Mathematical Sciences and Informatics Journal
Vol. 5, No. 1, May 2024, pp. 10-20
http://www.mijuitm.com.my
DOI : 10.24191/mij.v5i1.890

Forecasting CO_2 Emissions in Malaysia using Group Method of Data Handling

Basri Badyalina
Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, Malaysia
basribdy@uitm.edu.my

Fatin Farazh Ya'acob
Fakulti Pengurusan dan Perniagaan, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, Malaysia
fatinfarazh@uitm.edu.my

Rabiatul Munirah Alpandi
Fakulti Pengurusan dan Perniagaan, Universiti Teknologi MARA Cawangan Selangor Kampus Puncak Alam, Malaysia
rabiatulmunirah@uitm.edu.my

Nur Diana Zamani
Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, Malaysia
nurdi958@uitm.edu.my

Muhammad Zulqarnain Hakim Abd Jalal
Kolej Pengajian Perkomputeran, Informatik dan Matematik, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, Malaysia
zulqarnainhakim @uitm.edu.my

Amir Imran Zainoddin
Fakulti Pengurusan dan Perniagaan, Universiti Teknologi MARA Cawangan Johor Kampus Segamat, Malaysia
amirimran @uitm.edu.my

Article Info

Article history:

Received Jan 08, 2024
Revised Feb 12, 2024
Accepted Apr 05, 2024

Keywords:

Forecasting
Group Method of Data Handling
Accuracy
CO_2 Emission
Malaysia

ABSTRACT

Critical environmental issues, such as climate change, pollution, and resource depletion, urgently require data-driven decision-making and accurate forecasting to guide sustainable policies and interventions. However, forecasting is a complex task that necessitates rigorous research to ensure precise predictions essential for addressing these environmental challenges effectively. To meet these forecasting challenges, this study utilized Group Method of Data Handling (GMDH) method, focusing on CO_2 emission in Malaysia. The analysis of the GMDH forecasting model for CO_2 emission provides distinct patterns in the behaviors of three input variables X1, defined as $(y_{t-2}, y_{t-3}, y_{t-5})$, X2 defined as $(y_{t-1}, y_{t-5}, y_{t-6}, y_{t-7})$ and X3 defined as $(y_{t-2}, y_{t-5}, y_{t-6}, y_{t-8}, y_{t-10})$. Notably, X2 consistently exhibits strong performance, whereas X1 and X3 face difficulties, particularly in the forecasting phase. The GMDH model demonstrates proficiency in adaptive self-organization, automatic feature extraction, managing non-linear relationships, and interpretability, enhancing its effectiveness in capturing complex patterns in CO_2 emission data. The observed decrease in performance for specific inputs during the forecasting process highlights the necessity for improving and adjusting the model and developing a detailed grasp of the dynamics of the variables involved.

Corresponding Author:
Fatin Farazh Ya'acob
Universiti Teknologi MARA Cawangan Johor Kampus Segamat, Malaysia
email: fatinfarazh@uitm.edu.my

PENERBIT PRESS
UNIVERSITI TEKNOLOGI MARA

1. Introduction

In recent years, global concerns about climate change and its associated environmental impacts have reached unprecedented levels, prompting nations worldwide to reevaluate their carbon footprint and seek sustainable solutions [1]. At the forefront of this challenge is Malaysia, a nation navigating the delicate balance between economic development and environmental preservation [2], [3]. As Malaysia's economy experiences rapid growth and its industrial landscape continues to expand, there is an urgent imperative to confront the escalating levels of carbon dioxide (CO_2) emissions, which stem from various sectors such as manufacturing, transportation, and energy production. This pressing issue not only poses significant environmental challenges, including climate change and air pollution, but also underscores the need for comprehensive strategies and policies aimed at mitigating CO_2 emissions while ensuring sustainable economic development.

Forecasting CO_2 emissions is important for several reasons, and it plays a crucial role in addressing environmental, economic, and policy challenges. Policymakers rely on accurate forecasts to make informed decisions about environmental policies and regulations [4], [5]. Forecasting provides the necessary data to design effective measures for emissions reduction and sustainable practices [6]. Many nations aim to achieve sustainable development, which involves economic progress without compromising environmental health [7].

Forecasting CO_2 emissions helps identify sustainable pathways for development. Forecasting helps allocate resources efficiently by identifying sectors or industries with the highest carbon footprint. This information aids in prioritizing efforts to reduce emissions in areas that contribute significantly to overall carbon output. Forecasting can drive research and development in clean and renewable technologies by providing insights into future energy demands and emission scenarios [8]. This, in turn, promotes innovation and the adoption of cleaner energy sources. Forecasting results can be used to raise public awareness about the environmental impact of various activities. This knowledge empowers individuals and businesses to make environmentally conscious choices. Understanding future CO_2 emission trends helps in assessing climate-related risks. This information is valuable for developing strategies to adapt to potential changes in weather patterns, sea levels, and other climate-related impacts.

