Journal of Physics: Conference Series

Prediction of CO_2 emissions in Saudi Arabia using Nonlinear Grey Bernoulli Model NGBM (1,1) compared with GM (1,1) model

To cite this article: Z F Althobaiti and A Shabri 2022 *J. Phys.: Conf. Ser.* **2259** 012011

Prediction of CO_2 emissions in Saudi Arabia using Nonlinear Grey Bernoulli Model NGBM (1,1) compared with GM (1,1) model

Z F Althobaiti[1] and A Shabri[2*]

[1] Department of Statistics, Faculty of Science, University of Tabuk, 47512, Tabuk, Saudi Arabia.
[2] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia

Email: zalthebiti@ut.edu.sa

Abstract. One of the most critical solution for tackling the challenges of global warming and climate change is to study and know the accurate prediction of carbon dioxide (CO_2) emissions. Thus, aid to develop appropriate strategic plans that will reduce future damages caused by these emissions into the atmosphere. This study utilizes annual time series data on CO_2 emissions in Saudi Arabia from 1970 to 2016. The goal of this study is to predict CO_2 emissions using the Nonlinear Grey Bernoulli model NGBM (1,1), and compared with the GM (1,1) model based on MAPE metrics to achieve a high-accuracy prediction. The NGBM (1,1) is a newly created grey model with wide ranging applications in diverse fields due to its precision in handling small time-series datasets with nonlinear variations. The NGBM (1,1) with power γ is a nonlinear differential equation that can control the predicted result and adjust the solution to fit the 1-AGO of previous raw data. Thus, the findings show that at sample sizes of N=10 and N=5, the Nonlinear Grey Bernoulli Model (NGBM) is more precise than the Grey Model GM (1, 1). The findings could help the government develop future economic policies.

Keywords— CO_2 emissions, Global warming, Grey Model (GM), Prediction, Nonlinear Grey Bernoulli Model (NGBM), Saudi Arabia

1. Introduction

In recent years, one of the major topics on international political plans for global warming has been climate change. This is because of greenhouse gas emissions, mainly carbon dioxide (CO_2) in the atmosphere [1], [2]. CO_2 is a type of greenhouse gas (GHGs) emitted due to human activities. Human activities are among the primary drivers of CO_2 emissions, with the most important being the generation of energy from coal, oil, and natural gas, and the use of petroleum products for transportation, aircraft, and vehicle trips.

Saudi Arabia is among the rich oil and industrial countries that will be severely affected by CO_2 emissions and leading to global warming. However, the resulting losses from CO_2 emissions exceed what was expected from the industries. This is in corroboration with the study of [3], who estimated that the size of the economic losses that will appear again in the economic results of developing countries, would be greater than their previous benefits from the fossil fuel economy. Though, the three largest countries that are much concerned of the climate change are the United States, Saudi Arabia, and

China were also ranks as the largest country in the world in terms of CO_2 emissions.

Another study by [4] also warned that failure to reduce greenhouse gas emissions would inevitably lead to sea-level rise, which are expected to bring severe economic consequences for the world. For instance, with temperatures reaching pre-industrial levels, floods from sea-level rise could cost humanity $14 trillion annually by 2100**.** Therefore, this has resulted in the prediction of CO_2 emissions, which is the most significant task in time series analysis. The prediction of CO_2 emissions involves predicting the values of the time series from the observed time series. CO_2 emissions predictions have become a global concern, as it has shown to assist in raising public knowledge about how to forestall environmental issues [5]. Therefore, to make a realistic estimate of Saudi Arabia's future CO_2 emissions, a fuller understanding of the country's previous CO_2 emission path is essential. This study aims to model and predict CO_2 emissions in Saudi Arabia using the Nonlinear Grey Bernoulli model NGBM (1,1) as compared with the GM (1,1).

