/10.1017/s0266466600010781

Birim, N. G., Turhan, C., Atalay, A. S., & Gokcen Akkurt, G. (2023). The Influence of Meteorological Parameters on PM10: A Statistical Analysis of an Urban and Rural Environment in Izmir/Türkiye. *Atmosphere*, *14*(3). https://doi.org/10.3390/atmos14030421

Cannon, S. (2023). Irreducibility of Recombination Markov Chains in the Triangular Lattice. In SIAM Conference on Applied and Computational Discrete Algorithms (ACDA23) (pp. 98–109). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611977714.9

de Souza e Silva, E., & Gail, H. R. (2000). Transient solutions for Markov chains. In *Computational probability* (pp. 43-79). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-4828-4_3

Di Giorgio, V., Langella, R., Testa, A., Djokic, S. Z., & Zou, M. (2020, August). First order non-homogeneous Markov chain model for generation of wind speed and direction synthetic time series. In 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS) (pp. 1-6). IEEE. https://doi.org/10.1109/pmaps47429.2020.9183446

Dominick, D., Latif, M. T., Juahir, H., Aris, A. Z., & Zain, S. M. (2012). An assessment of influence of meteorological factors on PM sub (10) and NO sub (2) at selected stations in Malaysia. Sustainable Environment Research, 22(5), 305-315.

https://doi.org/10.24191/mij.v7i1.8169

Emeis, S. (2014). Current issues in wind energy meteorology. Meteorological Applications, 21(4), 803-819. https://doi.org/10.1002/met.1472

Fountas, G., Anastasopoulos, P. C., & Mannering, F. L. (2018). Analysis of vehicle accident-injury severities: A comparison of segment- versus accident-based latent class ordered probit models with class-probability functions. Analytic Methods in Accident Research, 18, 15–32. https://doi.org/10.1016/j.amar.2018.03.003

Fujita, H. ., & Sasaki, J. . (2010). Mapping of Annual Extreme Wind Speed Analysis from 12 Stations in Peninsular Malaysia. WSEAS Press.

Grilli, L., & Rampichini, C. (2014). Ordered logit model. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 4510–4513). Springer. https://doi.org/10.1007/978-94-007-0753-5_2023

Jamalani, M. A., Abdullah, A. M., Azid, A., Ramli, M. F., Baharudin, M. R., Bose, M. M., Elhadi, R. E., Youssef, K. A. A. Ben, Gnadimzadeh, A., & Gumel, D. Y. (2016). Analisis PM10 bulanan di dalam udara di Lembah Klang, Malaysia. Malaysian Journal of Analytical Sciences, 20(5), 1159–1170. https://doi.org/10.17576/mjas-2016-2005-23

Johnston, C., McDonald, J., & Quist, K. (2020). A generalized ordered Probit model. Communications in Statistics - Theory and Methods, 49(7), 1712–1729. https://doi.org/10.1080/03610926.2019.1565780

Jones, M. T. (2005). Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk; IMF Working Paper 05/219; November 1, 2005. https://doi.org/10.5089/9781451862386.001

Jung, J. (2020). Estimating Markov transition probabilities between health states using US longitudinal survey data (Vol. 6). Working paper 2020.

Kar Yong, N., & Awang, N. (2017). Quantile regression for analysing PM10 concentrations in Petaling Jaya. In *Journal of Fundamental and Applied Sciences* (Vol. 13, Issue 2). https://doi.org/10.2139/ssrn.3724408

Li, Q., Leng, Y., Yao, H., & Pei, M. (2024). Assessment of transit bus electricity consumption using a random parameters approach. *Energy*, *307*. https://doi.org/10.1016/j.energy.2024.132811

Mayer, A., & Foster, M. (2015). Understanding recession and self-rated health with the partial proportional odds model: An analysis of 26 countries. *PLOS ONE, 10*(10), e0140724. https://doi.org/10.1371/journal.pone.0140724

Najid, S. K., Zaharim, A., Razali, A. M., Zainol, M. S., Ibrahim, K., & Sopian, K. (2009). Analyzing the east coast Malaysia wind speed data. International Journal of Energy and Environment, 3(2), 53-60.

