INTERNATIONAL JOURNAL OF CONSERVATION SCIENCE

ISSN: 2067-533X Volume 16, Issue 1, 2025: 149-164 www.ijcs.ro

DOI: 10. 36868/IJCS.2025.01.09

ANALYSIS OF AIR POLLUTION IN MALAYSIA: IMPLICATIONS FOR ENVIRONMENTAL CONSERVATION USING GRANGER CAUSALITY AND PEARSON CORRELATION

Zulkifli Abd RAIS[1], Norazrin RAMLI[1,2,*], Norazian Mohamed NOOR[1,2], Hazrul Abdul HAMID[3], Ahmad Zia UL-SAUFIE[4], Mohd Khairul Nizam MAHMAD[5]

[1] Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Arau 02600, Perlis, Malaysia
[2] Sustainable Environment Research Group (SERG) Centre of Excellence Geopolymer and Green Technology (CEGeoGTech), Universiti Malaysia Perlis, Arau 02600, Perlis, Malaysia
[3] School of Distance Education, Universiti Sains Malaysia, Gelugor 11800, Penang, Malaysia
[4] Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara (UiTM), Shah Alam 40450, Selangor, Malaysia
[5] Mining and Energy Resources Academy (MERA), Jalan Kuala Ketil, Parit Panjang, 09100 Baling, Kedah

Abstract

This study investigates the relationships between air pollutants (PM_{10}, SO_2, NO_2, O_3, CO) and meteorological factors (temperature, relative humidity, wind speed) across five states in Malaysia: Seberang Perai, Shah Alam, Nilai, Larkin and Pasir Gudang. Using time-series data from 2017 to 2021, we applied Granger causality and Pearson correlation to explore the predictive relationships and linear associations between these variables. Granger causality provided insights into temporal precedence, revealing significant predictive relationships such as temperature Granger-causing PM_{10} and O_3 in Nilai and Shah Alam. Meanwhile, Pearson correlation highlighted strong linear relationships, such as the positive correlation between PM_{10} and wind speed in Shah Alam and the negative correlation between humidity and O_3 across several stations. By comparing both methods, we show how combining Granger causality with Pearson correlation can enhance environmental modelling, offering a comprehensive approach to air pollution prediction. This integration provides robust insights into the dynamics of air quality, which are critical for developing effective pollution control strategies.

***Keywords**: Air pollutants; Granger causality; Pearson correlation; Meteorological factors; Malaysia; Predictive modelling*

Introduction

Granger causality is a powerful tool in time-series analysis, widely applied across disciplines such as economics, environmental studies and neuroscience to explore the temporal relationships between variables [1]. Developed by Clive Granger in 1969, this technique operates on the principle of temporal precedence, where the past values of one variable can improve the prediction of another variable's future values [2]. However, it is cru1cial to recognize that Granger causality does not imply a direct cause-and-effect relationship. Instead, it identifies whether changes in one variable are statistically significant in forecasting another. This distinction is important as the observed relationships might be influenced by external or

* Corresponding author: norazrin@unimap.edu.my

unmeasured factors [3]. Therefore, the results of Granger causality tests must be interpreted with caution, considering potential confounding variables and the assumptions underlying the test.

A known limitation of Granger causality is its vulnerability to spurious correlations, where variables appear causally linked due to shared external influences or coincidental patterns [4]. Additionally, the technique assumes linearity, which can limit its effectiveness when dealing with non-linear relationships often present in complex systems like air quality dynamics. In scenarios where non-linear interactions dominate, relying solely on Granger causality may yield incomplete or misleading conclusions [5, 6]. These challenges underscore the need for complementary methods that can address both linear and non-linear associations, as well as enhance the interpretability of time-series data.

Pearson correlation is another foundational statistical method used to measure the strength and direction of the linear relationship between two continuous variables. It plays a pivotal role in predictive modelling, particularly in feature selection, where it helps identify key variables that are strongly associated with the target variable [7]. In addition, Pearson correlation is valuable for diagnosing multicollinearity, a condition in which two or more predictor variables are highly correlated, potentially leading to redundancy and instability in regression models. By identifying these correlations early in the modeling process, researchers can simplify their models and improve reliability and interpretability.

Despite its advantages, Pearson correlation is limited in several key respects. First, it only captures linear relationships, making it less effective in cases where variables exhibit complex, non-linear interactions [8]. In the context of environmental modeling, where air pollution and meteorological factors often display intricate, multi-dimensional dynamics, Pearson correlation may fail to detect significant associations. Moreover, Pearson correlation does not account for the temporal ordering of variables, which is crucial when analyzing time-series data. It measures simultaneous correlations but cannot determine whether changes in one variable precede or predict changes in another [7, 9]. For these reasons, Pearson correlation alone is insufficient for capturing the full complexity of relationships in time-series data and its findings should be complemented with other analytical techniques like Granger causality.

In the field of air pollution studies, both Pearson correlation and Granger causality have been employed to investigate the relationships between pollutants and meteorological parameters. Pearson correlation provides a straightforward approach to identifying linear associations, such as the link between wind speed and particulate matter (PM_{10}), or the inverse relationship between humidity and ozone levels [10]. However, these correlations provide only a static snapshot of the data, lacking the temporal dimension needed to understand how meteorological factors influence pollution over time. This is where Granger causality offers added value, as it examines the temporal order of events to determine whether changes in one variable, such as temperature, can predict future changes in pollutant concentrations, such as PM_{10} or O_3 [11]. By identifying these temporal dynamics, Granger causality helps researchers gain a more comprehensive understanding of cause-and-effect relationships in air quality studies.

Given the distinct strengths and weaknesses of both Pearson correlation and Granger causality, this study aims to compare the utility of these two methods in analyzing the relationships between air pollutant concentrations (PM_{10}, SO_2, NO_2, O_3, CO) and meteorological parameters (temperature, humidity, wind speed) across five regions in Malaysia. The regions selected for this study—Seberang Perai, Shah Alam, Nilai, Larkin and Pasir Gudang—are characterized by diverse environmental conditions, ranging from industrial areas to urban and semi-urban regions, offering a comprehensive dataset for analysis. Time-series data from 2017 to 2021 is used to examine how these statistical methods can be applied to predict air pollution trends and enhance environmental modeling.