[6] and [7] termed the recently created Nonlinear Grey Bernoulli Model (NGBM (1,1) as precise in handling small time-series datasets with nonlinear variations. Also, in the book published by [8] termed the NGBM (1,1) as more flexible than the GM (1,1). This is because of the NGBM (1,1) model's versatility in determining annual unemployment statistics in various nations. This is used to assist governments in developing future labor and economic policy. In 2005, NGBM (1,1) was also employed to predict the foreign exchange rates of twelve of Taiwan's major trading partners. Both experiments mentioned above revealed that the NGBM (1,1) could increase the accuracy of the original GM (1,1) simulation and forecasting predictions. Recently, some researchers attempted to improve the NGBM (1,1) in various ways, such as [9] who used a particle swarm optimization approach to determine the parameter value of "n", and employed the model to predict the power load of the Hubei electric power network [10]. The genetic algorithm was used to optimize the parameters of the NGBM (1,1), which was then employed to predict economic developments in Taiwan's integrated circuit industry. A paper by [11], studied fuel combustion-related CO_2 emissions forecasting using a novel continuous fractional nonlinear grey Bernoulli model with grey wolf optimizer. The study is critical for framing and implementing reasonable plans and policies, owing to diverse national energy structures. Therefore, by simultaneously incorporating conformable fractional accumulation and derivative into the traditional NGBM (1,1) model, it can capture the nonlinear characteristics hidden in sequences. The author thus developed a novel continuous fractional NGBM (1,1) model, dubbed CCFNGBM (1,1), to accurately project CO_2 emissions from fuel combustion in China by 2023. GWO is used in the study to determine the developing coefficients to enhance the predictability of the newly provided model. However, by replacing the fractional derivative with the integer-order derivative, the model not only improves on the grey forecasting model, but it also provides decision-makers with more dependable forecasts.

This study thus aimed to predict CO_2 emissions in Saudi Arabia using the Nonlinear Grey Bernoulli model NGBM(1,1), and compared with the GM(1,1) based on MAPE metrics to achieve a high-accuracy prediction of CO_2 emissions in Saudi Arabia using the proposed model- NGBM(1,1).

2. Research Methodology (Framework)

To accomplish this research, the following research framework was espoused as presented in Figure 1.

