Forecasting Carbon Dioxide Emissions for Malaysia using Grey Model with Cramer’s Rule

Assif Shamim Mustaffa Sulaiman, Ani Shabri*

Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia

Abstract
This article analyses and forecasts carbon dioxide (CO₂) emissions in Malaysia for the 2014 to 2018 period. The study analysed the data using grey forecasting model with Cramer’s rule to calculate the best SOGM(2,1) model with the highest accuracy of precision compared to conventional grey forecasting model. According to the forecasted result, the fitted values using SOGM(2,1) model has a higher accuracy precision with better capability in handling information to fit larger scale of uncertain feature compared to other conventional grey forecasting models. This article offers insightful information to policymakers in Malaysia to develop better renewable energy instruments to combat the greater issues of global warming and reducing the fossil carbon dioxide emissions into the environment.

Keywords: carbon dioxide emissions, grey forecasting model, Cramer’s rule

Introduction
Carbon Dioxide (CO₂) is among the most dangerous gas particles released into the atmosphere that accumulates and forms the Greenhouse Gases (GHG) which affect the environment adversely and contribute to global warming. There have been alarming increases in global temperatures and rising sea levels [1]. The impact is more severe among developed and developing countries due to increased GHG emissions resulting from several activities such as burning of fossil fuels from industries and transportation as well as waste management site [2–4].

According to the United Nations Department of Economic and Social Affairs in Sustainable Development, it is reported that before the event of the COVID-19 pandemic takes place, the year of 2019 was the second warmest year ever recorded compared to the year of 2016. While the ongoing pandemic might cause GHG to drop around 6% in the year 2020, we are still short of 7.6% of reduction required to limit the global warming impact of 1.5 degree Celsius ensuring the event of climate change is reversible [5]. With only 85 countries having their own national natural disaster reduction strategies which aligned with the Sendai Framework, the effect of climate change remains to exacerbate the frequency and severity of natural disasters such as massive wildfires, extreme hurricanes, floods, and droughts affecting more than 39 million people in the year of 2018.

Asia continent emits half of global CO₂ emissions, with a significant increase from 1440.06Mt in 1965 to 17586.01Mt in 2015 where China is the largest contributor in Asia, whereas Qatar were the largest contributor in the context of CO₂ emissions per capita [6]. According to Qiao et al, six out of ten largest carbon dioxide emitters in the world are within the Asia Pacific Economic Cooperation (APEC) region.
with four countries belong in the top five positions [7]. In ASEAN, the usage of mix-energy sources from fossil fuels such as coal, oil and gas were among the largest GHG contributors, putting ASEAN region at risk of severe impact of global warming as it is expected to reach 2.3 billion tons by the year 2040 [8]. The Climate Risk Index 2019 report also stated that two out of ten countries most influenced by climate change which are Vietnam and Thailand were ranked fifth and tenth, were both located in the region of ASEAN, while the 2020 version report ranked Philippines as second in the Climate Risk Index due to Typhoon Mangkhut was considered to be one of the most disastrous typhoon ever recorded in the year of 2018 [9], [10]. As energy sector in Malaysia are among the highest carbon dioxide emissions contributor, the projection of electricity sector are moving in an unsustainable path as the demand of electricity increasing each year with the projection of carbon dioxide in the year 2050 ranging from 200 Mt up to 400 Mt [11]. In recent years, agricultural sectors are also contributing to the amount of GHG emissions globally with 11.8% and 8.22% alone are generated by Malaysia in 2016 [12]. It is forecasted that if no immediate plan of action being implement, the global trend will increase significantly by 58% in the year 2050.

Over the years, researchers from all around the world have performed studies related to the impacts of climate change, particularly on the study of CO$_2$ emissions forecast as the result concerns of acute global warming effect, as well as hazardous health-related issues that are linked to the poor air quality. There are diverse methods that can be carried out in the study of CO$_2$ emissions, each with its own pros and cons depending on the type and variation of the data irrespective of using a computer simulation-based analysis or a statistical analysis models. Abdullah & Pauzi reported that the most used method CO$_2$ emissions forecast are neural network (NN) (47.76%), followed by grey model (14.39%), computer based simulation and Intergovernmental Panel on Climate Change (IPCC) method (9.52% respectively), optimal growth model and fuel analysis (7.14% respectively), and others (4.76%) [13].