In the dynamic landscape of data-driven decision-making, forecasting holds a pivotal role in anticipating future trends and making informed choices across various domains. The application of data driven model has been implemented in may research areas [8-13]. Among the array of predictive modelling techniques, the Group Method of Data Handling (GMDH) model stands out as a powerful tool for unravelling complex relationships within datasets [14-17]. While the GMDH algorithm has been recognized for its potential in forecasting across various domains, its application within the environmental science areas, particularly for forecasting CO2 emissions, has not been thoroughly investigated. This study addresses the research gap by evaluating the potential of the GMDH model in forecasting CO_2 emission from fossil fuels and industries for Malaysia.

This study provides two key contributions. Firstly, it introduces a GMDH model that incorporates three specific groups of input variables related to CO2 emissions. Secondly, through empirical analysis, the findings show that the forecasting accuracy of this model for CO2 emissions is significantly influenced by the categorization of input variables.

2. Literature Review

GMDH, rooted in the principles of self-organization and adaptive learning, offers a versatile approach to forecasting by automatically selecting and optimizing models through a process of continuous improvement. The Group GMDH model is effective in automatic variable selection through a iterative procedure that assesses the significance of input variables from a vast array of potential variables. This process minimizes dimensionality by emphasizing the most informative predictors, hence enhancing the efficiency and interpretability of the model [18], [19]. GMDH is highly efficient in identifying nonlinear relationships in data, which is essential for modeling intricate systems and phenomena found in various real-world applications like economics, engineering, and environmental sciences [20-22]. One of the notable strengths of GMDH is its robustness to noisy data. GMDH can effectively enhance model performance in difficult circumstances by iteratively adjusting to filter out noise and manage uncertainties or mistakes in the dataset [20], [23]. GMDH shows adaptation to diverse data kinds and architectures, making it adaptable and suitable for varied domains and datasets. GMDH may adapt its model structure to match the unique attributes of the dataset, improving its flexibility and usefulness across continuous, discrete, or categorical data types [24-26]. GMDH models demonstrate exceptional prediction accuracy and generalization performance, particularly when dealing with datasets that contain nonlinear interactions and intricate patterns. GMDH's repeated refinement process enables the model to enhance its predictive accuracy and adjust to the data distribution, resulting in dependable predictions and strong generalization to new data [18], [27], [28]. GMDH's adaptability makes it a helpful tool for various applications in different fields, offering customized insights and forecasts to meet specific analytical requirements. This inherent flexibility leads to model architectures that are efficient and effective in precisely capturing the intrinsic relationships within the data without requiring manual intervention [29]. GMDH allows incremental learning, enabling the model to continually develop and enhance its performance with fresh data over time. This feature guarantees that the model stays current and adaptable to shifting data trends, making it ideal for dynamic settings and changing datasets. This manuscript delves into the integration of GMDH within the context of forecasting, exploring its applicability and effectiveness in predicting outcomes, particularly in the domain of environmental science. As industries and nations grapple with the challenges posed by climate change and environmental sustainability, the need for accurate and reliable forecasting methods becomes increasingly apparent. The GMDH algorithm, with its ability to handle non-linear relationships and adapt to evolving data patterns, emerges as a promising candidate for forecasting endeavours. This introduction sets the stage for a comprehensive exploration of GMDH in the realm of forecasting, elucidating its underlying principles, methodological nuances, and the potential it holds for enhancing the precision and reliability of predictions. By focusing on GMDH's application in the specific context of forecasting CO_2 emissions in Malaysia, this research aims to contribute valuable insights to the broader discourse on predictive modelling, offering a pathway for robust and adaptive forecasting methodologies in the face of complex, real-world challenges.

3. Methodology

3.1 Data

In the contemporary discourse of environmental sustainability, understanding and managing carbon dioxide (CO_2) emissions stand as imperative components in the pursuit of a greener and more resilient future. As a nation at the crossroads of rapid industrialization and ecological conservation, Malaysia's CO_2 emission data serves as a critical lens through which to analyze the intricate interplay between economic development and environmental impact. The CO_2 data used in this study is ranging from 1890 to 2022. The data is obtained from Penn World Table v10.0 via Our World in Data . The data shows the CO_2 emission from fossil fuels and industries for Malaysia [30].