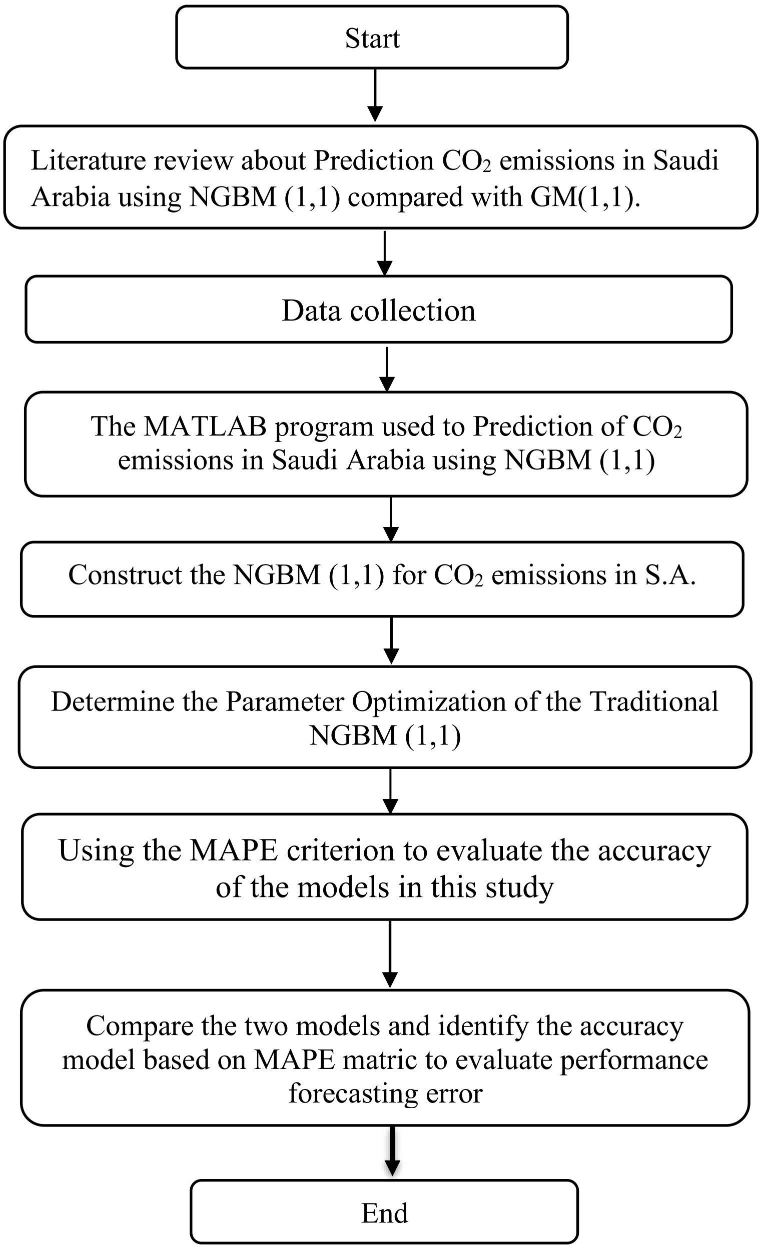

Figure 1. Flow chart of NGBM (1,1)

2.1. The Basic NGBM (1,1)

The GM (1,1) method requires obtaining initial data to generate a regular creation sequence for constructing the model. Though, the generative model predicts the original data processing data. The nonlinear Bernoulli grey prediction model is based on the GM (1,1) and the differential equation of the modeling to enhance prediction accuracy. This model is commonly utilized by [12] and [13]. Also, [6] proposed the Nonlinear Bernoulli Grey Model NBGM(1,1) to improve prediction accuracy when compared to the original GM(1,1) model. To achieve this, the following sequence was proposed.

Step 1: Create a starting sequence depending on the data collected.

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

Where $x^{(0)}(i)$ is the baseline data (state = 0) for time i .

That $x^{(0)}$ is a non-negative sequence, and that n is the sample size. Thus, four data can create and operate a GM (1, 1) model.

Step 2: From the start sequence $x^{(0)}$, generate the first-order Accumulated Generating Operation (AGO) sequence $x^{(1)}$.

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots\ldots, x^{(1)}(n)\right) \quad , \qquad n \geq 4 \tag{2}$$

where $x^{(1)}(k)$ is derived as the following formula:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \quad , \quad k = 1, 2, 3, \ldots. n \tag{3}$$

Step 3: Calculate the first-order AGO sequence's mean value.

The following is the definition of the average sequences generator:

$$z^{(1)} = \left(z^{(1)}(1), z^{(1)}(2), \ldots\ldots, z^{(1)}(n)\right)$$

In which $z^{(1)}(k)$ is the background value sequence taken to be the mean generation of consecutive neighbors of $x^{(1)}$ where

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \ , \ k = 2, 3, \ldots., n \tag{4}$$

The NGBM(1, 1) model is represented as;

$$x^{(0)}(k) + a\, z^{(1)}(k) = b(z^{(1)}(k))^{\gamma}, \quad \gamma \neq 1 \tag{5}$$

Which is the whitening equation of the NGBM (1, 1) model

Step 4: Define the sequence $x^{(1)}$ first-order differential equation is:

$$\frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b(x^{(1)})^{\gamma} \tag{6}$$

The nonlinear parameter γ is given as one, while the linear parameters a and b are determined using the least-squares approach.

