Time Series Prediction of CO_2 Emissions in Saudi Arabia Using ARIMA, GM(1,1), and NGBM(1,1) Models

Z. F. Althobaiti
Department of Statistics, Faculty of Science, University of Tabuk,
Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia

A. Shabri
Department of Mathematical Sciences, Faculty of Science,
Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia

The investigation of economic aspects of gas emissions in terms of its volume and consequences is very important, given the current increasing trend. Therefore, the prediction of carbon dioxide emissions in Saudi Arabia becomes necessary. This study uses annual time series data on CO_2 emissions in Saudi Arabia from 1970 to 2016. The study built the prediction model of CO_2 emissions in Saudi Arabia by using Autoregressive Integrated Moving Average (ARIMA), Grey System GM and Nonlinear Grey Bernoulli Model (NGBM), and comparing their efficiency and accuracy based on MAPE metric. The results revealed that Nonlinear Grey Bernoulli Model (NGBM) is more accurate than the other prediction models. The results may be useful to Saudi Arabian government in the development of its future economic policies. As a result, five policy recommendations have been proposed, each of which could play a significant role in the development of acceptable Saudi Arabian climate policies.

Keywords: *annual time series data, Autoregressive Integrated Moving Average (ARIMA), CO_2 emissions, global warming, Grey Model (GM), Nonlinear Grey Bernoulli Model (NGBM), prediction, 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 CO_2 in the atmosphere (Hossain et al. 2017, Bonga & Chirowa 2014). CO_2 is a type of greenhouse gas (GHGs) emitted due to human activities.

Human activities are among the primary drivers of carbon dioxide 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 one of the wealthy oil and industrial nations disposed to carbon dioxide emissions, thus exacerbate global warming. Accordingly, the resulting economic losses from CO_2 emissions are more than those anticipated by the industries. This is in corroboration with the study of Ricke et al. (2018), 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. Nevertheless, the three largest countries that are much concerned of the climate change are the United States, Saudi Arabia, and China, which have been ranked in terms of carbon dioxide emissions.

Another study by Jevrejeva et al. (2018) also warned that failure to reduce greenhouse gas emissions would inevitably lead to sea-level rise, which would have severe economic consequences in the world. For instance, with temperatures reaching pre-industrial levels, floods from sea-level rise could cost society $14 trillion yearly by 2100. Therefore, the prediction of CO_2 emissions, which is the most significant task in time series analysis become necessary. Predictions are extremely essential in many fields such as sciences, economics, agriculture, meteorology, medicine, engineering, and others. The prediction of CO_2 emissions involves predicting the values of the time series from the observed time series. The prediction of CO_2 emissions have become a global concern, as it has shown to assist in raising public knowledge about how to forestall environmental issues (Abdullah & Pauzi 2015). Therefore, to make a realistic estimate of Saudi Arabia's future CO_2 emissions, a fuller understanding of the most suitable prediction models is essential.

Many predictive models, such as ARIMA and gray models have been used by researchers to predict CO_2 emissions. For instance, Nyoni & Bonga, (2019) studied forecasting of CO_2 emissions in India. In the study, ARIMA(2,2,0) model was determined to have the best fit for projecting yearly CO_2 in India for the next 13 years, with an estimate of 3.89 million kt by 2025. Also, Chigora et al. (2019) carried out a research on univariate approach using Box-Jenkins to forecast CO_2 emissions for Zimbabwe's tourism destination vibrancy. The ARIMA(10,1,0) model, which focuses on the amount of carbon dioxide (CO_2) emission in Zimbabwe from 1964 to 2014, was employed to have the most suitable model for forecasting yearly CO_2 emissions for the next 10 years, with the model indicating that it will be around 15,000 kt by 2024. Similarly, Nyoni & Mutongi, (2019) predicted carbon dioxide emissions in China from 1960 to 2017, using autoregressive integrated moving average (ARIMA) models. With a prediction of 10,000,000 kt by 2024, the ARIMA (1, 2, 1) model proved to be the most suitable model for forecasting yearly total CO_2 emissions in China for the next ten years.

