Optimized Non-linear Multivariable Grey Model for Carbon Dioxide Emissions in Malaysia

Assif Shamim Bin Mustaffa @ Sulaiman
Department of Mathematical Science,
Universiti Teknologi Malaysia,
Skudai, MALAYSIA
assifshamim@graduate.utm.my

Ani Bin Shabri
Department of Mathematical Science,
Universiti Teknologi Malaysia,
Skudai, MALAYSIA
ani@utm.my

2021 International Congress of Advanced Technology and Engineering (ICOTEN) | 978-1-6654-1224-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOTEN52080.2021.9493552

***Abstract*— This paper analyses the relationship between carbon dioxide emissions with the energy consumption from the year 2005 to 2014 in Malaysia by introducing an optimized non-linear multivariable grey, NGM(1,N) model by establishing a power exponent term for its subsequent relevant factors. The aim of this research is to improve the existing NGM(1,N) model by solving the effect of non-linear properties which is able to correlate among the consequent factors based on the selection of power exponent optimization. This paper will also introduce the transformed NGM(1,N) known as TNGM(1,N) model that produces a more accurate result compared to NGM(1,N) model that prompted simulated output. The power exponent term value was determined using the generalized reduced gradient (GRG) method in Microsoft Excel Solver. It is proven that the TNGM(1,N) model performs the best and hence it serves as vital information for the government's environmental-related agencies and policymakers to focus on the effort to promote green efficient technology to society at large by reducing the releases of carbon dioxide emissions to the environment.**

Keywords—carbon dioxide emissions, energy consumption, optimized non-linear multivariable grey model, grey forecasting

I. Introduction

Grey system theory was proposed by Deng which specialises in the short-term forecast [1]. The word grey suggested by the author indicated that the system contains uncertain and incomplete information from its data points [2]. Hence, the aim of implying the grey model is to close the gap between the known and unknown information in the system and search for its uniformity when forecasting. The specialty of grey model is through the superior method called the accumulated generating coefficient (AGO) that will set the system into a monotonous characteristic. Throughout the years, many researchers have branch out the grey system theory into multiple area of studies such as medical, marine and biology, economic as well as environmental science field [3]–[6]. With many grey models have undergone several and multiple modification to fully optimized its capabilities in forecasting the short-term data, numerous research studies were conducted to improve the currently existing grey forecasting model.

This paper will focus on the study of carbon dioxide (CO_2) emissions using the grey forecasting model with a more suitable model along with a reliable forecasting result. There are numerous methods that can be applied when considering the effect of CO_2 emissions studies such as neural network, computer-based simulation, grey model and others [7], [8]. Some studies conducted on the CO_2 emissions often uses multiple variables which are included in the model building process commonly known as the multivariable grey model or GM(1,N) model [9]. With multiple researches has been performed in the study of forecasting CO_2 emissions, it is important that we utilise the advantage of implementing the grey forecasting model that brings value when there are restriction in acquiring large amount of data to forecast.

A study conducted using an improved Verhulst GM(1,N) with Fourier series for CO_2 emissions in China uses bilateral foreign direct investment (FDI) as its subsequent relevant factors [10]. Although this model is able to precisely illustrate the correlation among the variables, the data fluctuation error still occurs when uses big sample to model and forecast. Another study of non-linear GM(1,N) model with multi-kernel using variable such as the raw and clean coal industry was used to forecast the CO_2 emissions in China [11]. This model was able to minimize the computational error stimulated by the increase of the nonlinearity trend in the system. However, this model suffered from numerical instability when the level of degree increases when using the polynomial kernel function which consequently resulting in zero.

As major development of economy means more energy consumption needed to power through the vast majority of the industry, a comprehensive study on the environmental science field has become a vital topic to be discussed thoroughly. The usage of fossil fuel increases the release of carbon dioxide emissions into the environment, worsen almost all aspect of life including the habitat's ecosystem and threaten the health of all living creatures [12], [13]. Consequently, the impact on the global warming also deteriorated throughout the years. Hence, the study of the CO_2 emissions has become important to ensure the impact of global warming are still reversible.