Step 5: Assuming the power exponent g is already known, the NGBM(1,1) with the last two parameters are determined as follows:

$$[a,b]^T = \left(B^T B\right)^{-1} B^T Y$$

In which T is the matrix transpose. As a result:

$$\begin{bmatrix} Y = \left[x^{(0)}(2), x^{(0)}(3), \ldots\ldots, x^{(0)}(n)\right]^T \\ B = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^{\gamma} \\ -z^{(1)}(3) & (z^{(1)}(3))^{\gamma} \\ . & \\ . & \\ -z^{(1)}(\mathrm{n}) & (z^{(1)}(\mathrm{n}))^{\gamma} \end{bmatrix} \end{bmatrix} \tag{7}$$

Step 6: the following is the solution to the whitening equation:

$$\hat{x}^{(1)}(k+1) = \left\{ \frac{b}{a} + \left[\left(x^{(0)}(1)\right)^{1-\gamma} - \frac{b}{a}\right] e^{-(1-\gamma)ak} \right\}^{\frac{1}{1-\gamma}} \tag{8}$$

Step 7: Compute the original sequence's prediction value:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), \quad k = 2, 3, \ldots m \tag{9}$$

The NGBM model is a substantial nonlinear grey prediction model in which the power exponent is crucial in grey systems theory. The NGBM model is The GM (1,1) model, especially when γ = 0. The NGBM model is the grey Verhulst model (GMV) when γ = 2. Thus, the GM(1,1) and GMV models, in particular, can be considered as versions of the NGBM model. On the other side, the NGBM model can be thought of as a combination of the GM and GMV models. Therefore, the effectiveness of the NGBM model involves specific approaches that may be employed to identify the appropriate power exponent value, which matches the actual data. As a result, the NGBM model can adequately describe the nonlinear properties of real data and improve simulation and prediction accuracy. [14] used the core principle of information overlap in grey systems to determine the estimated arithmetic of power exponent in the NGBM model. The non-linear programming approach can then be used to calculate the power exponent to minimize mean absolute percentage error (MAPE) [15].

2.2 Parameter Optimization of the Traditional NGBM (1,1)

The traditional NGBM (1,1) help to determine the expected values for the optimization problem. However, [16] proposed a relatively simple iterative method for determining the optimal γ.

$$\min_{\gamma} MAPE = \frac{1}{n-1}\sum_{k=2}^{n}\left|\frac{\hat{x}^{(0)}(k)-x^{(0)}(k)}{x^{(0)}(k)}\right|\times 100\% \quad (10)$$

Where

$$\gamma \neq 1$$

$$[B_1, B_2]^T = \left(B^T B\right)^{-1} B^T Y \quad B, Y \ \ in \ \ (7)$$

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), \ \ k = 2,3,....m$$

3. Model Evaluation

The Mean Absolute Percentage Error (MAPE) was used to evaluate the accuracy of the model in this study. This is a widely used criterion for determining the accuracy of predictions. This is presented below;

$$MAPE = \frac{1}{n}\left(\left|\frac{\sum_{i=1}^{n} x_i - \hat{x}_i}{x_i}\right|\right) \times 100\% \quad (11)$$

Where MAPE refers to Mean Absolute Percentage Error, $\hat{x}_i$ is the predicted value, x_i is the actual value, andproposed a criterion for MAPE as shown [17] denotes the number of data observations. n in Table 1.

Table 1. MAPE for model evaluation.

MAPE (%)	Forecasting power
>50	Inaccurate
20-50	Reasonable
10-20	Good
<10	Highly accurate

Source: [17]

Hence, for a good forecast, the obtained MAPE. It should be as small as possible [18].

4. Data Processing and Analysis

This study is based on 47 yearly CO_2 emissions (kt) observations in Saudi Arabia from 1970 to 2016. The World Bank's online database, which is respected for its trustworthiness and integrity worldwide, provided all the data employed for analysis. The analysis involves using Nonlinear Grey Bernoulli Model (NGBM) and Grey Model (GM) to predict CO_2 emissions. Using the two models, NGBM (1,1) and GM (1,1), Excel 2016 was used to create a database of annual CO_2 emissions in Saudi Arabia from 1970 to 2016.