In the context of conservation in Malaysia, integrating Granger causality and Pearson correlation into environmental research offers valuable insights into the complex relationships between air pollutants and meteorological factors, particularly in a rapidly urbanizing and industrializing nation. Malaysia faces unique environmental challenges, including seasonal haze episodes, emissions from deforestation and urban heat island effects, which have significant implications for public health and biodiversity [3]. By employing these statistical methods, air quality forecasting can be significantly improved, enabling better preparedness for haze events and reducing their impacts on vulnerable ecosystems, such as national parks and conservation areas. Furthermore, the insights gained from this approach can inform the development of targeted pollution control policies, such as stricter emissions regulations and sustainable urban planning initiatives. These strategies not only address pollution sources but also support long-term environmental sustainability. Additionally, understanding the interplay between pollutants and meteorological factors can help anticipate how changes in climatic conditions may affect air quality, guiding conservation efforts such as reforestation to enhance carbon sequestration and mitigate pollution levels. Enhanced air quality monitoring also benefits Malaysia's eco-tourism sector, which depends on clean air and well-preserved natural environments to attract visitors. By leveraging such an analytical framework, Malaysia can better balance economic development with environmental conservation, fostering a healthier and more sustainable future for its people and ecosystems [12].

Experimental part

Materials and Methods

The study focuses on five air quality monitoring stations located across different regions of Malaysia: Seberang Perai, Shah Alam, Nilai, Larkin and Pasir Gudang. These stations were selected due to their varied environmental conditions, which include industrial, urban and semi-urban areas. The coordinates of the stations are listed in Table 1, covering northern, central and southern regions of the country. This distribution allows for an assessment of air pollution trends in diverse climatic and topographical conditions [13].

Table 1. Latitude and Longitude for five monitoring station's location

Station ID	State	Location	Coordinate
CA07P	Seberang Perai, Pulang Pinang	Sek. Keb. Cenderawasih, Taman Inderawasih, Perai	N05° 23.470' E100° 23.213'
CA20B	Shah Alam, Selangor	Sek. Keb. Taman Tun Dr. Ismail Jaya, Shah Alam	N03° 06.287' E101° 33.368'
CA23N	Nilai, Negeri Sembilan	Taman Semarak (Phase II), Nilai	N02° 49.246' E101° 48.877'
CA33J	Larkin, Johor	Teacher Education Temenggong Ibrahim Campus, Larkin, Johor Bahru	N01° 29.815' E103° 43.617'
CA34J	Pasir Gudang, Johor	Sek. Men. Keb. Pasir Gudang 2, Pasir Gudang, Johor Bahru	N01° 28.225' E103° 53.637'

The regions represented in this study exhibit distinct pollution sources. Shah Alam, for example, is an urban centre with heavy vehicular traffic, while Pasir Gudang and Larkin are industrial hubs [14]. By contrast, Seberang Perai and Nilai represent a mix of residential and industrial activities. These varying conditions make the locations suitable for a broad analysis of air quality parameters and their relationship with meteorological factors.

Site Description

The five air quality monitoring stations are strategically distributed across northern, central and southern Malaysia, ensuring a comprehensive evaluation of air quality in different climatic and pollution contexts as shown in figure 1.

Seberang Perai, located in Pulau Pinang, is a rapidly urbanizing area characterized by a combination of industrial and residential zones. It experiences moderate industrial emissions coupled with vehicular pollution, making it a prime location for studying the interplay between urban pollution and meteorological conditions.

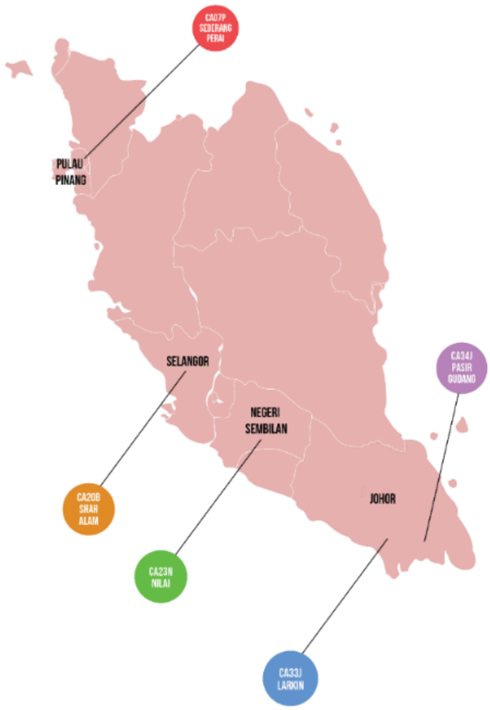


Fig 1. Location of research area

Similarly, Shah Alam in Selangor, a highly urbanized commercial and industrial hub within the Klang Valley, faces significant air quality challenges due to heavy traffic emissions, construction activities and industrial pollutants. Nilai in Negeri Sembilan offers a contrasting semi-urban environment, where both industrial and residential activities contribute to local pollution levels [15]. Additionally, Nilai is affected by transboundary pollution from neighbouring regions, providing a diverse dataset to analyze pollutant dispersion under varying meteorological influences. In Johor, the urban area of Larkin, near Johor Bahru, faces air quality issues driven by moderate industrial activity and vehicular emissions, with these challenges escalating alongside urbanization and industrialization. Finally, Pasir Gudang in Johor, one of the most heavily industrialized regions in Malaysia, is dominated by petrochemical industries. The air quality in this region is primarily shaped by emissions from industrial processes, making it a crucial site for investigating the impact of industrial pollution in relation to meteorological factors.

The geographical diversity of these locations allows for a broad analysis of air pollution trends under varying environmental conditions, from coastal to inland regions and across different land use types. The details and coordinates of the stations are summarized in Table 1.

Data Collection and Preliminary Data Processing

Data on five key air pollutants— PM_{10}, SO_2, NO_2, O_3 and CO —and three meteorological parameters—temperature, humidity and wind speed—were collected from the Department of Environment Malaysia. Monthly averages for each pollutant and meteorological parameter were computed from raw daily data spanning from 2017 to 2021. The corresponding symbols and units for the air quality parameters was present in Table 2.

The dataset underwent preprocessing to handle missing values and outliers. Statistical techniques, including data imputation and outlier detection, were used to clean the data, ensuring that the analysis remained robust [13]. The descriptive statistics of the data—such as mean, standard deviation, skewness and kurtosis—were calculated to summarize the characteristics of each variable at each station.