Lotfalipour et al. (2013) using the Grey and ARIMA models, estimated that CO_2 emissions in Iran will reach 925.68 million tons in 2020, up to 66% from 2010. Also, employing a differential model to predict CO_2 emissions in Iran, the author used the grey system and Autoregressive Integrated Moving Average, and compared them with the RMSE, MAE, and MAPE metrics models. Based on the findings, the ideal degree of Hannan – Rissanen and Box – Jenkins for ARIMA, the ARIMA(1, 1, 2) model was developed. Even though MAPE metrics for three models were less than 10% accuracy of prediction, the grey system confirmed that the three models demonstrated predicting accuracy. Thus, based on the GM (1, 1) estimates, CO_2 emissions was revealed to reach 925.68 million tons in 2020, representing a 66 percent increase over 2010. Besides, Ho, (2018) has also investigated the grey model.

Chen, (2008) and Chen et al. (2008) 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 Liu et al. (2004) termed 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 Zhou et al. (2009) 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. The genetic algorithm was used in (Hsu 2009) to optimize the parameters of the NGBM(1, 1), which was then employed to predict economic developments in Taiwan's integrated circuit industry. Moreover, studies by Xie et al. (2021) projected fuel combustion-related CO_2 emissions 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 was also 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.

The findings of these studies imply that ARIMA, GM (1, 1), and NGBM(1, 1) models has continued to prove to be the most suitable model for predicting yearly CO_2 emissions and could form the underlying basis for predicting CO_2 emissions in Saudi Arabia. In this regard, this study intends to evaluate the accuracy of the predicting models in order to obtain the most precise data prediction.

2. Research Methodology

Three predicting models: ARIMA model, grey model and NGBM(1,1) are used in this study. The reasons why these three models were chosen is firstly due to the ARIMA model, which is a conventional forecasting model that produces more reliable and accurate forecasts. Also, it has the benefit of being able to employ a combination of auto regression, difference, and moving average of different orders to generate the ARIMA(p, d, q) model, which can convey multiple types of information of time series. Secondly, GM(1,1) does not necessitate a large sample size, and the effect of short-term prediction is good. Thirdly, ARIMA model and grey model can be directly compared on the same base. The NGBM(1, 1) is a newly created grey model with wide range of applications in diverse fields. This is due to its precision in handling small time-series datasets with nonlinear variations.

2.1. Autoregressive Integrated Moving Average (ARIMA)

The prediction using ARIMA models statistical method is usually viewed as providing more accurate predictions than econometric methodologies (Song et al. 2003). Also, in terms of forecasting performance, ARIMA models outperformed the multivariate models (Du Preez & Witt 2003). Moreover, ARIMA models outperform naive models and smoothing approaches in terms of overall performance (Goh & Law 2002). ARIMA models were created in the 1970s by Box and Jenkins, and its identification, estimation, and diagnostics method is based on the notion of parsimony (Asteriou & Hall 2015). That is; when the original time series is not stationary, the first order difference process ΔY or second order differences $\Delta^2 Y$, and so on, can be investigated. While, If the differenced process is a stationary process, ARIMA model of that differenced process can be found in practice if differencing is applied, usually $d = 1$, or maybe $d = 2$, is enough. The general form of the ARIMA(p, d, q) can be represented by a backward shift operator as.

$$\phi(B)\,\Delta^d Y_t = \theta(B)\,\varepsilon_t$$

The general autoregressive moving average process with AR order p and MA order q can be written as

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ (the p order AR operator)

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ (the p order AR operator)

$\Delta^d = (1 - B)^d$

These processes can be written briefly as: $Y_t \sim$ ARIMA(p, d, q) where ϕ is the autoregressive component's parameter estimate, θ is the moving average component's parameter estimate, Δ is the difference operator, d is the difference, and B is the backshift operator (Box et al. 2015).

2.2. ARIMA model

The ARIMA model is one of the most widely used statistical models for time series forecasts (Box et al. 2015). Its forecast principle is to transfer a non-stationary time series into a stationary time series first. As a result, the dependent variable will be described as a model that only yields its lag value, as well as the actual and lag values of the random error term. The following are the steps in the prediction phase (Wang et al. 2018):

Phase 1: Smooth the time data with a differential tool. In theory, stationarity ensures that the fitted curve formed by sampling time series can continue inertially along the present form in the future, i.e., the data's mean and variance should not be significantly changed.