4.1 Empirical Results

The CO_2 emissions (kt) in Saudi Arabia were forecasted using the NGBM (1,1) and GM(1,1) models. The model's predicting effect was evaluated by comparing predictions to actual values. MAPE was used to assess the models' performance, and the findings revealed that the NGBM (1,1) model performed better in predicting than the GM (1,1) model. Table 3, Figure 1, and Figure 2 show the forecasting of these two models.

This study divides the data into two categories, with 80 % being used for modeling and 20 % for predicting. This means that the year from 2002 to 2011 is used for modeling at N=10 and from 2012 to 2016 for simulation and forecast at N=5 by using Microsoft Excel. When comparing the modeling data at N=10 with the simulation and prediction at N=5, the findings revealed that the model's MAPE values are 3.79%. In comparison, the MAPE values for simulation and forecast data are 8.63%, as shown in Table 3. It is known that the lower the MAPE value, the more accurate the model. Thus, this indicates that the smaller size of data influences the MAPE value for simulation and forecast data, which increases the value of the model. As for the GM (1,1) model, when comparing the modeling data at N=10 with the simulation and forecast at N=5, revealed that the MAPE value for modeling is 4.42%.

In comparison, the MAPE value for simulation and forecast data is 6.02%. This shows that the smaller size of data influences the MAPE value for simulation and forecast data, which increases the model's value. This is because it is well known that the lower the MAPE value, the more accurate the model.

Accordingly, the parameter computation categorized GM (1,1) into two variables; a and b for ordinary least squares calculation, and the output is actual GM (1,1). In contrast, only variable a and b must be simulated with γ=0. The other is determined using the three unknown NGBM (1,1) variables a, b, and γ, as given in Table 2. After computing the Grey Model and Non-linear Grey Bernoulli Model with raw historical data, forecast or predict generation. Table 2 shows the variables for each model that are compatible with the best response resultant value. The GRG Nonlinear method of optimization, first devised by Leon Lasdon and Alan Waren, was used to determine the value of the index (Power Exponent) γ [19]. Its implementation as a FORTRAN software for addressing small to medium-sized issues and some computational findings solved the nonlinear optimization problem. Thus, the value of MAPE was calculated using the NGBM (1,1) at each data point to be predicted by setting the minimum value of MAPE in Eq. (10). This is done by varying the index value between -10 and 10 for each data point to be predicted [20]. Table 2 shows how the GRG Nonlinear optimization approach was carried out using the Microsoft Excel Solver software.

Table 2. Extract Variable of GM (1,1) and NGBM (1,1) Models, N=10.

System	Forecasting Parameter	Coefficient	MAPE
GM (1,1)	$a = -0.0580$, $b = 229.464$	$\gamma = 0$	4.42%
NGBM (1,1)	$a = -0.0783$, $b = 315.420$	$\gamma = -0.2$	3.79%

However, Table 3 compares the NGBM (1,1) model's predicting values for CO_2 emissions in Saudi Arabia from 2002 to 2016 with the GM (1,1) model. That shows a high predictive consequence by modifying the index γ using the GRG Nonlinear approach. Table 3 also shows the MAPE values for CO_2 emissions in Saudi Arabia, demonstrating that using the NGBM (1,1) model, the optimum index value $\gamma = -0.2$ is 3.79 % at N=10, compared to the Traditional GM(1,1) mode, which is 4.42 % at N=10.

Accordingly, to evaluate the prediction accuracy, MAPE measures were employed after the NGBM (1,1) and GM (1,1) models of predicted carbon dioxide emissions. The findings show that the prediction accuracy of the NGBM (1,1) model is higher than the GM(1,1) model, as shown in Table 3. This is because it has the lowest rate of error. Though, the MAPE metrics in both models are less than 10%. The GM (1,1) has a MAPE of 4.42 %, while NGBM (1,1) has a MAPE of 3.79 %. This shows that as compared to the GM (1,1) model, the NGBM (1,1) model can improve prediction performance. As a result, the prediction value of NGBM(1,1) differs significantly from that of GM(1,1). This study, therefore, demonstrated that the Mean Absolute Percentage Error (MAPE) is around 3.79% in NGBM (1,1), which implies that the model is about 96.21% accurate in prediction. While GM(1,1) is approximately 95.58%.