3.2 Group Method of Data Handling

GMDH model is based on the principle of heuristic of self-organization to identify a mathematical model between input and output data. GMDH can solve also the modeling problem that has data. The flow chart that describes the GMDH model is shown in Figure 1. The GMDH is an advanced and adaptable modelling approach that combines statistical analysis and machine learning. GMDH stands out from standard modelling techniques due to its capacity to adapt and enhance its structure in response to the intrinsic patterns seen in the data. GMDH is especially suitable for scenarios characterized by large, complex datasets that may encompass a wide range of factors. The process functions by sequentially choosing and merging variables, creating clusters, and improving the model's structure to boost its ability to make accurate predictions. The iterative method enables GMDH models to adjust to the inherent complexity of the data, automatically recognizing important features and relationships. An important characteristic of GMDH is its ability to detect non-linear relationships and capture complex patterns within the information. GMDH's versatility renders it important in diverse fields such as banking, engineering, and scientific research, where the capacity to comprehend complex relationships is vital for effective decision-making. Essentially, GMDH signifies a fundamental change in modelling, where the structure of the model is not planned but instead arises from the data itself. The data-driven nature of GMDH models enables them to effectively handle various datasets, making them a versatile and potent tool for predictive modelling and analysis in scenarios where conventional methods may be inadequate.

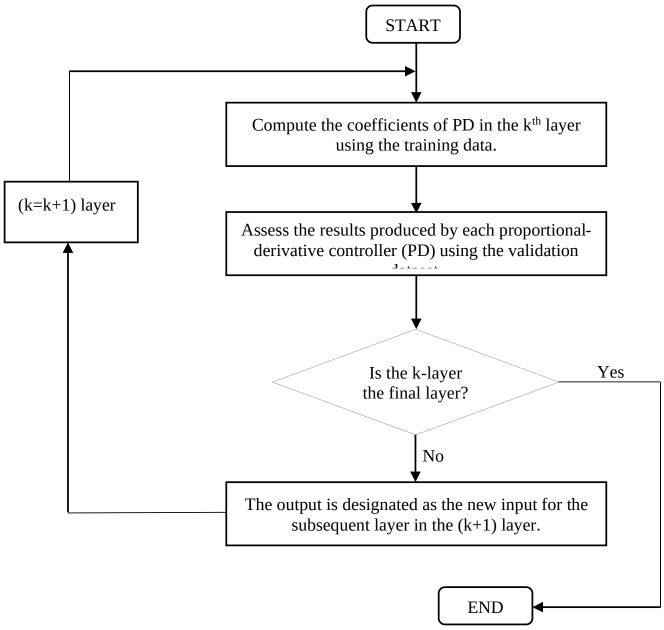

Figure 1. Group Method of Data Handling model framework

Step 1: In the modeling process, the input variables are denoted as $X = \{x_1, x_2, ..., x_p\}$, and the output variable is represented by y, with p being the total number of inputs. If deemed necessary, normalization of input data is carried out as part of the preprocessing steps.

Step 2: The available data is partitioned into training and forecasting datasets. The training dataset comprises 80% of the total dataset, while the forecasting dataset makes up the remaining 20% [31], [32]. The training dataset is employed to construct the partial description (PD), manifested in the form of a quadratic regression polynomial. The PD is shown in Equation 1.

$$\hat{w}_k = v_0 + v_1 x_i + v_2 x_j + v_3 x_i x_j + v_4 x_i^2 + v_5 x_j^2 \tag{1}$$

The coefficients of the PD $(v_0, v_1, v_2, v_3, v_4, v_5)$ are estimated using the least square method. The training dataset is then employed to evaluate the estimated PD. In this step, the total number of regression polynomials is determined.

$$U = p(p-1)/2 \tag{2}$$

Therefore, at current layer GMDH model contains U estimates of y_k where $\hat{y}_k = \hat{w}_k$.

Step 3: The identification of the optimal new variable for the next layer involves the elimination of the weakest variable based on specific selection criteria. These criteria are rooted in performance indices, such as the mean square error (MSE), which quantifies how well the values $\hat{y}_k$ align with the experimental output y. Among several variables, the single best variable, determined by the smallest mean square error (MSE) on the training dataset, is chosen as the new input variable. This selected variable is then combined with the existing input variables $\{x_1, x_2, ..., x_p, \hat{y}_k\}$ with $p = p+1$ The MSE is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{i,k})^2 \tag{3}$$

Step 4: Validate the termination criterion by evaluating the optimal outcome of the present layer with the previous layer. This entails evaluating if the set of equations in the model may be further improved. The iterative computation terminates if the mean square error (MSE) on the current layer for the validation dataset increases or stops decreasing compared to the previous layer. If the mean squared error (MSE) on the current layer lowers compared to the prior layer, the process is repeated by going back to stages 2 and 3. The iteration concludes when there is no further improvement, completing a realization of the network.