Table 2. Air Quality Parameters with Corresponding Symbols and Units

Parameter	Symbol	Unit
Particulate Matter < 10µm	PM_{10}	µg/m^3
Sulfur Dioxide	SO_2	ppm
Nitrogen Dioxide	NO_2	ppm
Oxygen	O_3	ppm
Carbon Dioxide	CO	ppm
Temperature	T	°C
Relative Humidity	RH	%
Wind Speed	WS	m/s

Granger Causality Analysis

Granger causality was used to assess whether one-time series could predict another. The test was performed on pairs of air pollutant concentrations and meteorological variables. Specifically, the test examined whether the historical values of meteorological factors like temperature or wind speed could help predict future pollutant concentrations. Granger causality is based on the Vector Autoregression (VAR) model, where each variable is regressed against its own past values and the past values of the other variables in the system.

The null hypothesis of the Granger causality test assumes that no causality exists between the tested variables. If the null hypothesis is rejected at a significant level, it indicates that one variable Granger-causes the other. The analysis was performed using the statistical software EViews, with results interpreted using graphical depictions of causality

The analysis is grounded in the theoretical framework of autoregressive models, which explore how a variable's current value is influenced by its past values. In the context of Granger causality, these models are used to determine whether the past values of one variable (X) significantly improve the prediction of another variable (Y), beyond what Y's own history can predict. While Granger causality does not imply true causation, it serves as a powerful statistical tool for identifying predictive relationships in time series data [2, 16].The test involves comparing models that include and exclude the past values of X to evaluate whether adding X enhances the prediction of Y. This approach is crucial in time series analysis, where understanding the temporal relationships and directional influences between variables is essential for accurate forecasting [17]:

$$y_i = \alpha_0 + \sum_{j=1}^{m} \alpha_j y_{i-j} + \sum_{j=1}^{m} \beta_j x_{i-j} + \varepsilon_i \tag{1}$$

In the final phase of the study, Granger causality tests were applied to examine the causal relationships between air pollutants and meteorological parameters. Before conducting these tests, the optimal lag length for the analysis was determined using the Akaike Information Criterion (AIC) in EViews software, ensuring a balance between model complexity and goodness of fit.

With the selected lag order, an unrestricted Vector Autoregression (VAR) model was estimated [18]. The Granger causality tests were then used to assess whether the past values of one variable provided predictive information about another. A p-value of 0.05 or lower led to the rejection of the null hypothesis of no Granger causality, indicating a significant causal relationship [5]. The results were visualized to illustrate the direction and strength of these causal connections across different monitoring stations.

Pearson Correlation Analysis

The application of Pearson correlation in predictive modelling serves a crucial function in feature selection. By analyzing the correlations between predictor variables and the target variable, researchers can identify relevant features for inclusion in predictive models [7]. High correlation coefficients suggest potential predictors that may significantly improve the overall performance of the model. Additionally, Pearson correlation is instrumental in assessing multicollinearity, a condition characterized by elevated correlations among predictor variables, which can lead to instability and redundancy in predictive models [7, 19]. Identifying and addressing multicollinearity through correlation analysis enhances both the reliability and interpretability of predictive models.

In air pollution research, the incorporation of Pearson correlation into predictive models is a valuable tool for elucidating the complex relationships within air quality data. Its capability to quantify linear associations between variables provides critical insights into the dynamics of air quality [10].

To examine the linear relationships between each pollutant and meteorological factor, Pearson correlation coefficients were calculated. This measure assesses the strength of the association between two continuous variables, yielding values that range from -1 (indicating a perfect negative correlation) to +1 (indicating a perfect positive correlation), with a value of zero signifying no linear relationship [20]. A Pearson correlation matrix was constructed for each monitoring station, providing a clear view of the pairwise correlations among pollutants and meteorological factors. The analysis was conducted using IBM SPSS Statistical Software Version 29. To fully leverage the potential of Pearson correlation, a careful and multidimensional approach is required, taking into account several key factors. A fundamental aspect of Pearson correlation is its role in variable selection and dimensionality reduction. Although it aids in identifying relevant variables, managing a large number of predictors necessitates the use of techniques such as Principal Component Analysis (PCA) or other feature selection methods. This strategic approach streamlines the modeling process and effectively addresses multicollinearity issues, leading to increased computational efficiency [21].

In air quality studies, where distinct temporal patterns are prevalent, temporal considerations are paramount. Analyzing correlations across various time intervals allows models to capture seasonality, diurnal variations and long-term trends, ensuring that predictive models reflect the dynamic nature of air quality data [22]. It is also important to recognize the limitations of Pearson correlation in detecting non-linear relationships; thus, the integration of complementary techniques is essential. Approaches such as polynomial regression or machine learning algorithms, which excel at revealing complex non-linear patterns, can enhance the understanding of relationships between variables, contributing to the overall robustness of predictive models [23].

Furthermore, considering causality is vital. While Pearson correlation identifies associations, it does not confirm causation. Employing techniques such as Granger causality testing or structural equation modeling can facilitate a more nuanced exploration of causal relationships among variables, thereby enriching the predictive modeling process [9].

Statistical Software and Tools

All analyses were performed using industry-standard statistical software tools. The EViews software was employed for Granger causality testing due to its advanced capabilities in time-series analysis, particularly for handling multivariate datasets. IBM SPSS Statistics (Version 29) was used for Pearson correlation analysis and descriptive statistics, as well as for data preprocessing tasks such as handling missing values, detecting outliers and generating visualizations.

Data Visualization and Interpretation

Graphical representations of the results were critical for interpreting the complex relationships between pollutants and meteorological factors. Time-series plots were used to depict trends in pollutant concentrations and meteorological parameters over the study period, while correlation heatmaps provided a clear visualization of the strength of linear relationships. Causality diagrams, derived from the Granger causality tests, were employed to map the direction and strength of predictive relationships, helping to uncover the temporal dynamics underlying the air quality data.

Results and Discussion

Descriptive Analysis

The descriptive statistics for each pollutant and meteorological parameter were computed to establish a foundational understanding of the data. Table 3 presents a summary of the average concentrations of pollutants (PM_{10}, SO_2, NO_2, O and, CO) and the meteorological parameters (temperature, relative humidity, wind speed) across the five monitoring stations with 54 of the sample size.

The analysis shows that PM_{10} levels were consistently highest in Nilai, an industrial area, with a mean concentration of 34.23μg/m³. In contrast, Seberang Perai recorded the lowest PM_{10} mean at 25.39μg/m³. A noticeable spike in PM_{10} levels occurred across all stations in October 2019, likely due to the influence of industrial emissions and meteorological conditions, particularly during periods of lower wind speeds and higher temperatures, which limited the dispersion of pollutants.