Table 3. Forecast value and Percentage Error (PE).

Year	Actual value	GM (1,1)		NGBM (1,1)	
		Predicted Value	PE (%)	Predicted Value	PE (%)
2002	326.407	314.32	3.70%	299.21	8.33%
2003	327.272	333.11	1.78%	316.38	3.33%
2004	395.834	353.02	10.81%	336.01	15.11%
2005	397.642	374.13	5.91%	358.00	9.97%
2006	432.739	396.49	8.38%	382.35	11.64%
2007	387.777	420.196	8.36%	409.126	5.51%
2008	430.175	445.314	3.52%	438.439	1.92%
2009	468.965	471.934	0.63%	470.433	0.31%
2010	518.491	500.146	3.54%	505.275	2.55%
2011	499.878	530.043	6.03%	543.162	8.66%
MAPE (2002-2011)		4.42%			3.79%
2012	564.842	534.679	6.03%	502.516	12.47%
2013	541.047	555.664	2.59%	525.569	2.74%
2014	601.046	577.473	4.36%	552.736	8.93%
2015	647.111	600.137	7.82%	583.585	10.57%
2016	563.449	623.691	9.31%	617.945	8.42%
MAPE (2012-2016)			6.02%		8.63%

Besides, Figure 2 and Figure 3 show the actual value with the forecasted value of CO_2 emissions in Saudi Arabia for NGBM (1,1) and GM (1,1) from 2002 to 2011. It was observed from the graphs that there is an increase in the CO_2 emissions in the forecast values than the actual values.

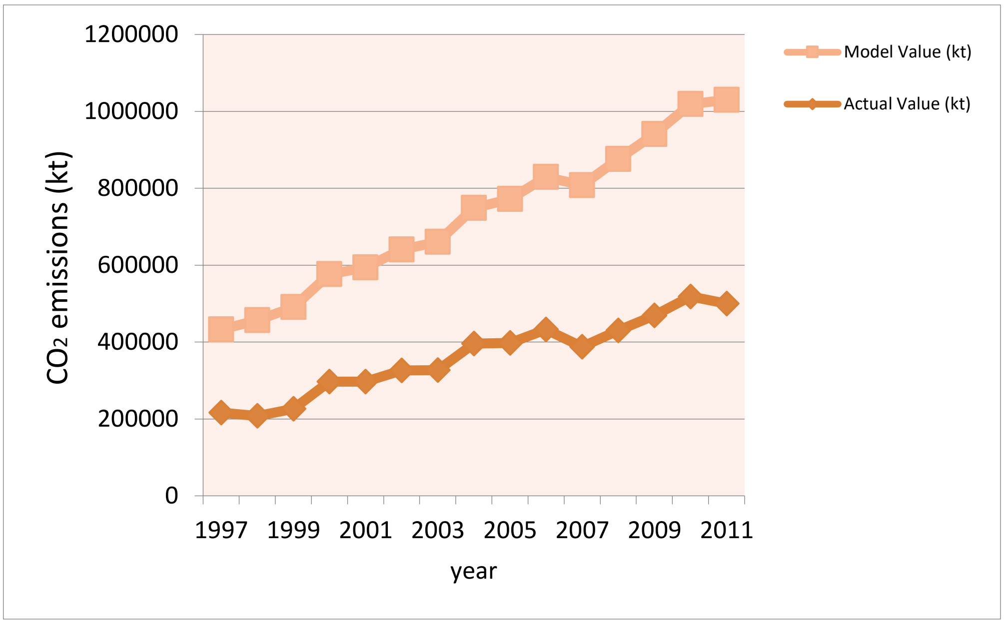

Figure 2. Actual and predicted CO_2 emissions in S.A for GM (1,1).