3.3 Model Performance

Climate change mitigation efforts heavily rely on accurate forecasting of CO_2 emissions. Assessing the performance of CO_2 emission forecasting models is crucial for making informed decisions and implementing targeted interventions. Accurate assessment of forecasting models is essential to ensure the reliability of predictions. One key aspect of this assessment is the measurement of errors, which quantifies the disparity between predicted values and actual observations. Root Mean Squared Error (RMSE) a widely-used metric that provides a comprehensive view of the prediction accuracy. It calculates the square root of the average of the squared differences between predicted and observed values (refer Equation 4). RMSE penalizes larger errors more significantly, making it sensitive to outliers. Mean Absolute Error (MAE) is another commonly used metric for error measurement in time series analysis. Unlike RMSE, MAE computes the average absolute differences between predicted and observed values. Coefficient of determination evaluates the proportion of the variance in the dependent variable (actual values) that is predictable from the independent variable (predicted values). It is a measure of how well the model explains the variability in the data. values range from 0 to 1, where 1 indicates a perfect fit. Mathematically, the RMSE, MAE and coefficient of determination can be expressed by Equation 4, Equation 5 and Equation 6 respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)}{n}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{n} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \tag{6}$$

where y_i is the observed data, $\hat{y}_i$ is the predicted data, $\overline{y}$ is the mean of the observed data and n the number of data.

4. Results and Discussion

Amidst the growing worldwide conversation on climate change, countries such as Malaysia are becoming more aware of the crucial need of precise CO_2 emission forecasting in shaping sustainable development trajectories. In Malaysia, the unique interplay of industrial growth, urbanization, and environmental policies accentuates the importance of meticulous result analysis in CO_2 emission forecasting. This research serves the purpose of providing information for national plans and is in line with Malaysia's commitment to international climate initiatives. Table 1 shows the descriptive statistics of the data. The CO_2 emission data is measure in megatons (Mt).

Table 1. Descriptive statistics for CO_2 emission

Mean	Standard Error	Variance	Kurtosis	Skewness	Minimum	Maximum
48.793 Mt	6.952 Mt	6428.832 Mt	1.542 Mt	1.689 Mt	0.007 Mt	291.071 Mt

The mean CO_2 emission in the dataset is approximately 48.793 Mt, with a standard error of about 6.952 Mt, indicating some variability in the sample means. The standard error of about 6.952 Mt provides information about the precision of the mean estimate. A higher standard error indicates that there is more variability or uncertainty in the average value. The data has a sample variance of approximately 6428.832 Mt, suggesting a spread in CO_2 emission values. The positive kurtosis (1.542 Mt) indicates relatively heavy tails in the distribution, while the positive skewness (1.689 Mt) suggests a right-skewed distribution. The range of CO_2 emissions in the dataset varies from a minimum of 0.007 Mt to a maximum of 291.071 Mt, reflecting a diverse range of emissions. The plot of the data is shown in Figure 2.