Table 3 presents descriptive statistics for various locations from 2017 to 2021, illustrating how broader environmental factors influence air quality. The 2020 Movement Control Order (MCO) notably led to a significant reduction in NO_2 levels in urban areas like Shah Alam, attributed to decreased transportation and industrial activity [24, 25]. However, the data also highlight the susceptibility of these areas to extreme pollution events, such as the notable spikes in PM_{10} and SO_2 concentrations recorded in October 2019, driven largely by seasonal weather conditions and industrial processes. These observations underscore the need for ongoing monitoring and targeted interventions to effectively manage air pollution, particularly in urban and industrial areas [26]. The variation in pollutant levels across different locations and time periods emphasizes the importance of developing localized strategies to address specific pollution sources and mitigate the effects of extreme pollution events on the environment.

Table 3. The descriptive analysis of the concentrations from 2017 to 2021

Station	Concentration	Seberang Perai	Shah Alam	Nilai	Larkin	Pasir Gudang
Minimum, μg/m³	PM_{10}	16.185	**20.991**	20.905	17.579	15.502
Maximum, μg/m³		43.696	89.365	**102.887**	72.508	63.236
Mean, μg/m³		25.387	32.667	**34.234**	28.194	26.039
Standard Deviation, μg/m³		6.758	10.138	**12.615**	8.336	8.497
Coefficient of variation		0.266	0.310	**0.368**	0.296	0.326
Skewness		1.153	3.570	**3.379**	3.005	1.787

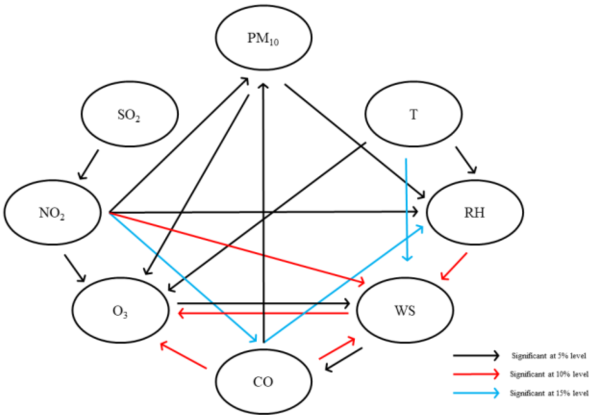
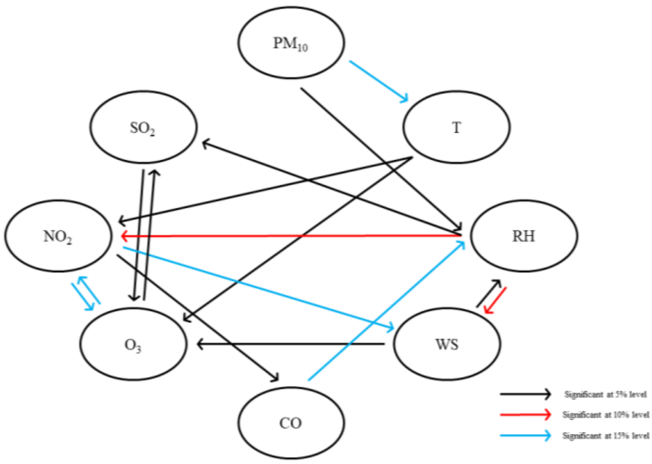
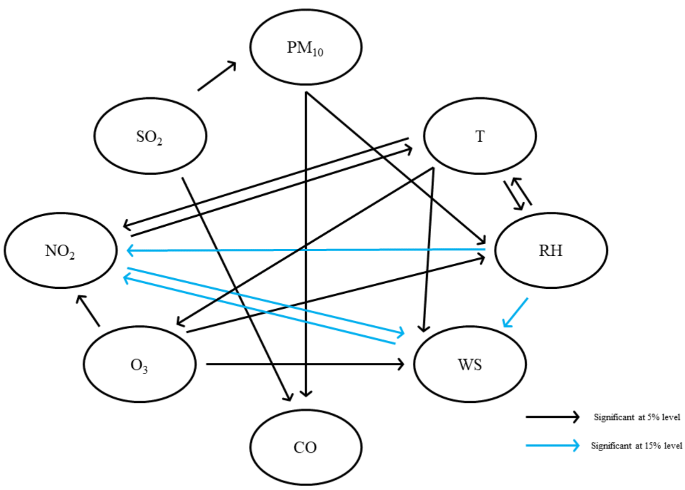
Station	Concentration	Seberang Perai	Shah Alam	Nilai	Larkin	Pasir Gudang
Kurtosis		1.089	**18.262**	16.227	14.350	5.600
Minimum, ppm	SO_2	0.001	0.001	0.001	0.001	0.001
Maximum, ppm		0.002	0.002	0.004	0.003	**0.010**
Mean, ppm		0.001	0.001	0.001	**0.002**	**0.002**
Standard Deviation, ppm		0.0003	0.0002	0.0005	0.0005	**0.0013**
Coefficient of variation		0.3	0.2	0.5	0.25	**0.65**
Skewness		0.473	0.187	4.234	0.063	**4.269**
Kurtosis		-0.757	0.746	23.370	0.223	**23.504**
Minimum, ppm	NO_2	0.005	**0.007**	0.006	0.004	0.004
Maximum, ppm		0.015	**0.022**	0.018	0.018	0.018
Mean, ppm		0.009	**0.016**	0.013	0.012	0.011
Standard Deviation, ppm		0.002	0.003	0.003	0.003	**0.004**
Coefficient of variation		0.22	0.19	0.23	0.25	**0.36**
Skewness		**0.483**	-0.408	0.040	0.036	-0.495
Kurtosis		**1.314**	0.290	-0.845	-0.788	-0.863
Minimum, ppm	O_3	0.004	**0.014**	0.005	0.009	0.008
Maximum, ppm		0.031	**0.038**	0.024	0.026	0.025
Mean, ppm		0.016	**0.020**	0.010	0.015	0.014
Standard Deviation, ppm		**0.007**	0.005	0.004	0.004	0.004
Coefficient of variation		**0.44**	0.25	0.4	0.27	0.29
Skewness		0.075	1.306	**1.451**	0.667	0.517
Kurtosis		-0.734	1.860	**2.074**	0.863	-0.032
Minimum, ppm	CO	0.473	**0.531**	0.372	0.188	0.410
Maximum, ppm		1.023	**1.263**	1.184	0.975	0.905
Mean, ppm		0.715	**0.829**	0.594	0.601	0.657
Standard Deviation, ppm		0.101	0.150	0.127	**0.218**	0.122
Coefficient of variation		0.14	0.18	0.21	**0.36**	0.19
Skewness		0.348	0.240	**1.888**	-0.152	0.080
Kurtosis		0.678	-0.158	**7.703**	-1.313	-0.871
Minimum, °C	T	26.219	26.125	26.093	26.053	**27.039**
Maximum, °C		29.768	28.644	29.174	28.977	**31.295**
Mean, °C		27.980	27.547	27.900	27.559	**29.454**
Standard Deviation, ppm		0.746	0.682	0.652	0.571	**1.108**
Coefficient of variation		0.027	0.025	0.023	0.021	**0.038**
Skewness		-0.032	-0.549	-0.295	**0.133**	-0.483
Kurtosis		-0.345	-0.749	-0.242	**0.311**	-0.523
Minimum, %	RH	67.083	70.400	70.361	**77.250**	72.416
Maximum, %		85.768	84.457	85.069	**88.491**	85.296
Mean, %		78.905	78.768	78.539	**84.397**	79.570
Standard Deviation, ppm		3.978	3.463	**3.469**	2.962	2.881
Coefficient of variation		**0.050**	0.044	0.044	0.035	0.036
Skewness		-0.877	-0.301	-0.451	-0.796	**-0.278**
Kurtosis		**0.782**	-0.600	-0.205	-0.308	-0.253
Minimum, m/s	WS	**0.973**	0.638	0.408	0.589	0.710
Maximum, m/s		1.550	1.204	**2.021**	1.978	1.689
Mean, m/s		**1.343**	0.864	0.945	0.900	1.151
Standard Deviation, ppm		0.155	0.136	**0.363**	0.303	0.246
Coefficient of variation		0.115	0.157	**0.384**	0.337	0.214
Skewness		-0.516	0.485	1.587	**1.707**	0.604
Kurtosis		-0.682	-0.270	2.411	**2.594**	-0.545