II. Methodology

A. Non-linear Multivariable Grey Model

Multivariable grey model or GM(1,N) model indicates first order multiple variables. As proposed by Deng, the grey model were able to deal with the uncertainties that possessed both known and unknown information in the grey area in the system [14]. As most environmental science data are in yearly form, GM(1,N) model is one of the most recommended model to forecast CO_2 emissions related data. Let the sequence system's characteristics data denoted as

$$X_1^{(0)} = \left(x_1^{(0)}(1), x_1^{(0)}(2), x_1^{(0)}(3), \ldots, x_1^{(0)}(n)\right) \tag{1}$$

and the subsequent relevant factors are denoted as

$$\begin{aligned} X_2^{(0)} &= \left(x_2^{(0)}(1), x_2^{(0)}(2), x_2^{(0)}(3), \ldots, x_2^{(0)}(n)\right) \\ X_3^{(0)} &= \left(x_3^{(0)}(1), x_3^{(0)}(2), x_3^{(0)}(3), \ldots, x_3^{(0)}(n)\right) \\ &\vdots \\ X_i^{(0)} &= \left(x_i^{(0)}(1), x_i^{(0)}(2), x_i^{(0)}(3), \ldots, x_i^{(0)}(n)\right) \end{aligned} \tag{2}$$

where i=1,2,…,N indicate number of variables and n indicate number of data entries for modelling sets of data. Next, a new sequence using the first order accumulated generating operation (1-AGO) for Eq. 1 and Eq. 2 are calculated as

$$X_i^{(1)} = \left(x_i^{(1)}(1), x_i^{(1)}(2), x_i^{(1)}(3), \ldots, x_i^{(2)}(n)\right) \tag{3}$$

and the sequence of $z_1^{(1)}$ is deduced as

$$z_1^{(1)} = \{z_1^{(1)}(2), z_1^{(1)}(3), z_1^{(1)}(4), ..., z_1^{(1)}(k)\} \quad (4)$$

Eq. 4 is called the mean equation of Eq. 2, and it is calculated as

$$z_1^{(1)}(k) = \omega x_1^{(1)}(k) + (1-\omega) x_1^{(1)}(k-1) \quad (5)$$

The value of ω is fixed at 0.5 when applying the GM(1,N) model. Therefore, the traditional form of a non-linear GM(1,N) model, or NGM(1,N) model is

$$x_1^{(0)} + az_z^{(1)}(k) = \sum_{i=2}^{N} b\{x_i^{(1)}(k)\}^{\chi_i} \quad (6)$$

Parameter a is the grey system's developing coefficient, whereas b_i is the control variable of the system. The power exponent of χ_i represent the non-linear, which signifies the non-linear effect of associated I variables in the system. The value of power exponent χ_i is assumed to be known during the modelling process and can be solved using any method of optimization. In this research, the power exponent χ_i is determined using the generalized reduced gradient (GRG) nonlinear method in Microsoft Excel Solver. When the value of χ_i equals to 1, the model will be reduced to the conventional GM(1,N) model. Hence, the parameter sequence of NGM(1,N) model is

$$\hat{a} = [a, b_1, b_2, ..., b_N]^T \quad (7)$$

By implementing the ordinary least square (OLS) method, parameters a, b_i , and power exponent χ_i can be calculated as

$$[a,b]^T = (B^T B)^{-1} B^T Y_N \quad (8)$$

where

$$B = \begin{bmatrix} -z_1^{(1)}(2) & \{x_2^{(1)}(2)\}^{\chi_i} & \cdots & \{x_N^{(1)}(2)\}^{\chi_i} \\ -z_1^{(1)}(3) & \{x_2^{(1)}(3)\}^{\chi_i} & \cdots & \{x_N^{(1)}(2)\}^{\chi_i} \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(n) & \{x_2^{(1)}(n)\}^{\chi_i} & \cdots & \{x_N^{(1)}(2)\}^{\chi_i} \end{bmatrix}, \text{ and}$$

$$Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{bmatrix} \quad (9)$$

The OLS parameter sequence in Eq. 7 satisfies the following conditions

(1) As $n = N+1$, $\hat{a} = B^{-1}Y_N$, where |B| ≠ 0

(2) As $n = N+1$, $\hat{a} = (B^T B)^{-1} B^T Y_n$, where |B| ≠ 0

(3) As $n = N+1$, $\hat{a} = B^T (B^T B)^{-1} Y_n$, where |B| ≠ 0

Assume $\hat{a} = [a, b_1, b_2, ..., b_N]^T = (B^T B)^{-1} B^T Y$, then the whitenization equation of NGM(1,N) model is

$$\frac{dx_1^{(1)}}{dt} + ax_1^{(1)} = \sum_{i=2}^{N} b_i \{x_i^{(1)}(k)\}^{\chi_i} \quad (10)$$