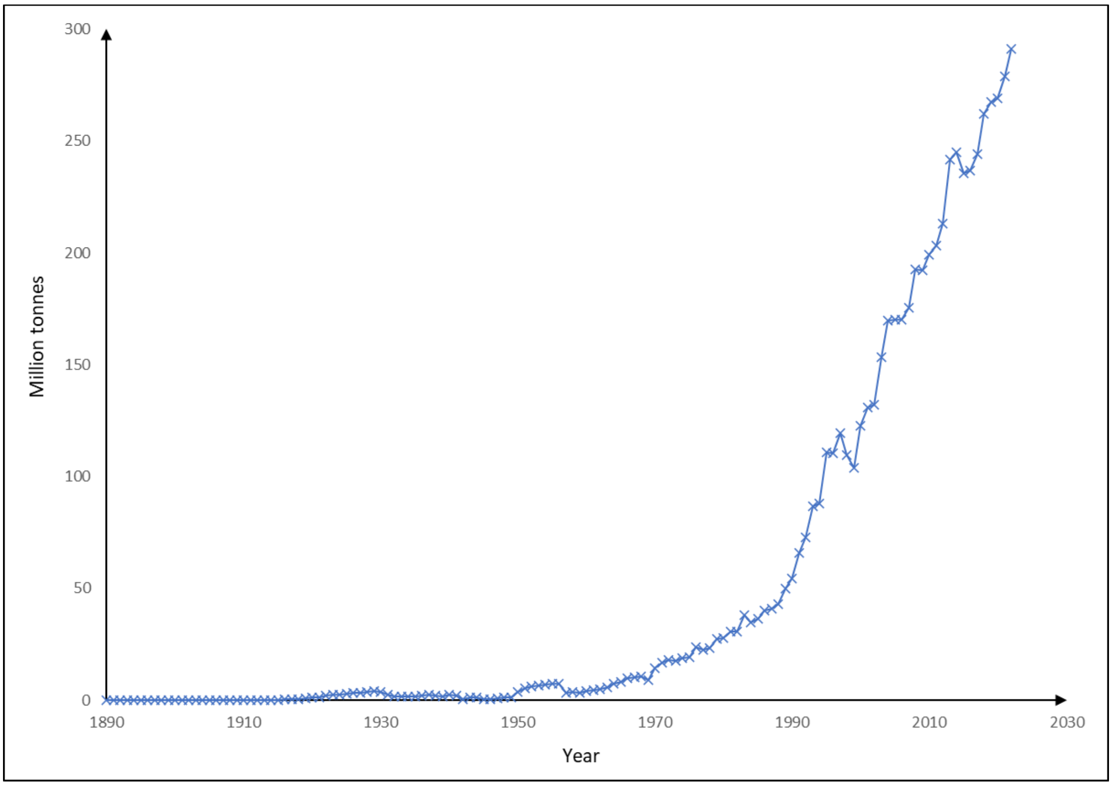

Figure 2. CO_2 emission from fossil fuel and industries in Malaysia

In the domain of forecasting, the selection of input variables significantly influences the model's effectiveness. In this study, the determination of input variables for the GMDH model involved a meticulous process combining a stepwise approach and trial-and-error methodology. The outcomes of this investigation, including the identified input variables, are presented in Table 2.

Table 2. Input variables for GMDH model

Input	Type of input
$X1$	$y_{t-2}, y_{t-3}, y_{t-5}$
$X2$	$y_{t-1}, y_{t-5}, y_{t-6}, y_{t-7}$
$X3$	$y_{t-2}, y_{t-5}, y_{t-6}, y_{t-8}, y_{t-10}$

There are three types of inputs that are X1, X2 and X3. In addition, in order to assure the strength and applicability of the GMDH model, a meticulous approach was used to divide the dataset into two separate sets: an 80% training set and a 20% forecasting set. The training set, which accounts for 80% of the data, was used to develop the model. This allowed the GMDH algorithm to learn and adjust to the underlying patterns in the majority of the dataset. Afterwards, the forecasting set including 20% of the data, which was not previously observed, was used to assess the model's ability to forecast outcomes. This approach, which is divided into two parts, is used to confirm the effectiveness of the model in applying patterns learnt during training to new data points that were not observed before. Partitioning the datasets into training and forecasting sets, with a larger proportion allocated to training, improves the model's ability to comprehend the complexities of the data while still being able to make precise forecasts for new scenarios during forecasting. The assessment of CO_2 emission forecasting is conducted through the utilization of error measurement metrics, including RMSE, MAE, and R^2. Table 3 presents the results of RMSE, MAE, and R^2 for each input variable, segmented into training and forecasting phases.

Table 3. Result of RMSE, MAE and R^2 for training and forecasting data set

Input	**Training**			**Forecasting**		
	RMSE	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**
$X1$	2.6813	1.7589	0.9835	18.6412	12.4196	0.9470
$X2$	1.6053	1.1028	0.9941	6.2227	4.5743	0.9883
$X3$	2.6706	1.5385	0.9881	14.8207	10.7534	0.9399

Referring to Table 3, the GMDH model demonstrates exceptional performance during the training phase when considering the X1 input, as evidenced by low RMSE and MAE values, with RMSE at 2.6813 and MAE at 1.7589. However, in the subsequent forecasting phase, there is a discernible escalation in both RMSE and MAE, implying a diminished predictive accuracy for the X1 input variables on the forecast dataset. Nevertheless, the associated R2 value of 0.9470 from the GMDH model with X1 input variables indicates a reasonably robust explanatory power. Conversely, the GMDH model with X2 input variables exhibits consistently impressive results across both training and forecasting phases, characterized by low RMSE (1.6053) and MAE (1.1028) values, underscoring the model's precision in capturing patterns associated with X2. Furthermore, the GMDH model demonstrates robust performance during the training phase with X3 as an input variable, reflected in low RMSE (2.6706) and MAE (1.5385) values. However, challenges arise during the forecasting phase, indicated by an increase in errors, with RMSE at 14.8207 and MAE at 10.7534, suggesting potential difficulties in generalizing the model to new data. Notably, while the GMDH model with X2 input variables maintains consistent performance in both training and forecasting, a significant performance decline is observed for the GMDH model with X1 and X3 as input variables, particularly during forecasting. This decline may signify challenges encountered by the model in extrapolating learned patterns to the forecast dataset for X1 and X3. Despite the observed challenges with specific inputs, the GMDH model exhibits notable strengths that contribute to its efficacy in predicting CO_2 emissions. The GMDH model's unique ability to autonomously self-organize its structure during training enables it to capture complex relationships within the data. This adaptability is particularly advantageous when dealing with intricate patterns and nonlinearities in time series data, providing a flexible framework for forecasting. Figure 3 illustrates the estimation of CO2 emissions based on the training dataset, utilizing input variables X1, X2, and X3. Meanwhile, Figure 4 depicts the CO2 emission estimation derived from the forecast dataset, employing the same set of input variables X1, X2, and X3.