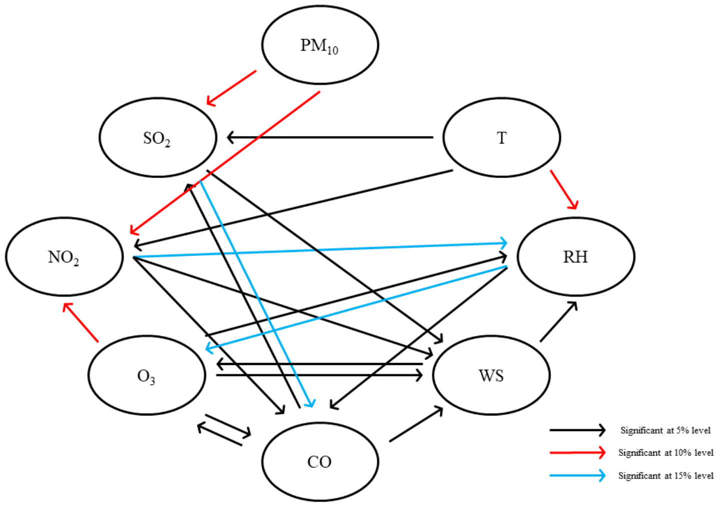
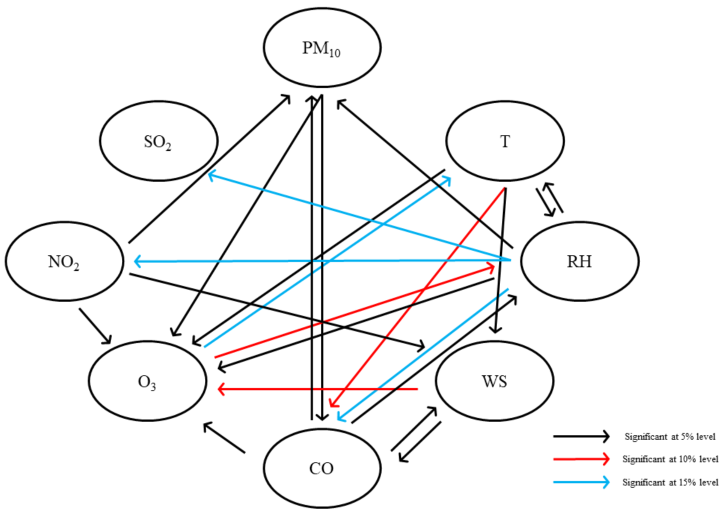
**Note: Bold denotes referred to the significant of the concentration

Granger Causality Analysis

Granger causality occurs when the past and present values of a specific characteristic provide valuable information for predicting its future behavior in a time series. To explore the causal relationships among the parameters PM_{10}, SO_2, NO_2, O_3 and CO, a Granger causality test was conducted. The results of these tests for each monitoring station are presented in Table 4.

Table 4. The result Granger Causality Diagram for all the monitoring stations

Monitoring Stations	Granger Causality Diagram
Seberang Perai, Pulau Pinang	
Shah Alam, Selangor	
Nilai, Seremban	

Monitoring Stations	**Granger Causality Diagram**
Pasir Gudang, Johor	
Larkin, Johor	
*Black lines indicate causality relationships with p-values less than 0.05 *Red lines represent relationships with p-values between 0.05 and 0.10 *Blue lines indicate causality relationships with p-value more than 0.10	

In the context of conservation efforts in Malaysia, the findings on the relationship between meteorological conditions and pollutant levels in suburban, urban and industrial regions underscore the importance of location-specific strategies for air quality management. In Seberang Perai, where ozone (O_3) levels are strongly influenced by temperature, addressing heat-related pollution could be critical for safeguarding agricultural activities and local ecosystems. Similarly, the moderate influence of temperature on SO_2 and NO_2 concentrations suggests that mitigating temperature fluctuations through urban greening or improved land-use planning could contribute to better air quality [27]. Shah Alam, an urbanized area, exhibits a strong link between relative humidity and NO_2 levels, as well as a moderate relationship between humidity and O_3. These findings suggest that increasing vegetation cover to regulate local humidity levels could not only improve air quality but also support urban biodiversity [28].