Phase 2: Create a model that is autoregressive (AR). The autoregressive model is a way of forecasting itself using the variable's historical result data, and it describes the link between current value and previous value. It has the following formula:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \quad (1)$$

where y_t represents the current value, μ indicates the constant term, p denotes the order, ϕ_i is the autocorrelation coefficient, and ε_t represents the error.

Phase 3: Create a model based on moving averages (MA). In the autoregressive model, the moving average model concentrates on the accumulation of error components. Random fluctuations in forecasts can be successfully eliminated. It has the following formula:

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2)$$

where θ_i is the MA formula's correlation coefficient.

Phase 4: Create an autoregressive moving average model by combining AR and MA(ARMA). The following is the exact formula. The orders of the autoregressive and moving average models, respectively, are p and q in this formula. The correlation coefficients of the two models, ϕ_i and θ_i, respectively, must be solved.

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{3}$$

2.3. The Box – Jenkins Methodology

The subjective evaluation of plots of auto-correlograms and partial auto-correlograms of the series is used to identify models in the Box-Jenkins process (Meyler et al. 1998). The initial step in model selection is to vary the series to attain stationarity. The researcher will then assess the correlogram to identify the right sequence of the AR and MA components. Because there are no clear–cut guidelines for determining whether AR and MA components are appropriate. Though, this method of selecting AR and MA components is skewed toward the use of personal judgement. As a result, prior experience is essential. The next step is to estimate the preliminary model, which is followed by diagnostic testing. This is accomplished by creating residuals and analyzing whether they fulfil the parameters of a white noise process, which is common in diagnostic testing. If this is not the case, the model must be re-specified, and the method must be restarted from the second stage. The process may continue indefinitely until a suitable model is produced (Nyoni 2018). This procedure is clearly illustrated in Figure 1.

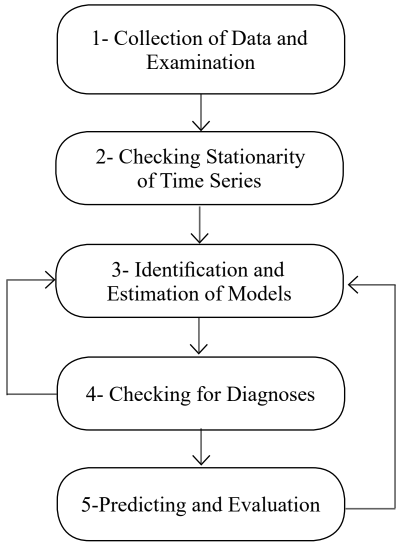

Figure 1. Procedure for ARIMA Forecasting

2.4. Grey Model, GM(1,1)

GM(1,1) denotes a grey forecasting model with one variable and one order. The following is the general steps for creating a grey forecasting model:

Step 1: Create an initial sequence based on observed data.

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(\mathrm{n})) \quad (4)$$

where $x^{(0)}(i)$ denotes the baseline data (state = 0) for the time i

The sample size is n, and the non-negative sequence is $x^{(0)}$. Four data points can be used to develop and build the GM (1, 1) model.

Step 2: Using the initial sequence $x^{(0)}$, to generate the first-order Accumulated Generating Operation (AGO) sequence $x^{(1)}$

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots\ldots, x^{(1)}(n)\right) \quad , \quad \mathrm{n} \geq 4 \quad (5)$$

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

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \quad (6)$$

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

The average sequences generator's definition is as follows:

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

The average value of the sequential data $z^{(1)}(k)$ is define as follows;

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \quad k = 2, 3, \ldots, n \quad (7)$$

Step 4: Assume the first-order differential equation for the sequence $x^{(1)}$ is as follows:

$$\frac{dx^{(1)}(k)}{dk} + a\, x^{(1)}(k) = b$$

Then its difference equation is shown as:

$$x^{(0)}(k) + a\, z^{(1)}(k) = b \quad (8)$$

where a and b are the estimated parameters of the grey forecasting model.