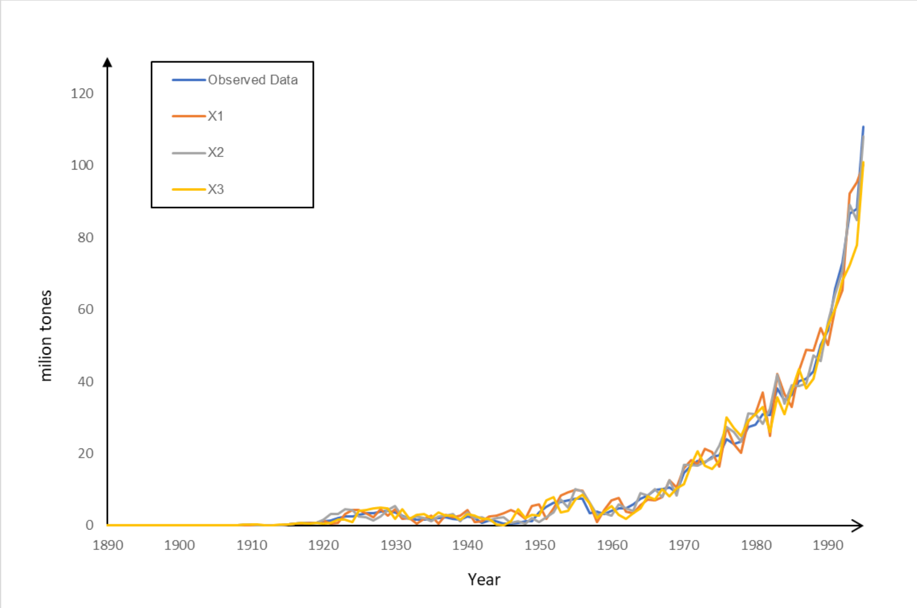

Figure 3. CO_2 emission estimation for training data set for X1, X2 and X3 input variables

Figure 4. CO_2 emission estimation for forecast data set for X1, X2 and X3 input variables

Based on the results, it is certain that the input variable X2 is the best choice for the GMDH model in predicting CO_2 emissions in Malaysia. Therefore, X2 will be utilised to forecast CO_2 emissions for both the years 2023 and 2024. The 2023 forecast utilises a single-step estimation, whereas the 2024 forecast employs a two-step estimation method utilizing the GMDH model. Table 4 displays the projected CO2 emission figures for the years 2023 and 2024.

Table 4. Input fort GMDH model

Year	Forecast CO_2 Emission
2023	301.3055 Mt
2024	317.0915 Mt

5. Conclusion

The primary objective of this research is to forecast CO_2 emissions using the Group Method of Data Handling (GMDH) model in Malaysia. A comprehensive evaluation of several input factors, combined with thorough evaluation of the training and forecasting phases, has been conducted. The analysis has provided useful insights into the model's capabilities, limitations, and its performance in various temporal settings. The unique patterns displayed by the input variables X1, X2, and X3 have been noticed in both the training and forecasting phases. Although X2 continuously shown strong performance, both X1 and X3 encountered difficulties, particularly in the domain of forecasting. The GMDH model demonstrated its capacity to adaptively self-organize, automatically extract features, handle non-linear relationships, and provide interpretability. These qualities enhance its usefulness in capturing complex patterns in CO_2 emission data. Although X2 maintained its performance, there was a noticeable decline in the prediction accuracy of X1 and X3 over the forecast period. The decrease in accuracy highlights the difficulties in using previously acquired patterns to forecast new sets of data, highlighting the significance of comprehending the distinct characteristics of each variable. Forecasting future emission trends is valuable for policymakers, industries, and environmentalists as it enables them to develop efficient plans for mitigating emissions, implement steps to reduce emissions, and promote sustainable practices. Forecasting CO_2 emissions not only supports environmental conservation but also aligns with global initiatives to combat climate change and achieve emission reduction goals. By comprehending the intricacies of CO_2 emission patterns, individuals and organizations can take proactive measures to address environmental concerns, fostering a shared dedication to a more sustainable and resilient future.

Acknowledgment

The authors would like to thank Universiti Teknologi MARA for the support given to this study.

Conflict of Interest

The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

References

[1] S. Hoffmann, W. Lasarov, H. Reimers, and M. Trabandt, "Carbon footprint tracking apps. Does feedback help reduce carbon emissions?," *Journal of Cleaner Production,* p. 139981, 2023, doi.org/10.1016/j.jclepro.2023.139981.

[2] E. Chin, R. Maharjan, and N. Emalya, "Navigating Soil Erosion Challenges in Malaysia: Insights, Prospects, and Solutions," *Civil and Sustainable Urban Engineering,* vol. 3, no. 2, pp. 138–147-138–147, 2023, doi.org/10.53623/csue.v3i2.339.

[3] M. Qamruzzaman, "Clean energy-led tourism development in Malaysia: Do environmental degradation, FDI, Education and ICT matter?," *Heliyon,* vol. 9, no. 11, 2023, doi.org/10.1016/j.heliyon.2023.e21779.