In industrial areas such as Nilai, Larkin and Pasir Gudang, the intricate interactions between pollutants and meteorological factors highlight the need for targeted pollution control measures. For instance, in Nilai, the two-way relationship between SO_2 and PM_{10}, influenced by wind speed, calls for stricter emissions regulations for industries and better urban planning to optimize air circulation. In Larkin, where NO_2 and CO drive changes in PM_{10} and relative humidity influences both PM_{10} and O_3, strategies like afforestation and improved public transportation could help reduce pollutant levels and enhance air quality. Pasir Gudang presents even more complex dynamics, with pollutants such as PM_{10} impacting multiple others, including NO_2, O_3 and CO, alongside the influence of wind speed and O_3 on SO_2 levels [29]. This underscores the need for integrated air quality monitoring systems and adaptive pollution control policies to address the multifaceted interactions in such industrial hubs.

By incorporating these insights into conservation planning, Malaysia can tailor its environmental management strategies to the unique challenges of each location, ensuring the protection of its ecosystems and promoting sustainable urban and industrial development [3].

Pearson Correlation Analysis

In Malaysia, air pollution is a significant concern in both urban and industrial areas, where various sources of emissions adversely affect air quality and the surrounding ecosystems as shown in Table 5. Urban areas like Shah Alam and Seberang Perai exhibit distinct challenges due to the high volume of vehicular emissions and urban activities. In Shah Alam, the strong correlations between carbon monoxide (CO) and nitrogen dioxide (NO_2) ($r = 0.815$) and particulate matter (PM_{10}) with CO ($r = 0.638$) indicate that traffic-related pollutants are a major contributor to the region's air quality issues [30]. These pollutants pose a threat to urban green spaces and water bodies, which are essential for regulating temperatures and supporting biodiversity. To address these concerns, increasing urban tree cover, enhancing public transportation systems and promoting the use of low-emission vehicles can significantly reduce pollution and improve environmental quality. Similarly, in Seberang Perai, moderate correlations between PM_{10} and NO_2 ($r = 0.604$) and a strong relationship between ozone (O_3) and NO_2 ($r = 0.642$) underscore the impact of both vehicular and industrial emissions on air quality. These pollutants endanger the region's mangrove ecosystems and agricultural lands, which are critical for biodiversity conservation and carbon sequestration. Sustainable transportation and cleaner industrial practices are necessary to reduce NO_2 emissions and safeguard these vital ecosystems.

In contrast, industrial areas such as Nilai, Larkin and Pasir Gudang face pollution challenges stemming primarily from industrial and vehicular emissions. In Nilai, the exceptionally strong correlation between CO and PM_{10} ($r = 0.850$) reflects the significant influence of both industrial and vehicular emissions on the region’s air quality. The pollution from these sources threatens nearby forests and rivers, which are important habitats for a wide range of species. To mitigate these effects, stricter regulations on industrial emissions and the establishment of green buffers around industrial zones would not only reduce air pollution but also help preserve these ecosystems. Larkin exhibits moderate correlations between PM_{10} and NO_2 ($r = 0.632$) and CO ($r = 0.475$), indicating that urban emissions significantly contribute to air pollution. These pollutants degrade nearby green spaces and biodiversity, emphasizing the need for sustainable urban planning, such as the creation of low-emission zones and the expansion of green infrastructure. Finally, Pasir Gudang shows moderate correlations between PM_{10} and sulfur dioxide (SO_2) ($r = 0.576$) and NO_2 ($r = 0.569$), highlighting the influence of industrial activities on air quality. The region’s coastal and marine ecosystems are particularly vulnerable to pollution, which can disrupt biodiversity and weaken natural coastal defenses. Strengthening air quality monitoring systems and implementing sustainable industrial practices are essential to protect these ecosystems and ensure their long-term resilience.

Overall, the findings across these urban and industrial areas demonstrate the complex relationship between air pollution, human activities and environmental degradation. Tailored interventions that address the specific sources of pollution in each region are essential to reduce emissions, protect ecosystem and promote sustainable development across Malaysia's urban and industrial landscapes.

Table 5. Correlation coefficient matrix of air pollutants concentration and meteorological parameters

Seberang Perai, Pulau Pinang

	PM_{10}	SO_2	NO_2	O_3	CO	T	RH	WS
PM_{10}	1							
SO_2	0.343	1						
NO_2	0.604	0.499	1					
O_3	0.572	0.533	0.642	1				
CO	0.353	0.423	0.681	0.430	1			
T	0.554	0.338	0.420	0.770	0.229	1		
RH	-0.426	0.018	-0.034	-0.491	0.098	-0.592	1	
WS	0.464	-0.040	0.140	0.403	-0.100	0.581	-0.540	1

1
0.8
0.4
0
-0.4
-0.8
-1

Shah Alam, Selangor

	PM_{10}	SO_2	NO_2	O_3	CO	T	RH	WS
PM_{10}	1							
SO_2	0.285	1						
NO_2	0.589	0.577	1					
O_3	0.332	-0.063	-0.004	1				
CO	0.638	0.420	0.815	0.038	1			
T	0.243	0.185	0.209	-0.096	0.178	1		
RH	-0.401	-0.533	-0.341	-0.312	-0.188	-0.173	1	
WS	0.240	0.316	-0.016	0.293	-0.147	0.165	-0.688	1

1
0.8
0.4
0
-0.4
-0.8
-1

Nilai, Negeri Sembilan

	PM_{10}	SO_2	NO_2	O_3	CO	T	RH	WS
PM_{10}	1							
SO_2	0.249	1						
NO_2	0.499	-0.011	1					
O_3	0.125	-0.023	0.422	1				
CO	0.850	0.324	0.644	0.208	1			
T	-0.227	0.025	-0.367	0.264	-0.316	1		
RH	-0.181	-0.148	0.233	-0.355	-0.041	-0.538	1	
WS	0.047	0.088	-0.425	-0.149	-0.105	0.018	-0.499	1

1
0.8
0.4
0
-0.4
-0.8
-1

Pasir Gudang, Johor

	PM_{10}	SO_2	NO_2	O_3	CO	T	RH	WS
PM_{10}	1							
SO_2	0.576	1						
NO_2	0.569	0.602	1					
O_3	0.201	-0.168	-0.520	1				
CO	0.498	0.423	0.710	-0.146	1			
T	0.013	-0.142	-0.686	0.684	-0.521	1		
RH	-0.573	-0.387	0.007	-0.521	0.062	-0.463	1	
WS	0.222	-0.115	0.056	0.011	-0.172	0.218	-0.286	1

1
0.8
0.4
0
-0.4
-0.8
-1

Larkin, Johor

	PM_{10}	SO_2	NO_2	O_3	CO	T	RH	WS
PM_{10}	1							
SO_2	0.231	1						
NO_2	0.632	0.229	1					
O_3	0.341	-0.059	-0.286	1				
CO	0.475	-0.127	0.710	-0.134	1			
T	-0.057	0.039	-0.430	0.590	-0.242	1		
RH	-0.408	0.031	0.133	-0.756	-0.011	-0.489	1	
WS	0.370	0.397	0.158	0.279	0.054	0.234	-0.490	1

1
0.8
0.4
0
-0.4
-0.8
-1

Comparison of Granger Causality and Pearson Correlation

Granger causality and Pearson correlation are widely used tools in analyzing relationships between variables, but they serve different purposes. Pearson correlation measures the strength and direction of linear relationships between two variables at a given moment. For instance, in Shah Alam, a significant positive correlation was found between temperature and PM_{10} concentrations, suggesting that higher temperatures are linked with increased particulate matter. However, Pearson correlation does not indicate whether one variable influences the other or if time-dependent relationships exist.

