FACULTY OF CIVIL ENGINEERING & TECHNOLOGY FKTA
Universiti Malaysia Perlis (UniMAP)

FKTA Postgraduate Colloquium 2025

UNIVERSITI MALAYSIA PERLIS UniMAP

Evaluating Artificial Neural Networks with Granger Causality for PM10 Prediction in Seberang Perai, Penang

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

KEYWORDS

PM10
Air Pollution ANN
Granger Causality Forecasting
Urban Area

ABSTRACT

Air pollution, characterized by elevated concentrations of Particulate Matter (PM10), is a major environmental and health concern in urban- industrial regions of Malaysia. This study specifically investigates the predictive performance and stability of an Artificial Neural Network (ANN) model for PM10 concentrations in Seberang Perai, Penang, when its input features are selected using the Granger Causality variable selection method. Granger Causality focuses on temporal precedence by assessing whether past values of one variable improve the prediction of another. Model performance was assessed using the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Normalised Absolute Error (NAE), and the coefficient of determination (R2). Results indicated that the ANN model with Granger Causality input selection yielded poor and unstable performance (RMSE = 8.897, NAE = 0.962, R2 = 0.226). Critically, the MAE produced an inconsistent negative value (MAE = -1.9628), suggesting instability in the ANN predictions under this input selection method. These findings demonstrate that, for this specific urban-industrial context, Granger Causality is not a suitable variable selection technique for leveraging the non-linear predictive capabilities of an ANN for PM10 forecasting.

1. INTRODUCTION

Malaysia's rapid urban and industrial development has intensified air quality challenges, particularly concerning particulate matter with an aerodynamic diameter of less than 10 µm (PM10) [1]. The PM10 has been linked to respiratory illnesses, cardiovascular diseases, and broader environmental degradation, making its accurate forecasting essential for safeguarding public health, guiding policy-making, and supporting effective urban planning. Understanding the relationship between meteorological parameters and pollutant concentrations is a key step toward reliable prediction. The statistical tool Granger causality is often used in this context, as it focuses on temporal precedence, assessing whether past values of one variable can improve the prediction of another's future values [2]. However, Granger causality is limited by its linearity assumption, which can overlook the non-linear dynamics often found in atmospheric systems. Artificial Neural Networks (ANN), conversely, can model non-linear and high- dimensional interactions, potentially capturing associations invisible to purely linear techniques [3]. This study's objective is to specifically evaluate the predictive performance and stability of an ANN model for PM10 concentrations in Seberang Perai, Penang, when its input features are selected exclusively using the Granger Causality variable selection technique. Seberang Perai, a mixed urban- industrial area in the northern region, provides a meaningful basis for analysis due to its unique environmental and emission profiles [4].

2. METHODOLOGY

2.1 Site Description

The study focused on the air quality monitoring station in Seberang Perai, Penang. Seberang Perai is a mixed urban-industrial area. The station details are summarised in Table 1:

Table 1 Monitoring station's location

Station ID	CA07P
State	Seberang Perai, Pulang Pinang
Location	Sek. Keb. Cenderawasih, Taman Inderawasih, Perai
Coordinate	N05° 23.470', E100° 23.213'

*Corresponding author: noorulamilin@studentmail.unimap.edu.my

2.2 Modelling Approach: Artificial Neural Networks (ANN)

The predictive model employed was the Artificial Neural Network (ANN). ANN is a non-linear machine learning technique inspired by biological neural networks, consisting of interconnected layers of processing nodes capable of learning complex patterns [3]. The model was chosen to exploit its ability to capture intricate, non-linear interactions between pollutants and meteorological parameters influencing PM10 formation and dispersion [5], [6]. Specifically, the ANN's flexibility allows it to handle high-dimensional datasets and capture subtle dependencies that linear models typically miss.

2.3 Variable Selection Technique: Granger Causality

The input features for the ANN model were chosen exclusively using the Granger Causality analysis. This method was applied to capture predictors with significant temporal precedence influencing PM10 concentrations [2]. Unlike correlation, Granger causality examines whether past values of one variable can improve the forecast of another, which is particularly valuable for time-series data to reveal lagged relationships between atmospheric variables and PM10 levels [7]. By using this method, the study aims to assess how incorporating the temporal dimension, despite the linearity assumption of the test, impacts the performance of the non-linear ANN model [8], [9].