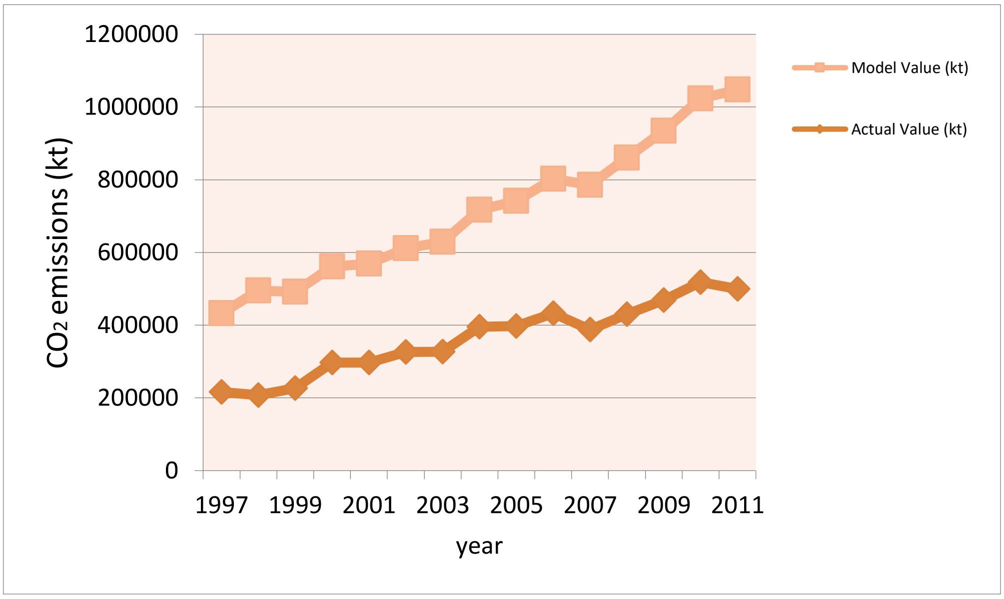

Figure 3. Actual and predicted CO_2 emissions in S.A for NGBM (1,1).

Accordingly, Figure 4 compares CO_2 emission predicting values in Saudi Arabia from 2002 to 2016 using the NGBM (1,1) model with the GM (1,1) model. As shown in Figure 4, there is a convergence between the predicted values of the two models.

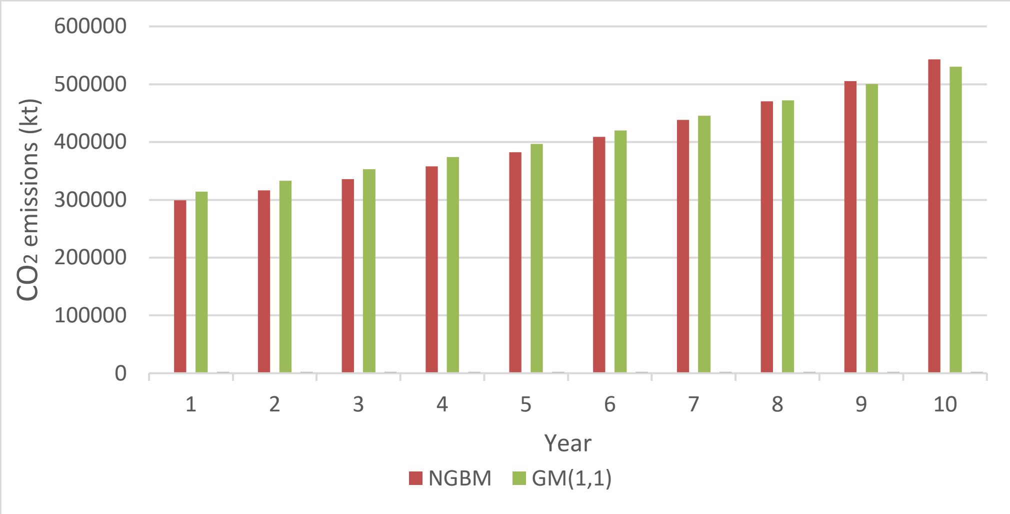

Figure 4. Predicted values of CO_2 emissions in S.A for two model.