Granger causality focuses on identifying temporal relationships by testing whether past values of one variable can predict future values of another. This makes it particularly valuable for understanding dynamic environmental interactions. In the Shah Alam study, Granger causality revealed that temperature could predict future PM_{10} concentrations. This method provides insights into how changes in one factor precede changes in another, which is critical for time-series data, though it does not measure the strength of the relationship.

By integrating both approaches, a more comprehensive understanding of air quality dynamics is achieved. Pearson correlation highlights immediate linear relationships, while Granger causality identifies predictive temporal links [31]. This dual perspective is crucial for areas like Shah Alam, where air quality is influenced by complex interactions between meteorological factors and pollutants.

Applying these methods in Malaysia can significantly enhance strategies for addressing air pollution. Understanding both the present and predictive relationships between weather patterns and pollutants allows for better forecasting and management [20]. This is especially important during haze episodes, where proactive measures can mitigate public health impacts and support environmental conservation efforts.

Conclusions

In conclusion, this study effectively illustrates the utility of combining Granger causality and Pearson correlation to explore the complex interactions between air pollutants and meteorological factors across five regions in Malaysia from 2017 to 2021. The Granger causality analysis revealed significant predictive relationships, such as temperature Granger-causing PM_{10} and O_3 in Nilai and Shah Alam and relative humidity Granger-causing NO_2 in Shah Alam. These temporal dynamics highlight how changes in meteorological factors can precede and influence air pollutant levels.

The Pearson correlation analysis provided valuable linear associations, such as the strong positive correlation between wind speed and PM_{10} ($r = 0.75$) in Shah Alam and the inverse relationship between humidity and O_3 across multiple stations, with r-values ranging from -0.49 to -0.76. These correlations emphasize how meteorological conditions directly affect pollutant concentrations.

The integration of Granger causality and Pearson correlation provides a more nuanced understanding of air quality dynamics across Malaysia's diverse regions, including urban centers like Shah Alam and industrial areas such as Pasir Gudang. For example, historical data reveals significant spikes in PM_{10} concentrations during haze episodes, with one notable incident occurring in October 2019, where several monitoring stations recorded levels exceeding 150 $\mu g/m^3$, a level classified as unhealthy by the Department of Environment Malaysia. By using Granger causality, it was possible to identify temporal patterns and potential predictors of such events, while Pearson correlation highlighted immediate relationships between meteorological variables like temperature and humidity with particulate matter concentrations.

This dual-method approach underscores the importance of continuous air quality monitoring and tailored intervention strategies. The predictive insights from Granger causality are particularly valuable for proactive air pollution management, enabling policymakers to anticipate and mitigate extreme pollution events. For instance, localized strategies, such as stricter emissions controls in industrial zones or public advisories during peak pollution periods, can be developed based on region-specific data.

Overall, this integrated analytical framework serves as a foundation for enhancing Malaysia's air quality management efforts. It aligns with the nation's goals for sustainable development and public health protection by providing actionable insights into the complex interplay of environmental and meteorological factors. Such an approach not only improves current mitigation strategies but also supports long-term planning to address the growing challenges of air pollution in Malaysia's rapidly urbanizing and industrializing landscape.

Acknowledgments

This research was supported by Ministry of Higher Education Malaysia (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2023/TK08/UNIMAP/02/24) which made this research possible. Sincere thanks are also extended to the Department of Environment (DOE) Malaysia for granting access to the comprehensive and detailed datasets that were essential to the analysis and findings.

References

[1] D. Yang, H. Chen, Y. Song, Z. Gong, *Granger Causality for Multivariate Time Series Classification*, **Proceedings - 2017 IEEE International Conference on Big Knowledge, ICBK 2017**, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 103–110. DOI: 10.1109/ICBK.2017.36.

[2] L. Barnett, A.K. Seth, *The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference*, **Journal of Neuroscience Methods**, **223**, 2021, pp. 50–68, DOI: 10.1016/j.jneumeth.2013.10.018.

[3] N. Amila, M. Ridzuan, N. Noor, N. Alis, A. Rahim, I. Amani, M. Jafri, A. Zia, U. Saufie, M. Remy, R. Mohd, A. Zainol, M. Anuar Kamaruddin, G. Deak, *Modeling of Particulate Matter (PM_{10}) During High Particulate Event (HPE) In Klang Valley, Malaysia*, **International Journal of Conservation Science**, **13**(3), 2022, pp. 1065-1078.

[4] J.M. Rohrer, *Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data*, **Advances in Methods and Practices in Psychological Science**, **1**(1), 2020, pp. 27–42, DOI: 10.1177/2515245917745629.

[5] A. Shojaie, E.B. Fox, *Granger Causality: A Review and Recent Advances*, **Annual Review of Statistics and Its Application**, **9**(1), 2022, pp. 289–319, DOI: 10.1146/annurev-statistics-040120-010930.

[6] A. Shojaie, E.B. Fox, *Annual Review of Statistics and Its Application Granger Causality: A Review and Recent Advances*, **Annual Review of Statistics and Its Application, 9**(1), 2022, pp. 289-319, DOI: 10.1146/annurev-statistics-040120.

[7] P. Schober, L.A. Schwarte, *Correlation coefficients: Appropriate use and interpretation*, **Anesth Analg**, **126**(5), 2019, pp. 1763–1768, DOI: 10.1213/ANE.0000000000002864.

[8] Z. Rezaee, S. Aliabadi, A. Dorestani, N.J. Rezaee, *Application of time series models in business research: Correlation, association, causation*, **Sustainability**, **12**(12), 2020, DOI: 10.3390/SU12124833.