Step 5: The parameters a and b are calculated using the least-squares method (OLS).

$$\hat{a} = [a,b]^T = \left(B^T B\right)^{-1} B^T Y \tag{9}$$

$$Y = [x^{(0)}(2), x^{(0)}(3),..., x^{(0)}(n)]^T$$

$$B = \begin{bmatrix} -\frac{1}{2}\left(x^{(1)}(1)+x^{(1)}(2)\right) & 1 \\ -\frac{1}{2}\left(x^{(1)}(2)+x^{(1)}(3)\right) & 1 \\ . & \\ . & \\ -\frac{1}{2}\left(x^{(1)}(n-1)+x^{(1)}(n)\right) & 1 \end{bmatrix}$$

Step 6: Under the initial condition $x^{(1)}(1) = x^{(0)}(1)$, the solution of the grey differential equation produces:

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a} \tag{10}$$

Step 7: The first-order inverse accumulated generating operation can be used to get the forecast values $\hat{x}^{(0)}(k+1)$ (IAGO).

$$\hat{x}^{(0)}(k+1) = x^{(1)}(k+1) - x^{(1)}(k) \tag{11}$$

2.5. 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 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 Wang et al. (2011) and Xu et al. (2015). Also, Xie et al. (2021) 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.

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

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 Acumulated Generating Operation (AGO) sequence $x^{(1)}$.

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

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

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

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

The following is the definition of the average sequences generator:

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

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

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

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

$$x^{(0)}(k) + a\, z^{(1)}(k) = b\, (z^{(1)}(k))^{\gamma}, \gamma \neq 1 \tag{12}$$

which is the whitening equation of the NGBM(1, 1) model.

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

$$\frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b(x^{(1)})^{\gamma} \tag{13}$$

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:

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

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

$$Y = \left[x^{(0)}(2), x^{(0)}(3), ..., x^{(0)}(n) \right]^T$$

$$B = \begin{bmatrix} -z^{(1)}(2) & \left(z^{(1)}(2)\right)^{\gamma} \\ -z^{(1)}(3) & \left(z^{(1)}(3)\right)^{\gamma} \\ \cdot & \\ \cdot & \\ -z^{(1)}(n) & \left(z^{(1)}(n)\right)^{\gamma} \end{bmatrix} \tag{14}$$

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

$$\hat{x}^{(1)}(k+1) = \left\{ \frac{b}{a} + \left[\left(x^{(0)}(1) \right)^{1-\gamma} - \frac{b}{a} \right] e^{-(1-\gamma)ak} \right\}^{\frac{1}{1-\gamma}} \tag{15}$$

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

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), \quad k = 2, 3, \ldots, m. \tag{16}$$

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 $\gamma = 0$. The NGBM model is the grey Verhulst model (GMV) when $\gamma = 2$. 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. Wang et al. (2009) 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) (Wang et al. 2012).

2.5.1. Parameter Optimization of the Traditional NGBM(1,1)

The traditional NGBM(1,1) help to determine the expected values for the optimization problem. However, Pao et al. (2012) proposed a relatively simple iterative method for determining the optimal γ.

$$\min_{\gamma} MAPE = \frac{1}{n-1}\sum_{k=2}^{n}\left|\frac{\hat{x}^{(0)}(k)-x^{(0)}(k)}{x^{(0)}(k)}\right| x100\% \tag{17}$$

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:

$$MAPE = \frac{1}{n}\left(\left|\frac{\sum_{i=1}^{n} x_i - \hat{x}_i}{x_i}\right|\right) x100\% \tag{18}$$

where MAPE refers to Mean Absolute Percentage Error, $\hat{x}_i$ is the predicted value, x_i, is the actual value, and the number of data observations n as shown in Table 1.

Table 1. The MAPE Criteria of Prediction Precision

MAPE (%)	≤10	10-20	20-50	≥50
prediction precision	Highly accurate	Good	Reasonable	inaccurate

Source: (Lewis, 1982)

Hence, for a good forecast, the obtained MAPE should be as small as possible (Agrawal & Adhikari, 2013)

4. Results and Discussions

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 ARIMA, Nonlinear Grey Bernoulli Model (NGBM) and Grey Model (GM) to predict CO_2 emissions. Figure 2 shows that CO_2 emissions (Y) has been increasing from 1970 to 2016, indicating that the trend is not stationary. This implies that the mean and variance of the data are changing over time. Accordingly, the data was divided into two parts: training and testing (forecasting). The data from 2002-2011 was used for training, while the data from 2012 -2016 was used for testing.