2.4 Performance Metrics

The reliability and accuracy of the ANN model were quantified using four standard statistical performance indicators: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Normalised Absolute Error (NAE), and the Coefficient of Determination (R^2) [10]. RMSE measures the square root of the average squared differences, heavily penalizing large deviations, which is important for capturing extreme pollution events [11], [12]. MAE calculates the average of the absolute differences, providing a straightforward interpretation of the average prediction error. NAE, obtained by normalizing the MAE by the mean of the observed values, provides a dimensionless indicator for cross- comparison [13]. Finally, R^2 measures the proportion of variance in observed PM10 concentrations that the model's predictions can explain, suggesting the overall fit of the model. These four metrics provided a robust evaluation framework.

3. RESULTS AND DISCUSSION

3.1 Predictive Performance of ANN with Granger Causality

The predictive performance of the Artificial Neural Network (ANN) model, specifically when using the Granger Causality variable selection method for PM10 concentrations in Seberang Perai, Penang, is summarised in Table 2:

Table 2 PM10 concentrations in Seberang Perai, Penang

RMSE	MAE	NAE	R^2
8.897	1.9628	0.962	0.226

The ANN model, when input variables were selected via Granger Causality, yielded the poorest and most unstable performance in Seberang Perai. Specifically, the model achieved a high RMSE of 8.897 and a minimal explanatory power with $R^2 = 0.226$. Critically, the model produced inconsistent outcomes, including an unphysical negative Mean Absolute Error (MAE) of -1.9628, which suggests significant instability in the predictions under this input configuration. This result clearly indicates that Granger causality is not a suitable variable selection method for the ANN model in this specific urban-industrial context as shown in Figure 1.

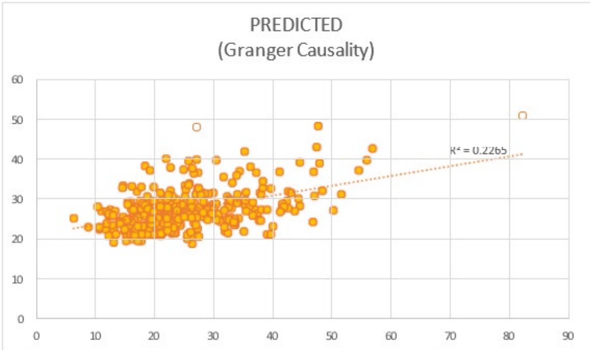

Figure 1. Scatter plot with regression lines showing the PM_{10} concentrations in Seberang Perai, Penang, using Artificial Neural Network (ANN) models.

3.2 Discussion of Model Instability and Performance

The results clearly demonstrate that the ANN model with the Granger Causality input selection yielded poor and unstable performance in predicting PM10 concentrations in Seberang Perai [3], [8]. The model recorded an of, and a very low coefficient of determination (R^2) of 0.226, indicating that it explained only a minimal fraction of the total PM10 variance. The large Normalised Absolute Error (NAE) of 0.962 further confirms the low predictive accuracy of this configuration. Most critically, the Mean Absolute Error (MAE) produced an inconsistent negative value (MAE = -1.9628). As MAE is derived from the average of absolute differences, it must mathematically be non-negative. This inconsistent result strongly suggests significant instability and unreliability in the ANN predictions when the input features are determined exclusively by Granger causality analysis. This weak performance suggests that the temporal relationships captured by Granger causality alone cannot sufficiently explain PM10 variability in Seberang Perai, despite the region's complex coastal and urban factors [1], [14]. The overall failure of this configuration implies that the non-linear dynamics of PM10 concentrations are not adequately isolated by a purely temporal and linear-assumed variable selection method [2], [15].

4. CONCLUSION

This focused evaluation demonstrated that the Artificial Neural Network (ANN) model, when combined with Granger Causality for variable selection, is unsuitable for PM10 prediction in Seberang Perai, Penang [9], [16]. The model exhibited poor predictive capability (RMSE = 8.897, R^2 = 0.226) and, most critically, produced an inconsistent negative Mean Absolute Error (MAE = -1.9628), highlighting significant model instability under this specific input configuration. The findings imply that for this urban-industrial context, the non-linear capabilities of ANN are not effectively leveraged by a purely temporal, linear-assumed variable selection method like Granger Causality [17]. The pollutant–meteorology relationships are likely governed by more simultaneous and consistent sources, which favour alternative feature selection methods [18]. Future research should explore non-temporal variable selection methods or expand datasets with additional nonlinear predictors to test whether ANN can achieve stronger performance.

ACKNOWLEDGEMENT

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

REFERENCE

[1] L. Liang, Z. Wang, and J. Li, "The effect of urbanization on environmental pollution in rapidly developing urban agglomerations," *J Clean Prod*, vol. 237, Nov. 2019, doi: 10.1016/j.jclepro.2019.117649.

[2] A. Shojaie and E. B. Fox, "Granger Causality: A Review and Recent Advances," *Annu Rev Stat Appl*, vol. 9, no. 1, pp. 289–319, Mar. 2022, doi: 10.1146/annurev-statistics-040120- 010930.