Therefore, the whitenization solution for NGM(1,N) model is

$$x_1^{(1)}(k) = e^{-at}\left[\sum_{i=2}^{N}\int b_i\{x_i^{(1)}(k)\}^{\chi_i} + x_1^{(1)}(0) - \sum_{i=2}^{N} b_i\{x_i^{(1)}(k)\}^{\chi_i}\, dt\right]$$

$$= e^{-at}\left[x_1^{(1)}(0) - t\sum_{i=2}^{N} b_i\{x_i^{(1)}(0)\}^{\chi_i} + \sum_{i=2}^{N} b_i\{x_i^{(1)}(0)\}^{\chi_i} e^{at} dt\right] \quad (11)$$

The forecasted values from 1-AGO is expressed as

$$\hat{x}_1^{(1)}(k+1) = \left(x_1^{(1)}(1) - \frac{1}{a}\sum_{i=2}^{N} b_i\{x_i^{(1)}(k+1)\}^{\chi_i}\right)e^{-at} + \frac{1}{a}\sum_{i=2}^{N} b_i\{x_i^{(1)}(k+1)\}^{\chi_i} \quad (12)$$

The value of $X_1^{(1)}(0)$ is from $X_1^{(0)}(1)$. Lastly, the restored values of inverse accumulation of NGM(1,N) model is

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k) \quad (13)$$

In this next section, we will focus on an enhance transformed of NGM(1,N) model. In NGM(1,N) model, the results of the whitenization time response prompted approximation values of the model from Eq. 12 cannot be used for forecasting purposes as it induces larger error. In order to further solve this concerns, Eq. 12 will be restored for a more precise during the simulation and forecast calculation. For transformed NGM(1,N) or TNGM(1,N) model, Eq. 6 can be deduced to

$$x_1^{(0)}(k) = \sum_{i=2}^{N} \beta_i \{x_i^{(1)}(k)\}^{\chi_i} - \alpha x_1^{(1)}(k-1) \quad (14)$$

where $\beta_i = \frac{b_i}{1+0.5a}$ and $\alpha = \frac{a}{1+0.5a}$. The proof of the background value of $X_1^{(0)}(k)$ are

$$z_1^{(1)}(k) = 0.5x_1^{(1)}(k-1) + 0.5x_1^{(1)}(k-1) + 0.5x_1^{(0)}(k)$$

$$= x_1^{(1)}(k-1) + 0.5x_1^{(0)}(k) \quad (15)$$

From Eq. (15), the background value of $X_1^{(0)}(k)$ is calculated as

$$x_1^{(0)}(k) = a\left[x_1^{(1)}(k-1) + 0.5x_1^{(0)}(k)\right] = \sum_{i=2}^{N} b_i\{x_i^{(1)}(k)\}^{\chi_i} \quad (16)$$

which subsequently is

$$(1+0.5a)x_1^{(0)}(k) = \sum_{i=2}^{N} b_i\{x_i^{(1)}(k)\}^{\chi_i} - ax_1^{(1)}(k-1) \quad (17)$$

Hence, the forecasted values of TNGM(1,N) model are simulated as

$$x_1^{(0)}(k) = \sum_{i=2}^{N} \frac{b_i}{1+0.5a}\{x_i^{(1)}(k)\}^{\chi_i} - \frac{a}{1+0.5a} x_1^{(1)}(k-1) \quad (18)$$

When the value of χ_i equals to 1, the model will also be reduced to the conventional GM(1,N) model.

B. Mean Absolute Percentage Error

To measure the accuracy of the models, the mean absolute percentage error (MAPE) is used to determine which model was able to produce a better forecasted result. The model with smaller value of MAPE therefore has a better forecast accuracy. The formula of MAPE is as shown

978-1-6654-1224-7/21/$31.00 ©2021 IEEE

$$\text{MAPE} = \frac{1}{m}\left(\left|\frac{\sum(\hat{x}^{(1)} - x^{(0)})}{x^{(0)}}\right|\right) \times 100\% \quad (19)$$

The value $\hat{x}^{(1)}$ indicate the forecasted value and $x^{(0)}$ is the original value. Also, the total number of data points are denoted as m.

III. Results and Discussions

This research will focus on the CO_2 emissions (Mt) for Malaysia from energy consumption (kg of oil equivalent per capita) for the years 2005 to 2014. Table 1 depicts original sequence for both the CO_2 emissions and energy consumption for Malaysia are taken from Department of Environment Malaysia and World Bank Data respectively.