[4] D. J. Fiorino, *Making environmental policy*. Univ of California Press, 2023.

[5] N. A. Mokhtar, B. Badyalina, Y. Z. Zubairi, A. S. M. Al Mamun, F. F. Ya'acob, and N. F. Shaari, "Functional Model of Wind Direction Data in Kuching, Sarawak, Malaysia," *Applied Mathematical Sciences,* vol. 16, no. 7, pp. 349-357, 2022, doi.org/10.12988/ams.2022.916816.

[6] H. Sheng, T. Feng, and L. Liu, "The influence of digital transformation on low-carbon operations management practices and performance: does CEO ambivalence matter?," *International Journal of Production Research,* vol. 61, no. 18, pp. 6215-6229, 2023, doi.org/10.1080/00207543.2022.2088426.

[7] Z. Guang-Wen, M. Murshed, A. B. Siddik, M. S. Alam, D. Balsalobre-Lorente, and H. Mahmood, "Achieving the objectives of the 2030 sustainable development goals agenda: Causalities between economic growth, environmental sustainability, financial development, and renewable energy consumption," *Sustainable Development,* vol. 31, no. 2, pp. 680-697, 2023, doi.org/10.1002/sd.2411.

[8] B. Badyalina, N. A. Mokhtar, A. I. F. Azimi, M. Majid, M. F. Ramli, and F. F. Yaa'coob, "Data-driven Models for Wind Speed Forecasting in Malacca State," *Matematika,* pp. 125-139, 2022, doi.org/10.11113/matematika.v38.n2.1404.

[9] E. Bwambale, F. K. Abagale, and G. K. Anornu, "Data-driven model predictive control for precision irrigation management," *Smart Agricultural Technology,* vol. 3, p. 100074, 2023, doi.org/10.1016/j.atech.2022.100074.
[10] L. C. Kerk, N. A. Mokhtar, P. Shamala, and B. Badyalina, "Global Optimization Method For Minimizing Portfolio Selection Risk," *Matematika,* pp. 115-123, 2022, doi.org/10.11113/matematika.v38.n2.1403.
[11] Z. Tarmudi, M. Y. Yusri, N. S. K. Arunah, and B. Badyalina, "Portfolio selection decision using fuzzy Delphi method with confidence analysis," in *AIP Conference Proceedings*, 2023, vol. 2746, no. 1: AIP Publishing, doi.org/10.1063/5.0153018.
[12] B. Badyalina and A. Shabri, "Flood estimation at ungauged sites using group method of data handling in Peninsular Malaysia," *Jurnal Teknologi,* vol. 76, no. 1, pp. 373-380, 2015, doi.org/10.11113/jt.v76.2640.
[13] F. F. Ya'acob *et al.*, " Competitiveness analysis of edible-birdnest (EBN) Ranching in Johor Bahru and Gua Musang, Malaysia," *INSIGHT Journal,* pp. 1-16, 2023.
[14] A. Piven, D. Darmoroz, E. Skorb, and T. Orlova, "Machine learning methods for liquid crystal research: phases, textures, defects and physical properties," *Soft Matter,* vol. 20, no. 7, pp. 1380-1391, 2024, doi.org/10.1039/d3sm01634j.
[15] M. Valipour, H. Khoshkam, S. M. Bateni, and C. Jun, "Machine-learning-based short-term forecasting of daily precipitation in different climate regions across the contiguous United States," *Expert Systems with Applications,* vol. 238, p. 121907, 2024, doi.org/10.1016/j.eswa.2023.121907.
[16] B. Badyalina, N. A. Mokhtar, N. A. M. Jan, and M. FADHIL, "Hydroclimatic Data Prediction using a New Ensemble Group Method of Data Handling Coupled with Artificial Bee Colony Algorithm," *Sains Malaysiana,* vol. 51, no. 8, pp. 2655-2668, 2022, doi.org/10.17576/jsm-2022-5108-24.
[17] R. Ray *et al.*, "Reliability analysis of reinforced soil slope stability using GA-ANFIS, RFC, and GMDH soft computing techniques," *Case Studies in Construction Materials,* vol. 18, p. e01898, 2023, doi.org/10.1016/j.cscm.2023.e01898.
[18] A. G. Ivakhnenko, "Polynomial theory of complex systems," *IEEE transactions on Systems, Man, and Cybernetics,* no. 4, pp. 364-378, 1971.
[19] M. Panahi, O. Rahmati, F. Rezaie, S. Lee, F. Mohammadi, and C. Conoscenti, "Application of the group method of data handling (GMDH) approach for landslide susceptibility zonation using readily available spatial covariates," *Catena,* vol. 208, p. 105779, 2022, doi.org/10.1016/j.catena.2021.105779.
[20] T. T. K. Tran *et al.*, "Enhancing predictive ability of optimized group method of data handling (GMDH) method for wildfire susceptibility mapping," *Agricultural and Forest Meteorology,* vol. 339, p. 109587, 2023, doi.org/10.1016/j.agrformet.2023.109587.
[21] F. U. Khan, F. Khan, and P. A. Shaikh, "Forecasting returns volatility of cryptocurrency by applying various deep learning algorithms," *Future Business Journal,* vol. 9, no. 1, p. 25, 2023, doi.org/10.1186/s43093-023-00200-9.
[22] B. Badyalina, A. Shabri, and M. F. Marsani, "Streamflow estimation at ungauged basin using modified group method of data handling," *Sains Malaysiana,* vol. 50, no. 9, pp. 2765-2779, 2021, doi.org/10.17576/jsm-2021-5009-22.
[23] Y. Qin, R. Langari, and L. Gu, "A new modeling algorithm based on ANFIS and GMDH," *Journal of Intelligent & Fuzzy Systems,* vol. 29, no. 4, pp. 1321-1329, 2015, doi.org/10.3233/ifs-141443.
[24] C. Bosah, S. Li, A. Mulashani, and G. Ampofo, "Analysis and forecast of China's carbon emission: evidence from generalized group method of data handling (g-GMDH) neural network," *International Journal of Environmental Science and Technology,* pp. 1-14, 2023, oi.org/10.1007/s13762-023-05043-z.
[25] A. Mazraeh, M. Bagherifar, S. Shabanlou, and R. Ekhlasmand, "A novel committee-based framework for modeling groundwater level fluctuations: A combination of mathematical and machine learning models using the weighted multi-model ensemble mean algorithm," *Groundwater for Sustainable Development,* vol. 24, p. 101062, 2024, doi.org/10.1016/j.gsd.2023.101062.
[26] F. Hadavimoghaddam *et al.*, "Modeling CO2 loading capacity of triethanolamine aqueous solutions using advanced white-box approaches: GMDH, GEP, and GP," *Discover Applied Sciences,* vol. 6, no. 2, p. 40, 2024, doi.org/10.1007/s42452-024-05674-y.
[27] C. Bosah, S. Li, A. Mulashani, and G. Ampofo, "Analysis and forecast of China's carbon emission: evidence from generalized group method of data handling (g-GMDH) neural