[3] M. Madhiarasan and M. Louzazni, "Analysis of Artificial Neural Network: Architecture, Types, and Forecasting Applications," *Journal of Electrical and Computer Engineering*, vol. 2022, 2022, doi: 10.1155/2022/5416722.

[4] S. Nazurah Mukhtar, A. Farhana Jamaludin, M. Hafiz, B. A. Razak, W. Arminda, and A. F. Saleh, "INVESTIGATING FACTORS INFLUENCING RESIDENTIAL LOCATION CHOICE USING PLS- SEM ANALYSIS: A CASE STUDY IN SEBERANG PERAI, PENANG, MALAYSIA," 2025.

[5] I. A. Pérez, M. Á. García, M. L. Sánchez, N. Pardo, and B. Fernández- Duque, "Key points in air pollution meteorology," Nov. 02, 2020, *MDPI AG*. doi: 10.3390/ijerph17228349.

[6] I. Kayes, S. A. Shahriar, K. Hasan, M. Akhter, M. M. Kabir, and M. A. Salam, "The relationships between meteorological parameters and air pollutants in an urban environment," *Global Journal of Environmental Science and Management*, vol. 5, no. 3, pp. 265–278, 2019, doi: 10.22034/gjesm.2019.03.01.D. Yang, H. Chen, Y. Song, and Z. Gong, "Granger Causality for Multivariate Time Series Classification," in *Proceedings - 2017 IEEE International Conference on Big Knowledge, ICBK 2017*, Institute of Electrical and Electronics Engineers Inc., Aug. 2021, pp. 103–110. doi: 10.1109/ICBK.2017.36.

[7] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference," *J Neurosci Methods*, vol. 223, pp. 50–68, Feb. 2021, doi: 10.1016/j.jneumeth.2013.10.018.

[8] D. I. Stern, "From Correlation to Granger Causality," 2021. [Online]. Available: http://ssrn.com/abstract=1959624http://ssrn.com/abstract=19596241.

[9] C. Audet, J. Bigeon, D. Cartier, S. Le Digabel, and L. Salomon, "Performance indicators in multiobjective optimization," *Eur J Oper Res*, vol. 292, no. 2, pp. 397–422, 2021, doi: https://doi.org/10.1016/j.ejor.2020.11. 016.

[10] S. Chen and Y. Ding, "Tackling Heavy Metal Pollution: Evaluating Governance Models and Frameworks," *Sustainability 2023, Vol. 15, Page 15863*, vol. 15, no. 22, p. 15863, Nov. 2023, doi: 10.3390/SU152215863.

[11] R. Munsif, M. Zubair, A. Aziz, and M. Nadeem Zafar, "Industrial Air Emission Pollution: Potential Sources and Sustainable Mitigation," in *Environmental Emissions*, IntechOpen, 2021. doi: 10.5772/intechopen.93104.

[12] T. Szandała, "Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks," in *Bio-inspired Neurocomputing*, A. K. Bhoi, P. K. Mallick, C.-M. Liu, and V. E. Balas, Eds., Singapore: Springer Singapore, 2021, pp. 203–224. doi: 10.1007/978-981-15-5495-7_11.

[13] I. A. M. Jafri, N. M. Noor, N. A. A. A. Rahim, A. Z. Ul-Saufie, Z. Hassan, and G. Deak, "Spatial and Temporal Analysis of Particulate Matter (PM10) in Urban-Industrial Environment during Episodic Haze Events in Malaysia," *EnvironmentAsia*, vol. 16, no. 1, pp. 111–125, Jan. 2023, doi: 10.14456/ea.2023.10.
[14] Z. Rezaee, S. Aliabadi, A. Dorestani, and N. J. Rezaee, "Application of time series models in business research: Correlation, association, causation," *Sustainability (Switzerland)*, vol. 12, no. 12, Jun. 2020, doi: 10.3390/SU12124833.
[15] P. Schober and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesth Analg*, vol. 126, no. 5, pp. 1763–1768, May 2019, doi: 10.1213/ANE.0000000000002864.
[16] J. Youn, H. Kim, and J. Lee, "Relationships between Thermal Environment and Air Pollution of Seoul's 25 Districts Using Vector Autoregressive Granger Causality," *Sustainability*, vol. 15, no. 23, p. 16140, Nov. 2023, doi: 10.3390/su152316140.
[17] A. Shojaie and E. B. Fox, "Annual Review of Statistics and Its Application Granger Causality: A Review and Recent Advances," 2021, doi: 10.1146/annurev-statistics-040120.