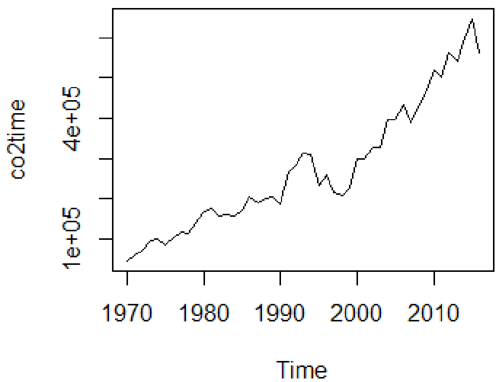

Figure 2. Time series of CO2 emissions in Saudi Arabia

4.1. ARIMA Model

To examine the stationarity of CO_2 emissions, Augmented Dickey- Fuller1 test (1981) was used. According to Table 2, the results of the (ADF) test of the time Series are not stationary in the level at which the calculated statistical significance levels are greater than the level of 0.05. The test results indicated that the time series has reached the stage of stationary after making its first difference. As indicated, the test's statistical significance is less than the 0.05 level.

Table 2. Augmented Dickey- Fuller test (ADF)

Result	Critical value of ADF	The test statistic	
Non- stationary	-1.7963	0.6552	CO_2
stationary	-3.9973	0.01814	d CO_2

ARIMA(1, 1, 0) with lower AIC is preferable than the one with a higher AIC values (Nyoni 2018). As a result, the ARIMA (1, 1, 0) model is selected as the best as shown in Table 3.

Table 3. Comparison of the Variants of the ARIMA Models

Box-Jenkins Model ARIMA(*p*,*d*,*q*)	AIC
ARIMA(2,1,2)	1097.67
ARIMA(0,1,0)	1094.491
ARIMA(1,1,0)	1093.599
ARIMA(0,1,1)	1094.78
ARIMA(2,1,0)	1095.573
ARIMA(1,1,1)	1095.563
ARIMA(2,1,1)	1098.859

In Table 4, the AR (1) component coefficients are negative and statistically significant at the 5%. This implies that historical CO_2 levels are relevant in describing current and future CO_2 levels in Saudi Arabia. Figure 3 shows that CO_2 emissions in Saudi Arabia are increasing throughout a 13-year period, from 2017 to 2030. Saudi Arabia's CO_2 emissions will reach 747241.6 kt by 2030. As a result, Saudi Arabia will continue to face issues related to global warming and climate change.

Table 4. Results of z Test Coefficients for ARIMA (1,1,0)

variable	coefficient	Standard Error	Z	p-value
AR(1)	-0.2678	0.1547	-1.7312	0.083422
Intercept(mean)	11689.3375	3851.7087	3.0348	0.002407 **

*The *, ** and *** means significant at 10%, 5% and 1% respectively.*

4.2. GM(1,1) and NGBM(1,1) models

The GM(1,1) and NGBM(1,1) models were employed to predict CO_2 emissions in Saudi Arabia. Equation (1) to Equation (6) are used to determine the parameters, develop coefficient *a,* and grey variable *b* for ordinary least squares calculation, and the output is actual GM (1, 1) only variable *a* and *b,* which must be simulated with $\gamma = 0$. The other is determined using the three unknown NGBM(1, 1) variables *a*, *b*, and γ, as given in Table 4. The GRG Nonlinear method of optimization, first devised by Leon Lasdon and Alan Waren, is used to determine the value of the index (Power Exponent γ) (Lasdon et al. 1978). Its implementation as a Fortran software for addressing small to medium-sized issues and some computational findings solved the Nonlinear Optimization Problem. As a result, 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 (Pao et al. 2012), and by varying the value of index between -10 and 10 for each data point to be forecasted (Mustaffa & Shabri 2020).

5. Comparative Study

Table 5. Predicted value and MAPE

Year	Actual value	GM(1,1), $\gamma = 0$ $a = -0.0580$, $b = 229.464$		NGBM(1,1), $\gamma = 0.2$ $a = -0.0783$, $b = 315.420$		ARIMA (1,1.0)	
		Predicted VALUE	PE(%)	Predicted VALUE	PE(%)	Predicted VALUE	PE(%)
2002	326.407	314.32	3.70%	299.21	8.33%	305.34	6.45%
2003	327.272	333.11	1.78%	316.38	3.33%	313.46	4.22%
2004	395.834	353.02	10.81%	336.01	15.11%	321.59	18.76%
2005	397.642	374.13	5.91%	358.00	9.97%	329.72	17.08%