network," *International Journal of Environmental Science and Technology,* vol. 21, no. 2, pp. 1467-1480, 2024, doi.org/10.1007/s13762-023-05043-z.
[28] J. Yan, J. Li, G. Bai, and T. Zong, "Identification of nonlinear system with time delay based on wavelet packet decomposition and Gaussian kernel GMDH network," *International Journal of Systems Science,* pp. 1-17, 2024, doi.org/10.1080/00207721.2024.2317354.
[29] M. Najafzadeh and R. Sheikhpour, "Local scour depth at piles group exposed to regular waves: On the assessment of expressions based on classification concepts and evolutionary algorithms," *Results in Engineering,* vol. 21, p. 101810, 2024, doi.org/10.1016/j.rineng.2024.101810.
[30] R. C. Feenstra, R. Inklaar, and M. P. Timmer, "The next generation of the Penn World Table," *American economic review,* vol. 105, no. 10, pp. 3150-3182, 2015, doi.org/10.1257/aer.20130954.
[31] W. Anggraeni *et al.*, "A hybrid EMD-GRNN-PSO in intermittent time-series data for dengue fever forecasting," *Expert Systems with Applications,* vol. 237, p. 121438, 2024, doi.org/10.1016/j.eswa.2023.121438.
[32] J. Wang, X. Niu, L. Zhang, Z. Liu, and X. Huang, "A wind speed forecasting system for the construction of a smart grid with two-stage data processing based on improved ELM and deep learning strategies," *Expert Systems with Applications,* vol. 241, p. 122487, 2024, doi.org/10.1016/j.eswa.2023.122487.

Biography of all authors

Picture	**Biography**	**Authoship contribution**
	Basri Badyalina is a senior lecturer from the Universiti Teknologi MARA Johor branch.	Analyzing data. Interpreting result and draft article.
	Fatin Farazh Ya'acob is a senior lecturer from the Universiti Teknologi MARA Johor branch.	Design the research work, interpretation, and draft article.
	Rabiatul Munirah Alpandi is a senior lecturer from the Universiti Teknologi MARA Selangor branch.	Collection of environmental data.
	Nur Diana Zamani is a senior lecturer from the Universiti Teknologi MARA Johor branch.	Development of modelling and programming.
	Muhammad Zulqarnain Hakim Abd Jalal is a senior lecturer from the Universiti Teknologi MARA Johor branch.	Development of modelling and programming.
	Amir Imran Zainoddin is a senior lecturer from the Universiti Teknologi MARA Johor branch.	Development of literature review.