[9] P. Damos, *Using multivariate cross correlations, Granger causality and graphical models to quantify spatiotemporal synchronization and causality between pest populations*, **BMC Ecology**, **16**, 2019, 33, DOI: 10.1186/s12898-016-0087-7.

[10] M. Khudari, N. Salwati, B. Othman, *The Impact Financial Development on Carbon Emissions: A Systematic Review Of The Literature*, **Journal of Research Administration**, **6**(1), 2024, pp. 1516-1545.

[11] X. Li, X. Liu, H. Li, R. Liu, Z. Zhang, Q. Lyu, L. Wen, *Research on Sinter Quality Prediction System Based on Granger Causality Analysis and Stacking Integration Algorithm,* **Metals**, **13**(2), 2023, Article Number 419, DOI: 10.3390/met13020419.
[12] S. Munir, M. Mayfield, D. Coca, L.S. Mihaylova, O. Osammor, *Analysis of air pollution in urban areas with airviro dispersion model-A case study in the city of Sheffield, United Kingdom*, **Atmosphere**, **11**(3), 2020, Article Number 285, DOI: 10.3390/atmos11030285.
[13] F.M. Saimi, F.M. Hamzah, M.E. Toriman, O. Jaafar, H. Tajudin, *Trend and linearity analysis of meteorological parameters in peninsular Malaysia,* **Sustainability**, **12**(22), 2020, Article Number 9533, DOI: 10.3390/su12229533.
[14] N.D.A. Halim, M. Latif, A. Mohamed, K. Maulud, S. Idrus, A. Azhari, M. Othman, N. Sofwan, *Spatial assessment of land use impact on air quality in mega urban regions, Malaysia,* ***Sustainable Cities and Society,*** **63**, 2020, Article Number 102436, DOI: 10.1016/j.scs.2020.102436.
[15] X. Zhang, L. Han, H. Wei, X. Tan, W. Zhou, W. Li, Y. Qian, *Linking urbanization and air quality together: A review and a perspective on the future sustainable urban development*, **Journal of Cleaner Production, 346**, 2022, Article Number 130988, DOI: 10.1016/j.jclepro.2022.130988.
[16] X. Song, A. Taamouti, *Measuring Granger Causality in Quantiles*, **Journal of Business & Economic Statistics**, **39**(4), 2021, DOI: 10.1080/07350015.2020.1739531.
[17] D.I. Stern, **From Correlation to Granger Causality,** 2021. Available: http://ssrn.com/abstract=1959624.
[18] W. Alahamade, I. Lake, C.E. Reeves, B. De La Iglesia, *A multi-variate time series clustering approach based on intermediate fusion: A case study in air pollution data imputation,* **Neurocomputing**, **490**, 2022, pp. 229–245, DOI: 10.1016/j.neucom.2021.09.079.
[19] M. Du, W. Liu, Y. Hao, *Spatial correlation of air pollution and its causes in northeast china,* **International Journal of Environmental Research and Public Health, 18**(20), 2021, DOI: 10.3390/ijerph182010619.
[20] H. Liu, W. Cui, M. Zhang, *Exploring the causal relationship between urbanization and air pollution: Evidence from China*, ***Sustainable Cities and Society***, **80**, 2022, Article Number 103783, DOI: 10.1016/j.scs.2022.103783.
[21] M.S. Callén, I. Martínez, G. Grasa, J.M. López, R. Murillo, *Principal component analysis and partial least square regression models to understand sorption-enhanced biomass gasification,* **Biomass Convers Biorefin**, **14**(2), 2024, pp. 2091–2111, DOI: 10.1007/s13399-022-02496-z.
[22] S. Pfenninger, I. Staffell, *Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data*, **Energy**, **114**, 2021, pp. 1251–1265, DOI: 10.1016/j.energy.2016.08.060.
[23] I. Ghosh, R.K. Jana, M.K. Sanyal, *Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms*, **Applied Soft Computing**, **82**, 2019, Article Number 105553, DOI: 10.1016/j.asoc.2019.105553.
[24] Z. Zhang, T. Xue, X. Jin, *Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities*, **Science of The Total Environment**, **741**, 2020, Article Number 140244, DOI: 10.1016/j.scitotenv.2020.140244.
[25] M. Sarfraz, M. Mohsin, S. Naseem, *A blessing in disguise: new insights on the effect of COVID-19 on the carbon emission, climate change, and sustainable environment*, **Environmental Science and Pollution Research: ESPR**, **29**(20), 2022, pp. 29651–29662, DOI: 10.1007/s11356-021-17507-1.
[26] A. Bekkar, B. Hssina, S. Douzi, K. Douzi, *Air-pollution prediction in smart city, deep learning approach*, **Journal of Big Data, 8**(1), 2021, DOI: 10.1186/s40537-021-00548-1.

[27] S. Shith, N.A. Ramli, A. Umirah, M. Nadzir, N.R. Awang, M.R. Ismail, *Drivers of nocturnal interactions between ground-level ozone and nitrogen dioxide*, **Global Nest Journal**, **25**(6), 2023, pp 149-155, DOI: 10.30955/gnj.004843.

[28] C.T. Sousa, N. Leme, M. Martins, F. Silva, T. Penha, N. Rodrigues, E. Silva, J. Hoelzemann, *Ozone trends on equatorial and tropical regions of South America using Dobson spectrophotometer, TOMS and OMI satellites instruments*, **Journal of Atmospheric and Solar-Terrestrial Physics**, **203**, 2020, Article Number 105272, DOI: 10.1016/j.jastp.2020.105272.

[29] J. Sentian, F. Herman, C.Y. Yih, J.C.H. Wui, *Long-term air pollution trend analysis in Malaysia*, **International Journal of Environmental Impacts**, **2**(4), 2019, pp. 309–324, DOI: 10.2495/EI-V2-N4-309-324.

[30] N. Ramli, M. Rubini, N.M. Noor, *Relationships between Air Pollutants and Meteorological Factors during Southwest and Northeast Monsoon at Urban Areas in Peninsular Malaysia*, **IOP Conference Series: Earth and Environmental Science**, **1303**(1), 2024, Article Number 012041, DOI: 10.1088/1755-1315/1303/1/012041.

[31] R. Cellmer, M. Bełej, A. Cichulska, *Identification of Cause-and-Effect Relationships in the Real Estate Market Using the VAR Model and the Granger Test*, **Real Estate Management and Valuation**, **27**(4), 2019, pp. 85–95, DOI: 10.2478/remav-2019-0038.

Received: August 21, 2024
Accepted: February 17, 2025