2006	432.739	396.49	8.38%	382.35	11.64%	337.84	21.93%
2007	387.777	420.196	8.36%	409.126	5.51%	345.97	10.78%
2008	430.175	445.314	3.52%	438.439	1.92%	354.09	17.69%
2009	468.965	471.934	0.63%	470.433	0.31%	362.22	22.76%
2010	518.491	500.146	3.54%	505.275	2.55%	370.35	28.57%
2011	499.878	530.043	6.03%	543.162	8.66%	378.47	24.29%
MAPE(2000-2011)			4.42%		3.79%	20.82%	
2012	564.842	534.679	5.34%	502.516	11.03%	368.6	34.74%
2013	541.047	555.664	2.70%	525.569	2.86%	394.73	27.04%
2014	601.046	577.473	3.92%	552.736	8.04%	402.85	32.98%
2015	647.111	600.137	7.26%	583.585	9.82%	410.98	36.49%
2016	563.449	623.691	10.69%	617.945	9.67%		
MAPE (2012-2016)			5.98%	8.28%		31.37%	

Source: Researcher's fieldwork

Table 5 demonstrated that the MAPE value for the NGBM(1,1) in modeling is 3.79%. In comparison, the MAPE value for simulation and forecast data is 8.63%, as shown in Table 5. This implies that the smaller data size influences the MAPE value for simulation data, and its value increases. It is known that the lower the MAPE value, the more accurate the model, and therefore the precise model is at N = 10 for NGBM(1,1).

According to the results, the GM(1,1) has a MAPE of 4.42 %, ARIMA has a MAPE of 20.82%, while NGBM(1,1) has a MAPE of 3.79 %. Compared to the GM(1,1) model and ARIMA 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) and ARIMA. 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% the highly accurate in prediction based on the MAPE criteria of prediction precision. While, GM(1,1) is around 4.42% approximately 95.58% highly accurate. But ARIMA(1,1,0) model is around 20.82%, about 79.18% reasonably accurate as presented in the MAPE criteria in Table 1. Consequently, Figure 3 shows the comparison of predictive data of these three models. The NGBM(1,1) model has outperformed than ARIMA(1,1,0) and GM(1,1) model. This is as a result that NGBM(1,1) model has the lower value of MAPE (3.79%) compared with GM(1,1) model (4.42 %) and ARIMA(1,1,0) model (20.82%) . Therefore, NGBM(1,1) delivers the best result among those considered and was used to predict CO_2 emissions in Saudi Arabia. It was also observed that CO_2 emission in Saudi Arabia is continuously increasing as shown in Figure 3. This implies that CO_2 emissions in Saudi Arabia will continue to rise over the next decade from 2017 to 2026, as presented in Figure 4, and the country will face the challenges of global warming, climate change, as well as clean and healthy environment.

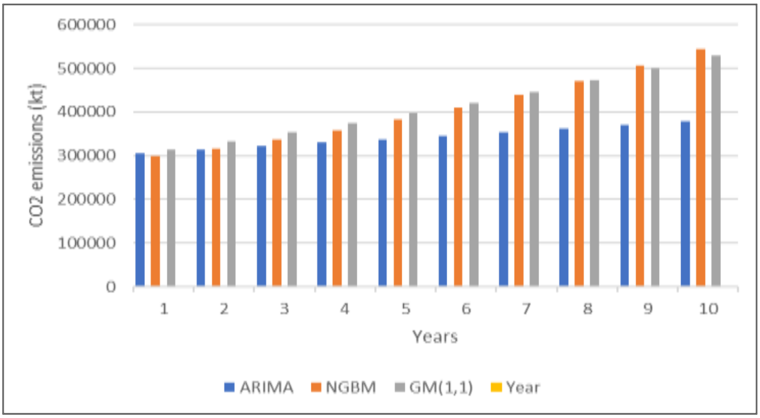

Figure 3. Comparison of predictive data, ARIMA(1,1,0),GAGM(1,1) and GM(1,1) in Saudi Arabia from 2002 to 2011