Statistical analysis forecast based on historical data is often applied by researchers particularly with statistical background to predict a certain event in future corresponding to the nature of the data. With numerous researchers looking for a new model to improve currently existing method, it is crucial for us to experiment with multiple methodology to make sure the research is up to date with the aspect of globalisation and modern landscape. Methods such as regression analysis, time series forecast models, as well as econometrics analysis usually requires a large sample size of data to make an inform and unbiased forecast result. However, due to the restriction of some of the data having smaller sample size (such as in the form of yearly data), it is important for the result reflecting the data in both the simulation which indicates the values based on the model and forecast output values which represent the forecasted values based on the simulation values. In the study of CO$_2$ emissions forecast, the usage of grey forecasting model has been widely used due to its property of generating operation mechanism that allows the data to create a regularity of the sequence suitable when there is a limitation in the total number of data set.

A study was conducted to forecast the CO$_2$ emissions, renewable energy consumption and economic growth in Vietnam using two different type of grey model namely GM(1,1) model, and a discrete GM(1,1) model also known as DGM(1,1) model [13], [14]. This study only uses 4 point of data sets from the years 2010 to 2014 to forecast all three variables using both type of grey models. The results indicated that both models perform well for all three variables with values of mean relative error (MRE) less than 5%, with DGM(1,1) model slightly outperform the traditional GM(1,1) model. This further prove that the grey model is the golden standard to forecast short-term data given the higher precision accuracy result.

The Mean Absolute Percentage Error (MAPE) is used to calculate the measurement accuracy of the model. This research will focus on the study of an optimized GM(2,1) model, called SOGM(2,1) model which improved the current method by structing a new grey system with an optimization method, the solving of reflection equation is derived on least square method, and introducing a new method when creating original values of the time response function. The SOGM(2,1) model with the application of Cramer’s rule outperformed other conventional grey forecasting model hence is in this research [15].
Methods

As proposed by Deng, the term “Grey” is referred to being incomplete or uncertain in the context of information, as opposed to the term “White” as having a complete and certain information. The information is considered to be poor when the system having insufficient information in order to explain the nature of the data structure. The grey system theory is widely used for short-term model forecast, to building of the grey model through the superior process of accumulated generating operation that set the system from its monotonous characteristics.

For this section, we will introduce a new form of an optimized grey model with two number of order of the system and one variable, known as structure optimized grey model or SOGM(2,1) model accomplish a better accuracy compared to other conventional grey forecasting available. Let the initial series of CO$_2$ emissions in Malaysia denoted as

$$X_i^{(0)} = \{x_i^{(0)}(1), x_i^{(0)}(2), x_i^{(0)}(3), \ldots, x_i^{(0)}(n)\}$$ (1)

where $i=1,2,\ldots,n$ is the total number of variables and $n$ indicate the number of data entries for modelling sets of data. As mentioned in section one, the superiority of using a grey system is it only requires a small sample size to attain a good forecasting result with $n$ can be as small as 4 and this research will show the example of it. The first order accumulated generating operation (1-AGO) to the Equation 1 is implemented to reduce the randomization and maintain a monotonous sequence of the system. Therefore, new sequence series of $X_i^{(1)}$ is obtained in Equation 2.

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

The construction of $Z_i^{(1)}$ background sequence

$$Z_i^{(1)} = \{Z_i^{(1)}(2), Z_i^{(1)}(3), Z_i^{(1)}(4), \ldots, Z_i^{(1)}(k)\}$$ (3)

is called the mean equation from Equation 2 and $k=1,2,\ldots,n$, shown in Equation 4.

$$Z_i^{(1)}(k) = \omega x_i^{(1)}(k) + (1-\omega) x_i^{(1)}(k-1)$$ (4)

where classically the grey forecasting model opts for the value of $\omega=0.5$.

Next, the grey differential equation for SOGM(2,1) model is denoted as

$$a_1 x_i^{(-1)}(k) + x_i^{(0)}(k) + a_2 z_i^{(1)}(k) = b$$ (5)

The parameters of $a_1$ and $a_2$ are referred to as lower and higher order parameters correspondingly, and $b$ is the control variable of the system. $z_i^{(1)}$ is called background sequence as indicated in Equation 3. It hence formed a sequence of parameters as

$$\hat{a} = [a_1, a_2, b]^T$$ (6)

The SOGM(2,1) model will opt the ordinary least square (OLS) method to estimate the parameter vectors $a_1$, $a_2$ and $b$ in $\hat{a}$ and is calculated using Equation 7 as follows

$$[a_1, a_2, b]^T = (B^T B)^{-1} B^T Y_N$$ (7)