Nizamani, Z., Thang, K. C., Haider, B., & Shariff, M. (2018). Wind load effects on high rise buildings in Peninsular Malaysia. IOP Conference Series: Earth and Environmental Science, 140(1). https://doi.org/10.1088/1755-1315/140/1/012125

O'Connell, A. A., & Liu, X. (2011). Model diagnostics for proportional and partial proportional odds models. Journal of Modern Applied Statistical Methods, 10(1), 139–175. https://doi.org/10.22237/jmasm/1304223240

https://doi.org/10.24191/mij.v7i1.8169

Pedersen, T. F., & Dahlberg, J. Å. (2024). Modelling of cup anemometry and dynamic overspeeding in average wind speed measurements. Atmospheric Measurement Techniques, 17(5), 1441-1461. https://doi.org/10.5194/amt-17-1441-2024

Rahim, N. A. A. A., Noor, N. M., Jafri, I. M., Ramli, N., Kamaruddin, M. A., & Deák, G. (2023). Predicting Particulate Matter (PM10) during High Particulate Event (HPE) using Quantile Regression in Klang Valley, Malaysia. In IOP Conference Series: Earth and Environmental Science (Vol. 1216, No. 1, p. 012003). IOP Publishing. https://doi.org/10.1088/1755-1315/1216/1/012003

Salleh, N. H. M., Hasan, H., & Yunus, F. (2020). Markov chain: First step towards heat wave analysis in Malaysia. Mathematics and Statistics, 8(2), 28–35. https://doi.org/10.13189/ms.2020.081305

Samsuddin, S., & Ismail, N. (2017). Transition probabilities of health states for workers in Malaysia using a Markov chain model. AIP Conference Proceedings, 1830. https://doi.org/10.1063/1.4980930

Samsuddin, S., & Ismail, N. (2019a). Estimating markov transition probabilities between health states in the social security Malaysia (SOCSO) dataset. International Journal of Innovative Technology and Exploring Engineering, 8(11 Special issue 2), 278–282. https://doi.org/10.35940/ijitee.K1043.09811S219

Samsuddin, S., & Ismail, N. (2019b). Markov chain model and stationary test: A case study on Malaysia Social Security (SOCSO). Sains Malaysiana, 48(3), 697–701. https://doi.org/10.17576/jsm-2019-4803-24

Schumacher, A. B., DeMaria, M., Knaff, J. A., Sampson, C. R., & Brown, D. (2010, May). Objective tropical cyclone warning guidance using Monte Carlo wind speed probabilities. In Preprints, 29th Conf. on Hurricanes and Tropical Meteorology, Tucson, AZ, Amer. Meteor. Soc. D (Vol. 15).

Sentian, J., Jemain, M. A., Gabda, D., Franky, H., & Wui, J. C. H. (2018). Long-term trends and potential associated sources of particulate matter (PM10) pollution in Malaysia. WIT Transactions on Ecology and the Environment, 230, 607–618. https://doi.org/10.2495/air180571

Shaadan, N., Jemain, A. A., Latif, M. T., & Deni, S. M. (2015). Anomaly detection and assessment of PM10 functional data at several locations in the Klang Valley, Malaysia. Atmospheric Pollution Research, 6(2), 365–375. https://doi.org/10.5094/apr.2015.040

Shamshad, A., Bawadi, M. A., Wan Hussin, W. M. A., Majid, T. A., & Sanusi, S. A. M. (2005). First and second order Markov chain models for synthetic generation of wind speed time series. Energy, 30(5), 693–708. https://doi.org/10.1016/j.energy.2004.05.026

Snyman, J., Fox, C., & Bryant, D. (2023). Parsimony and the rank of a flattening matrix. *Journal of Mathematical Biology*, *86*(3). https://doi.org/10.1007/s00285-023-01875-y

Spedicato, G. A., Tae, S., Kang, S., Bhargav, Y., & Yadav, D. (2014a). The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R. https://doi.org/10.32614/cran.package.markovchain

Tagliaferri, F., Hayes, B. P., Viola, I. M., & Djokić, S. Z. (2016). Wind modelling with nested Markov chains. Journal of Wind Engineering and Industrial Aerodynamics, 157, 118-124. https://doi.org/10.1016/j.jweia.2016.08.009

Zakaria, N. N., Othman, M., Sokkalingam, R., Daud, H., Abdullah, L., & Kadir, E. A. (2019b). Markov chain model development for forecasting air pollution index of miri, Sarawak. *Sustainability (Switzerland)*, *11*(19). https://doi.org/10.3390/su11195190

https://doi.org/10.24191/mij.v7i1.8169

Zakaria, N. N., Sokkalingam, R., Daud, H., & Othman, M. (2019a). Forecasting air pollution index in Klang by Markov chain model. International Journal of Engineering and Advanced Technology, 8(6 Spec), 635-639. https://doi.org/10.3390/su11195190

Zambri, I. A., & Shabri, A. (2024). *Forecasting Monthly PM10 Concentration in Petaling Jaya using Dynamic Regression Model* (Vol. 21). https://science.utm.my/