TABLE I. Original sequence of the series

Year	Carbon Dioxide Emissions (Mt)	Energy Consumption (kg of oil equivalent per capita)
2005	1867.2	2558.496
2006	1985.7	2522.331
2007	2533.4	2708.469
2008	2055.3	2777.664
2009	2585.3	2591.76
2010	2623.2	2601.449
2011	2746.1	2670.652
2012	2955.7	2678.534
2013	2955.8	2979.649
2014	3024.9	3003.456

We will compare the forecasted result for conventional GM(1,N) model, which is the GM(1,2) model, together with NGN(1,2) and TNGM(1,2) models for CO_2 emissions for Malaysia data to determine which model were able to produce a better forecasting result with least percentage error. First and foremost, we will use the data range from the years 2005 to 2011 for modelling purpose and 2012 to 2014 for forecasting purpose. Now, we let $X_1^{(0)}$ be the initial series for CO_2 emissions as in Eq. 1 and $X_2^{(0)}$ be the initial series for energy consumption data as in Eq. 2. Using the AGO, a new sequence of time series is generated as shown

$$X_1^{(1)} = (1867.2, 3851.9, \ldots, 16396.2) = \{X_1^{(1)}(n)\}_1^7 \quad (20)$$

$$X_2^{(1)} = (2558.5, 5080.8, \ldots, 18430.8) = \{X_2^{(1)}(n)\}_1^7 \quad (21)$$

All GM(1,2), NGN(1,2) and TNGM(1,2) models uses the OLS method to the mean equation of the sequence $z_1^{(1)}$ with value of ω equals to 0.5. The power exponent value of χ_i for NGN(1,2) and TNGM(1,2) models is solved by Microsoft Excel Solver generated using the GRG nonlinear optimization method. The process of selecting the ideal value of χ_i between -1 and 1 are based of the MAPE values of each model. Hence, the parameter's values of GM(1,2), NGN(1,2) and TNGM(1,2) models are shown in Table 2.

TABLE II. Values of parameters for GM(1,2), NGM(1,2) and TNGM(1,2) models

Model	Parameters a	b_2	χ_2
GM(1,2)	0.863361237	0.843938076	
NGM(1,2)	-0.065723451	3230.418903	-0.060262244
TNGM(1,2)	-0.023199704	616.5556241	0.137653751

Therefore, the simulated and forecasted accuracy analysis result was compared against the original sequence based on all GM(1,2), NGN(1,2) and TNGM(1,2) models as shown in Table 3 using the MAPE as indicator on the accuracy of the models in forecasting. The result showed the TNGM(1,2) model produces the smallest value of MAPE with 1.74% outperform both GM(1,2) model (4.37%) and NGM(1,2) model (3.01%) confirming that TNGM(1,2) model generates significantly better forecasted result, having the forecasted values, $\hat{x}^{(1)}$ imitate closely against the original values, $x^{(0)}$.

TABLE III. Original sequence and simulated result for GM(1,2), NGM(1,2) and TNGM(1,2) models

Year	Original Sequence	GM(1,2) Simulated Result	GM(1,2) APE (%)	NGM(1,2) Simulated Result	NGM(1,2) APE (%)	TNGM(1,2) Simulated Result	TNGM(1,2) APE (%)
2005	1867.2	1867.200		1867.200		1867.200	
2006	1985.7	1792.208		2123.394		2063.008	
2007	2533.4	2932.478		2162.687		2231.939	
2008	2055.3	3102.579		2250.111		2383.221	
2009	2585.3	2820.363		2370.178		2499.926	
2010	2623.2	2710.147		2500.398		2618.482	
2011	2746.1	2700.334		2642.013		2731.452	
2012	2955.7	2664.608	9.85%	2799.202	5.29%	2841.367	3.87%
2013	2955.8	2935.455	0.69%	2955.800	0.00%	2955.800	0.00%
2014	3024.9	2947.195	2.57%	3137.640	3.73%	3065.969	1.36%
MAPE (%)			**4.37%**		**3.01%**		**1.74%**

The core justification that NGM(1,2) model signify better forecast accuracy than the GM(1,2) model is due to the effect of non-linear properties which were able to correlate among the consequent factors based on the selection of power exponent optimization. However, NGM(1,2) model result remains as an approximation values whereas TNGM(1,2) model solution is more accurate in the sense of simulated values, consequently producing better result. Hence, the TNGM(1,2) model result will be set as the benchmark against

978-1-6654-1224-7/21/$31.00 ©2021 IEEE

all future grey forecasting model that uses the CO_2 emissions to conduct a short-term forecast.