Where
Let
\[ \hat{a} = [a_1, a_2, b]^T = (B^T B)^{-1} B^T Y_N \]

Then, we will now construct the grey whitenization equation of SOGM(2,1) model for equation
\[ a_1 x_1^{(-1)}(k) + x_1^{(0)}(k) + a_2 z_1^{(0)}(k) = b \]

as
\[ \hat{a}_1 \frac{d^2 x_1^{(-1)}(t)}{dt^2} + \frac{dx_1^{(0)}(t)}{dt} + \hat{a}_2 x_1^{(0)}(t) = \hat{b} \]  

(9)

By solving the grey whitenization equation and acquire the homogeneous differential equation of Equation 9 constructed in Equation 10 as follows

\[ \hat{a}_1 \frac{d^2 x_1^{(0)}(t)}{dt^2} + \frac{dx_1^{(0)}(t)}{dt} + \hat{a}_2 x_1^{(0)}(t) = 0 \]

(10)

The characteristic equation is denoted as
\[ \hat{a}_1 r^2 + r + \hat{a}_2 = 0 \]

(11)

To study the simulation function, we now let \( K = \hat{a}_1 \hat{a}_2 \). The main purpose of \( K \) is to verify for the system development pattern in SOGM(2,1) model. The criteria and justification of \( K \) function is explained in Table 1.

Based on Equation 11, the time response function can be determined by solving using the conventional quadratic formula to calculate the values of \( r_1 \) and \( r_2 \). These time response function serve to invigorate a specific pattern in the grey systems. When \( K > \frac{1}{4} \), it indicates a periodic trend that combines with power growth, and the time response function is denoted as
\[ x_1^{(0)}(t) = g(t) = e^{\gamma t} \left[ \left( \gamma c_1 + \beta c_2 \right) \cos \beta t + \left( \gamma c_2 - \beta c_1 \right) \sin \beta t \right]. \]

(12)

When \( K = \frac{1}{4} \), it indicates a non-linear development that combines linear and exponential growth in the system, and the time response function is denoted as
\[ x_1^{(0)}(t) = g(t) = \left( r c_1 + r c_2 t + c_2 \right) e^{\gamma t}. \]

(13)

Whereas when \( K < \frac{1}{4} \), it indicates a quasi-exponential law range with a definite deviation. The time response function when \( K < \frac{1}{4} \) is denoted as
\[ x_1^{(0)}(t) = g(t) = c_1 r_1 e^{\gamma t} + c_2 r_2 e^{\gamma t}. \]

(14)
Table 1. The criteria and justification of K function

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>K &gt; 1/4</td>
<td>The solution of SOGM(2,1) has a pair of complex conjugate roots, ( r_1, r_2 = \gamma \pm \beta )</td>
</tr>
<tr>
<td>K = 1/4</td>
<td>The solution of SOGM(2,1) has a pair of equal real roots, ( r_1, r_2 = -1/2a )</td>
</tr>
<tr>
<td>K &lt; 1/4</td>
<td>The solution of SOGM(2,1) has a pair of unequal real roots</td>
</tr>
</tbody>
</table>

Hence, the constant of the time response function can be established using the real data point according to the significance of the latest information possess in the system. After the algorithm is program, an evaluation of the time response function is determined used to develop the discrete simulation values. For SOGM(2,1) model, a total of two different initial terms is needed for the time response function. The priority is given to the latest information when handling a small data modelling according to the grey system theory.

Presume the time response function could be describe as \( \hat{x}^{(i)}(t) = g(t, c_1, c_2, ..., c_q) \) where \( 1 \leq q \leq n \).

Hence the restraint conditions for the function prioritizing the latest information is shown as

\[
\begin{align*}
g(n, c_1, c_2, ..., c_q) - x_1^{(0)}(n) &= 0 \\
&
\vdots \\

\end{align*}
\]

(15)

The criteria when \( q = 2 \) are

\[
\begin{align*}
g(n, c_1, c_2) - x_1^{(0)}(n) &= 0 \\
g(n-1, c_1, c_2) - x_1^{(0)}(n-1) &= 0.
\end{align*}
\]

(16)