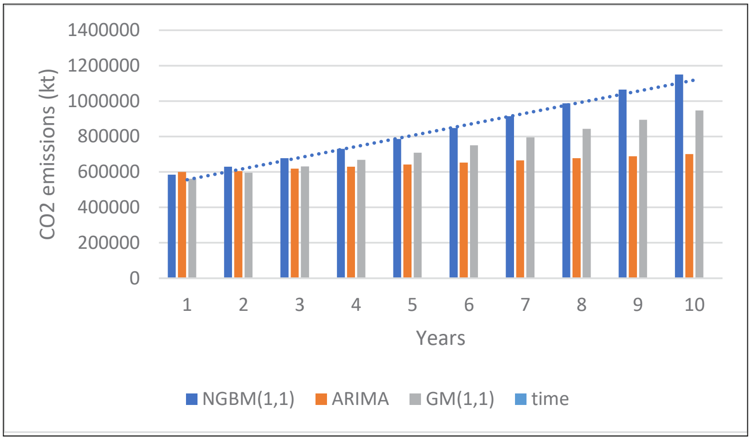

Figure 4. Comparison of predictive data, ARIMA(1,1,0),GAGM(1,1) and GM(1,1) in Saudi Arabia over the next decade from 2017 to 2026

6. Conclusion

This study concluded that NGBM(1,1) modelling is suitable 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 ARIMA(1,1,0) and 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.

To build on these findings and forecast the performance of other sectors, more investigation is recommended. Because this analysis exclusively forecasted CO_2 emissions in Saudi Arabia, this was proposed. CO_2 emissions are influenced by several causes, including the combustion of fossil fuels and the loss of vegetative cover. As a result, humans and ecosystems are affected, and future study will be able to use multi-factor Grey prediction models to develop more precise CO_2 emission projections.

Recommendations

Based on the findings, the following recommendations were made for Saudi Arabia to reach its goal of lowering carbon emissions:

1. Development of renewable energy sources. Although, Saudi Arabia has strong capabilities in solar and winds energy. It does not currently have a competitive sector in the area of renewable energy, so it must be developed.
2. The transition from coal to natural gas.
3. Reliance on nuclear technology to produce energy, which is used in nuclear power plants.
4. There is also a need to keep educating the Saudi people about the need of decreasing pollution levels.
5. The Saudi government should limit pollution by enacting policies such as raising taxes on polluting companies, particularly those that produce fossil fuels, in their everyday operations.

Acknowledgment

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

References

Abdullah, L., & Pauzi, H. M. 2015. "Methods in Forecasting Carbon Dioxide Emissions: A Decade Review." *Jurnal Teknologi* 75(1).

Agrawal, R. K., & Adhikari, R. 2013. *An Introductory Study on Time Series Modeling and Forecasting*. Nova York: CoRR.

Asteriou, D., & Hall, S. G. 2015. *Applied Econometrics*. Macmillan International Higher Education.

Bonga, W. G., & Chirowa, F. 2014. *Level of Cooperativeness of Individuals to Issues of Energy Conservation*. Available at SSRN 2412639.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Chen, C.-I. 2008. "Application of the Novel Nonlinear Grey Bernoulli Model for Forecasting Unemployment Rate. *Chaos, Solitons & Fractals* 37(1): 278–287.

Chen, C.-I., Chen, H. L., & Chen, S.-P. 2008. "Forecasting of Foreign Exchange Rates of Taiwan's Major Trading Partners by Novel Nonlinear Grey Bernoulli Model NGBM(1, 1)." *Communications in Nonlinear Science and Numerical Simulation* 13(6): 1194–1204.

Chigora, F., Thabani, N., & Mutambara, E. 2019. *Forecasting?2 Emission for Zimbabwe's Tourism Destination Vibrancy: A Univariate Approach using Box-Jenkins ARIMA Model*.

Du Preez, J., & Witt, S. F. 2003. "Univariate Versus Multivariate Time Series Forecasting: An Application to International Tourism Demand." *International Journal of Forecasting* 19(3): 435–451.

Goh, C., & Law, R. 2002. "Modeling and Forecasting Tourism Demand for Arrivals with Stochastic Nonstationary Seasonality and Intervention." *Tourism Management* 23(5): 499–510.

Hossain, A., Islam, M. A., Kamruzzaman, M., Khalek, M. A., & Ali, M. A. 2017. *Forecasting Carbon Dioxide Emissions in Bangladesh using Box-Jenkins ARIMA Models*. Department of Statistics, University of Rajshahi.