IV. Conclusion

The multivariable grey model is an effective forecasting method dealing with short-term multiple variables data. This paper managed to forecast the CO_2 emissions for Malaysia from energy consumption perspective using the GM(1,N) model, the nonlinear form of GM(1,N) named NGM(1,N) model, and the transformed model named TGNM(1,N) model. The power exponent value of χ_i for NGN(1,2) and TNGM(1,2) models are not known when estimating the parameters substituted into the OLS method, and it is determined when the value of MAPE has been calculated using the GRG nonlinear method in Microsoft Excel Solver to achieve an optimized result for the models. Based on the result, TNGM(1,2) model outperformed both GM(1,2) and NGM(1,2) models with smallest MAPE value. It can be summarized that by implementing the TNGM(1,2) model for CO_2 emissions data, the forecasted result able to be simulated with closer approximation values due to the effect of the non-linear properties able to correlate better among the subsequent factors based on the selection of optimized power exponent selected using the GRG nonlinear method which therefore can be concluded that the TNGM(1,2) model is one of the better model to forecast CO_2 emissions. Hence, the aim of this research is to highlight the severity of CO_2 emissions impacting the effect of global warming causing severe natural disaster events, specifically in Malaysia by finding the best forecasting model to analysed and forecast even with the restriction of having a small sample size data. Therefore, with little information that we have to conduct this study, we are still able to anticipate the future and enable to give dynamic information for the government's environmental-related agencies and policymakers to focus on the effort to promote green efficient technology to society at large. As this paper only includes one subsequent factor towards the CO_2 emissions factor, future research may include more factors that might have a direct impact on the CO_2 emissions ranging from economic to demographic factors influencing the main factor.

V. Acknowledgement

The authors sincerely expressed our gratitude to the Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia under Fundamental Research Grant Scheme (FRGS/1/2019/STG06/UTM/02/3) for fully supported and funded this research work.

References

[1] D. Ju-Long, "Control problems of grey systems," *Systems and Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[2] A. S. Mustaffa and A. Shabri, "An Improved Rolling NGBM(1,1) Model with GRG Nonlinear Method of Optimization for Fossil Carbon Dioxide Emissions in Malaysia and Singapore," in *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC 2020)*, 2020, pp. 32–37.

[3] N. Norouzi and M. Fani, "Black gold falls, black plague arise - An Opec crude oil price forecast using a gray prediction model," *Upstream Oil and Gas Technology*, vol. 5, no. May, p. 100015, 2020.

[4] R. Tian, Q. Shao, and F. Wu, "Four-dimensional evaluation and forecasting of marine carrying capacity in China : Empirical analysis based on the entropy method and grey Verhulst model," *Marine Pollution Bulletin*, vol. 160, no. March, p. 111675, 2020.

[5] X. Yang, J. Zou, D. Kong, and G. Jiang, "The analysis of GM (1, 1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China," *Medicine (United States)*, vol. 97, no. 34, 2018.

[6] B. Zeng, M. Zhou, and J. Zhang, "Forecasting the energy consumption of China's manufacturing using a homologous grey prediction model," *Sustainability (Switzerland)*, vol. 9, no. 11, 2017.

[7] K. Mason, J. Duggan, and E. Howley, "Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks," *Energy*, vol. 155, pp. 705–720, 2018.

[8] L. Wen and X. Yuan, "Forecasting CO2 emissions in Chinas commercial department, through BP neural network based on random forest and PSO," *Science of the Total Environment*, vol. 718, p. 137194, 2020.

[9] C. C. Lin, R. X. He, and W. Y. Liu, "Considering multiple factors to forecast CO2 emissions: A hybrid multivariable grey forecasting and genetic programming approach," *Energies*, vol. 11, no. 12, pp. 1–25, 2018.

[10] H. Jiang, P. Kong, Y. C. Hu, and P. Jiang, "Forecasting China's CO2 emissions by considering interaction of bilateral FDI using the improved grey multivariable Verhulst model," *Environment, Development and Sustainability*, no. 0123456789, 2020.

[11] H. Duan, D. Wang, X. Pang, Y. Liu, and S. Zeng, "A novel forecasting approach based on multi-kernel nonlinear multivariable grey model: A case report," *Journal of Cleaner Production*, vol. 260, p. 120929, 2020.

[12] E. T. Busayo, A. M. Kalumba, G. A. Afuye, O. Y. Ekundayo, and I. R. Orimoloye, "Assessment of the Sendai framework for disaster risk reduction studies since 2015," *International Journal of Disaster Risk Reduction*, vol. 50, no. September, p. 101906, 2020.

[13] M. A. Carrington, J. Ranse, and K. Hammad, "The impact of disasters on emergency department resources: review against the Sendai framework for disaster risk reduction 2015–2030," *Australasian Emergency Care*, no. xxxx, 2020.

[14] Z. X. Wang and D. J. Ye, "Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models," *Journal of Cleaner Production*, vol. 142, no. 2017, pp. 600–612, 2017.

978-1-6654-1224-7/21/$31.00 ©2021 IEEE