When \( K < 1/4 \), the values of \( c_1, c_2 \) satisfying criteria in Equation 15 ought to be

\[
\begin{align*}
c_1 &= \frac{r_1 e^{\gamma(n-1)} x_1^{(0)}(n) - r_2 e^{\beta(n-1)} x_1^{(0)}(n-1)}{r_1 r_2 e^{\gamma(n-1)} - r_1 e^{\gamma(n-1)} - r_2 e^{\beta(n-1)}}, \\

c_2 &= \frac{r_1 e^{\gamma(n-1)} x_1^{(0)}(n-1) - r_2 e^{\beta(n-1)} x_1^{(0)}(n)}{r_1 r_2 e^{\gamma(n-1)} - r_1 e^{\gamma(n-1)} - r_2 e^{\beta(n-1)}}.
\end{align*}
\]

(17)

Using Equation 15, the equation system for Equation 16 can be substitute with an initial sequence in the form of matrix as shown below

\[
\begin{bmatrix}
r_1 e^{\gamma n} & r_2 e^{\beta n} \\
r_1 e^{\gamma (n-1)} & r_2 e^{\beta (n-1)}
\end{bmatrix}
\begin{bmatrix}
c_1 \\
c_2
\end{bmatrix}
= 
\begin{bmatrix}
x_1^{(0)}(n) \\
x_1^{(0)}(n-1)
\end{bmatrix}.
\]

(18)

According to Cramer’s rule, the solution of equation is given as
The challenges of making a statistical analysis depends on the core of selecting an appropriate model which best assessing the circumstances of the study. It is fairly important to choose a suitable evaluation instrument to accurately summarize the effectiveness of the proposed model that “best fit” the quality of the model in the process of developing a forecasting model. The definition of model identification or model evaluation terms commonly used to critique the superiority of the proposed model and it varies for each type of model. In general, the grey forecasting model evaluated using several tools to explain the advantage of the proposed grey system against other existing models through model selection.

There are several methods of measurement that can be implemented to evaluate the effectiveness of the model in terms of its performance in modelling and forecasting. The chosen for the accuracy precision measurement in this study used is the MAPE which calculate the proximity between the actual value and the forecasted value, as shown

\[ \text{MAPE} = \frac{1}{n} \left( \sum \frac{|\hat{x}_i^{(1)} - x_i^{(0)}|}{x_i^{(0)}} \right) \times 100\% \]  

Where \( x_i^{(0)} \) indicates the actual value from of CO\(_2\) emissions data, \( x_i^{(1)} \) indicates the forecasted value of CO\(_2\) emissions based on the model value, and the number of observation of data is indicated by \( n \). The smaller the value of MAPE, the better the level of accuracy of the model. Table 2 shows the level of accuracy when using MAPE in the process of model performance evaluation.

<table>
<thead>
<tr>
<th>Level of Accuracy</th>
<th>( \leq 10% )</th>
<th>10% - 20%</th>
<th>20% - 30%</th>
<th>( \geq 30% )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highly Accurate</td>
<td>Good</td>
<td>Reasonable</td>
<td>Inaccurate</td>
</tr>
</tbody>
</table>

Results and discussion

This study uses CO\(_2\) emissions data for Malaysia from the years 2014 to 2018. As the data size is relatively small and cannot be reach with a common statistical requirement, the use of grey model is sufficient as an option. To generate the SOGM(2,1) model, we let \( X_1^{(0)} \) set as the original sequence of the data. Next, we let \( X_1^{(1)} \) being the accumulated generating operation (AGO) sequence and \( X_1^{(-1)} \) being the inverse accumulated generating operation (IAGO) sequence. Using the ordinary least square (OLS) method, we may construct SOGM(2,1) model by using the value of \( \omega \) as 0.5 for data from years 2012 to 2016, and the parameters vector of \( \hat{a} \) is obtained in Table 3.

Next, we let \( X_1^{(1)} \) being the AGO sequence and \( X_1^{(-1)} \) being the inverse accumulated generating operation (IAGO) sequence. As for this research, we will be applying the ordinary least square (OLS) method, we may construct SOGM(2,1) model by using the value of \( \omega \) as 0.5 for data from years 2012 to 2016, and the parameters vector of \( \hat{a} \) is obtained in Table 3.
Table 3. Parameters of $a_1$, $a_2$ and $b$ for Malaysia using OLS method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Malaysia</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>-0.934847928</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.000749347</td>
</tr>
<tr>
<td>$b$</td>
<td>219.8718315</td>
</tr>
</tbody>
</table>

Using all the values from parameters of vector in Table 3, we shall construct the grey whiteningization equation. Then, we solve the grey whiteningization equation and obtain the time response function and its characteristic equation is as follows:

$$-0.9348t^2 + r - 0.0007 = 0$$  \((21)\)

Solving the characteristics equation and acquire the solution of $r_1$ and $r_2$ by using the conventional quadratic formula is shown in Table 4.