Hsu, L. C. 2009. "Forecasting the Output of Integrated Circuit Industry Using Genetic Algorithm Based Multivariable Grey Optimization Models." *Expert Systems with Applications* 36(4). https://doi.org/10.1016/j.eswa.2008.11.004

Jevrejeva, S., Jackson, L. P., Grinsted, A., Lincke, D., & Marzeion, B. 2018. "Flood Damage Costs Under the Sea Level Rise With Warming of 1.5 C and 2 C." *Environmental Research Letters* 13(7): 74014.

Lasdon, L. S., Waren, A. D., Jain, A., & Ratner, M. 1978. "Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming." *ACM Transactions on Mathematical Software (TOMS)* 4(1): 34–50.

Lewis, C. D. 1982. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Butterworth-Heinemann.

Liu, S. F., Dang, Y. G., & Fang, Z. G. 2004. *The Theory of Grey System and Its Applications*. Science Press, Beijing.

Lotfalipour, M. R., Falahi, M. A., & Bastam, M. 2013. "Prediction of CO_2 Emissions in Iran Using Grey and ARIMA Models." *International Journal of Energy Economics and Policy* 3(3): 229.

Meyler, A., Kenny, G., & Quinn, T. 1998. *Forecasting Irish inflation using ARIMA models*.

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

Nyoni, T. 2018. *Modeling and Forecasting Naira/USD Exchange Rate in Nigeria: A Box-Jenkins ARIMA approach.*Nyoni, Thabani.

Nyoni, T., & Bonga, W. G. 2019. "Prediction of CO_2 Emissions in India Using Arima Models." *DRJ-Journal of Economics & Finance* 4(2): 1–10.

Nyoni, T., & Mutongi, C. 2019. *Modeling and Forecasting Carbon Dioxide Emissions in China Using Autoregressive Integrated Moving Average (ARIMA) Models*.

Pao, H.-T., Fu, H.-C., & Tseng, C.-L. 2012. "Forecasting of CO_2 Emissions, Energy Consumption and Economic Growth in China Using an Improved Grey Model. *Energy* 40(1): 400–409.

Ricke, K., Drouet, L., Caldeira, K., & Tavoni, M. 2018. Country-level Social Cost of Carbon. *Nature Climate Change* 8(10): 895–900.

Song, H., Witt, S. F., & Jensen, T. C. 2003. Tourism Forecasting: Aaccuracy of Alternative Econometric Models. *International Journal of Forecasting* 19(1): 123–141.

Wang, Q., Song, X., & Li, R. 2018. A Novel Hybridization of Nonlinear Grey Model and Linear ARIMA Residual Correction for Forecasting US Shale Oil Production. *Energy* 165: 1320–1331.

Wang, Z.-X., Hipel, K. W., Wang, Q., & He, S.-W. 2011. "An Optimized NGBM(1, 1) Model for Forecasting the Qualified Discharge Rate of Industrial Wastewater in China." *Applied Mathematical Modelling* 35(12): 5524–5532.

Wang, Z. X., Dang, Y. G., Liu, S. F., & Lian, Z. 2009. "Solution of GM (1, 1) Power Model and Its Properties. *Systems Engineering and Electronics* 31(10): 2380–2383.

Wang, Z. X., Dang, Y. G., & Zhao, J. J. 2012. "Optimized GM (1, 1) Power Model and Its Application." *Systems Engineering-Theory & Practice* 32(9): 1973–1978.

Xie, W., Wu, W.-Z., Liu, C., Zhang, T., & Dong, Z. 2021. "Forecasting Fuel Combustion-related CO_2 Emissions by a Novel Continuous Fractional Nonlinear Grey Bernoulli Model with Grey Wolf Optimizer." *Environmental Science and Pollution Research,* 1–17.

Xu, N., Dang, Y., & Cui, J. 2015. "Comprehensive Optimized GM (1, 1) Model and Application for Short term forecasting of Chinese energy consumption and production. *Journal of Systems Engineering and Electronics* 26(4): 794–801.

Zhou, J., Fang, R., Li, Y., Zhang, Y., & Peng, B. 2009. "Parameter Optimization of Nonlinear Grey Bernoulli Model using Particle Swarm Optimization." *Applied Mathematics and Computation* 207(2): 292–299.