5. Conclusion

This study aimed to predict CO_2 emissions in Saudi Arabia by employing the Nonlinear Grey Bernoulli model based on MAPE metrics models on the annual data from 2017 to 2021. This study concluded that NGBM (1,1) modeling is beneficial in predicting the future output of the system as it has a high level of accuracy. The prediction accuracy of the NGBM (1,1) model is estimated by Mean Absolute Percentage Error (MAPE). Generally, below 10% MAPE confirms that the NGBM (1,1) provides good prediction accuracy. Therefore, this study shows that NGBM (1,1) is more accurate than GM (1,1) by evaluating MAPE. The findings of this study are critical for the Saudi government, particularly in terms of medium- and long-term economic planning.

Acknowledgement

The authors would like to sincerely thank the University of Tabuk in Saudi Arabia for fully funding this research and Universiti Teknologi Malaysia (UTM) (University of Studying).

References

[1] Hossain A, Islam M A, Kamruzzaman M, Khalek M A and Ali M A 2017 *International conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment at Department of Statistics, University of Rajshahi* ***978-984-34-0996-6,*** *900-908*
[2] Bonga W G and Chirowa F 2014 *Social Science Research Network* ***2412639*** *1-26.*
[3] Ricke K, Drouet L, Caldeira K and Tavoni M 2018 *Nat. Clim. Chang.* **8** (10) 895–900.
[4] Jevrejeva S, Jackson L P, Grinsted A, Lincke D and Marzeion B 2018 *Environ. Res. Lett.* **13** (7) 740-14.
[5] Abdullah L and Pauzi H M 2015 *J. Teknol.* **75** (1) 67-82.
[6] Chen C I 2008 *Chaos, Solitons & Fractals* **37** (1) 278–287.
[7] Chen C I, Chen H L and Chen S P 2008 *Commun. Nonlinear Sci. Numer. Simul.* **13** (6) 1194–1204.
[8] Liu S F, Dang Y G and Fang Z G 2004 *The theory of grey system and its applications* (Science Press, Beijing).
[9] Zhou J, Fang R, Li Y, Zhang Y and Peng B 2009 *Appl. Math. Comput.* **207** (2) 292–299.
[10] Hsu L C 2009 *Expert Syst. Appl.*, 36 (4), doi: 10.1016/j.eswa.2008.11.004.
[11] Xie W, Wu,W Z, Liu C, Zhang T and Dong Z, 2021 *Environ. Sci. Pollut. Res.* **28** 38128-44.
[12] Wang Z X, Hipel K W, Wang Q and He S W 2011 *Appl. Math. Model.* **35** (12) 5524–5532.
[13] Xu N, Dang Y and Cui J 2015 *J. Syst. Eng. Electron.* **26** (4) 794–801.
[14] Wang Z X, Dang Y G, Liu S F and Lian Z 2009 *Syst. Eng. Electron.* **31** (10) 2380–2383.

[15] Wang Z X, Dang Y G and Zhao J J 2012 *Syst. Eng. Pract.* **32** (9) 1973–78.
[16] Pao H T, Fu H C and Tseng C L, 2012 *Energy* **40** (1) 400–409.
[17] Lewis, C. D, 1982, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. (Butterworth-Heinemann, Boston), 1-143.
[18] Agrawal R K and Adhikari R 2013, *An introductory study on time series modeling and forecasting* (*Nov. York CoRR*).
[19] Lasdon L S, Waren A D, Jain A and Ratner M 1978, *ACM Trans. Math. Softw.***4** (1) 34–50.
[20] Mustaffa A S and Shabri A 2020 *11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, 32–37.