Table 4. Solution of $r_1$ and $r_2$ for Malaysia from time response function

<table>
<thead>
<tr>
<th>Solution</th>
<th>Malaysia</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_1$</td>
<td>0.020073957</td>
</tr>
<tr>
<td>$r_2$</td>
<td>2.013387067</td>
</tr>
</tbody>
</table>

Next, we could calculate the value of $K$ for Malaysia is -0.0008. Since the value of $K < \frac{1}{4}$, the time response function for Malaysia is indicated as shown below

$$\hat{x}_1^{(0)}(t) = g(t) = -c_1 0.0007 e^{-0.0007 t} + c_2 1.0704 e^{1.0704 t}.$$  \((22)\)

Therefore, the time response function can be attained by calculate using the Cramer’s rule for the constants of $c_1$ and $c_2$ for Malaysia is as shown

$$\hat{x}_1^{(0)}(k) = g(k) = 216.73 e^{0.0201 t} + 0.0574 e^{2.0134 t}.$$  \((23)\)

The forecasted values for SOGM (2, 1) model can now be obtained based on Equation 23. Since the time response function of $K < \frac{1}{4}$, we will calculate the forecasted values of SOGM(2, 1) model as shown in the Equation 14. The value of $t$ is determined based on the number of order for each year, i.e. year 2014 will denote $t=1$, 2015 denotes $t=2$ and so forth.

Table 5 show that the comparison of forecasting value of CO$_2$ emissions from the years 2014 to 2018 using the SOGM(2, 1) model with comparison with GM(1, 1) model and DGM(2, 1) model. The result show that the value of MAPE using the SOGM(2, 1) model able to perform better compared to the other conventional grey forecasting model with MAPE value for CO$_2$ emissions in Malaysia using the SOGM(2, 1) model (0.74%) compared to GM(1, 1) model (1.05%) and DGM(2, 1) model (3.54%). The graph of the forecasted value is shown in Figure 1.
Table 5. Comparison of MAPE using SOGM(2,1) model vs. other conventional grey forecasting model for CO₂ emissions in Malaysia

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Value</th>
<th>Forecasted Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GM(1,1)</td>
</tr>
<tr>
<td>2014</td>
<td>220.19</td>
<td>220.19</td>
</tr>
<tr>
<td>2015</td>
<td>220.24</td>
<td>217.09</td>
</tr>
<tr>
<td>2016</td>
<td>216.33</td>
<td>219.81</td>
</tr>
<tr>
<td>2017</td>
<td>220.24</td>
<td>222.57</td>
</tr>
<tr>
<td>2018</td>
<td>228.04</td>
<td>225.37</td>
</tr>
<tr>
<td>MAPE</td>
<td>1.05%</td>
<td>3.54%</td>
</tr>
</tbody>
</table>

Conclusions

This paper focuses on the forecasting of CO₂ emissions for Malaysia for years 2014 to 2018 using the SOGM(2,1) model. The SOGM(2,1) model having certain key features such as it operates background sequence and inverse-AGO function to construct an optimized grey system with the estimation of parameters and solving the reflection equation using the OLS method. This model also introduces a new method that prioritizing the most recent information when creating original values of the time response function using the Cramer’s rule. Thus, this study proven that the SOGM(2,1) model has a better forecasting ability with better capability in handling information to fit larger scale of uncertain feature compared to other conventional grey forecasting models. Some future suggestion or improvement that can be taken into consideration for further enhancement on SOGM(2,1) is during the selection of parameters, replacing the traditional OLS method when generating the matrices of 1-AGO that cater to the needs and suitability of the data. This paper will, in hope, shed the light of the issues of global warming especially with the number of CO₂ emitted to the environment increase significantly throughout the last few decades, and hoping there will be changes made by the government official that will resolve this issues before it is too late to turn back.
Data availability

The data was generated from Key Energy Statistics, 2018, International Energy Agency database by filtering the information of carbon dioxide emissions in Malaysia for the years 2014 to 2018.

Conflicts of interest

The authors declare no conflict of interest on the subject of the publication of this paper.

Funding statement

Both authors would like to genuinely thank the institution of Universiti Teknologi Malaysia (UTM) for providing the research funding on this project under Grant Vot 5F271.

References