Carbon Dioxide Concentration Prediction Study Based on Grey Prediction Model and ARIMA Model

Zhanxinyu Zheng *, Xingkang Chen

Wuhan Britain-China School, Wuhan, China

* Corresponding Author Email: avaazzxy0409@gmail.com

Abstract. Concentrations of carbon dioxide began to soar after the Industrial Revolution and are now a serious threat to the global climate, and predicting future carbon dioxide levels is essential to balance economic prosperity and sustainable development. In this paper, the gray GM (1, 1) model and the ARIMA model are established to predict future carbon dioxide concentrations. The gray GM (1, 1) model predicts that carbon dioxide concentration levels will reach 685 ppm by 2132, and ARIMA models predict that carbon dioxide concentration levels will reach 685 ppm by 2103. The goodness-of-fit of the gray GM (1, 1) model and the ARIMA model were 0.99989 and 0.99998, respectively, and the ARIMA model was more accurate than the prediction results of the grey prediction model.

Keywords: Carbon Dioxide Concentration Prediction, ARIMA Model, Grey Prediction Model.

1. Introduction

Carbon dioxide is a significant heat-trapping gas that is produced by the extraction and combustion of fossil fuel, wildfires, and natural processes such as volcanic eruptions. According to the data provided by the National Aeronauticsn and Space Administration, human activities have increased atmospheric by approximately 50% since the beginning of industrial times (in the 18th century). Since modern recordkeeping began in 1995, the internal heat of the ocean has increased due to the ocean's 90 percent contribution to global warming. The surface holds the majority of the extra -energy, at a depth of 0 to 700 meters. As the excess heat and energy warm the ocean, the temperature change causes unprecedented casading effects such as ice melting, sea-level rise [1], marine heat waves and ocean acidification. The changes cause long-term effects on marine biodiversity, as well as the lives and livelihoods of coastal communities and beyond-including approximately 680 million people who live in low-lying coastal areas, nearly 2 billion who live in half og the world's coastal megacities, nearly half of the world's population (3.3 billion) who rely on fish for protein [2], and nearly 60 million people who work in fisheries and the aquaculture sector globally Predicting the concentration of carbon dioxide and its ensuing effect on ocean temperature relatively accurately is critical in balancing control of future economic and industrial development with environmental protection [3-4]. Evaluation of the impact of the year 2004 and prediction of the future trend of atmospheric concentration is also warranted. In addition, as the focus shifts to the temperature, we are requirednto describe the relationship between concentration levels and temperature over time. Finally, the overall practicability and sensitivity of the models should be analyzed in detail. To solve those problems, we will proceed as follows: using the most suitable model in step one to further predict future land-ocean temperatures changes; using grey relational analysis to analyze the relationship between concentrations and land-ocean temperatures since 1959 [5-6].

2. Model Theory, Implementation and Result of Grey GM (1, 1) Model

First of all, in order to confirm the feasibility of the data using the grey GM (1, 1) model, we should calculate, k=2, 3,..., if the value is in the interval, the grey GM (1, 1) model can be used.

Secondly, this paper let the terms in the original sequence for the 63 years 1959-2021 be respectively. Then this paper make an accumulation to generate a new sequence, where there are

$$x_k^{(1)} = \sum_{i=1}^{k} x_i^{(0)}, k = 1,2,3,\ldots,63 \tag{1}$$

Afterward, the mean series is generated

$$z_k^{(1)} = ax_k^{(1)} + (1-a)x_{k-1}^{(1)} \tag{2}$$

Where $0 \leq \alpha \leq 1$ is the weight. Usually $\alpha = 0.5$. Since the new sequence generated looks like a straight line, this paper can approximate this new sequence with an expression for a straight line.

Therefore, this paper construct a first-order ordinary differential equation to solve for the functional expression of the fitted curve. From this, the grey differential equation is established as

$$x_k^{(0)} + bz_k^{(1)} = c \tag{3}$$

The correspoding GM (1, 1) whitening differential equation is

$$\frac{\mathbf{dx^{(1)}}}{\mathbf{dt}} + \mathbf{bx_t}^{(\mathbf{1})} = \mathbf{c} \tag{4}$$

Shifting the term of the grey differential equation gives

$$-bz_k^{(1)} + c = x_k^{(0)} \tag{5}$$

a, b are parameters to be determined, the above equation can be written in form of a matrix in the form of $X\beta = Y$.

$$\begin{bmatrix} -z_2^{(1)} & 1 \\ -z_3^{(1)} & 1 \\ \cdots & 1 \\ -z_{63}^{(1)} & 1 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} x_2^{(0)} \\ x_{63}^{(0)} \\ \cdots \\ x_{63}^{(0)} \end{bmatrix} \tag{6}$$

Using the least-squares method, we can determined the parameter matrix β.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{7}$$

The resulting estimates of the parameters b, c are brought into the whitening equation to obtain the sequence $x_k^{(1)}$

$$\widehat{X_k^{(1)}} = (X_1^{(0)} - \frac{c}{b})e^{-\propto(k-1)} + \frac{c}{b} \tag{8}$$

Reducing to the original series yields the prediction function

$$\widehat{X_k^{(0)}} = (x_1^{(0)} - \frac{c}{b})e^{-\propto(k-1)}(1 - e^b), k = 1,2,3,\ldots,63 \tag{9}$$

When k=1, 2, 3...63, we can get the fitted value; when k > 63, we can get the predicted value.

This paper realize this model by using MATLAB. According to the predicted results, the CO2 concentration doubles from 1960 to 2100, rising from 300 to 600, with an overall slope increasing trend, indicating that the concentration is rising faster and faster with increasing time.

This paper can see that from the graph that the carbon dioxide concentration shows a relatively gradual upward trend which is less steep than the data predicted by polynomial fitting. Grey GM (1, 1) model predicts that in the year 2100, the carbon dioxide concentration is 592.04548 ppm in the atmosphere. This paper then predict the level of carbon dioxide concentration to the year 2150. The result shows that in the year 2132, carbon dioxide concentration is 685.567895553308 ppm, first exceeding 685 ppm.

2.1. Autoregressive Integrated Moving Average Model

1) The Principle of the ARIMA Model

The mathematical general formula of the ARIMA model is

$$\phi(L)\nabla^d Y_t = \theta(L)\epsilon_t \tag{10}$$

In equation (10), the Yt denotes the time series at moment t (corresponding to the ppm data for each year); L is the backshift operator, and d is the order. The auto-regressive coefficient can be written as

$$\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p \tag{11}$$

In equation (11), the parameter ϕ_1, ..., ϕ_p is the autocorrelation regression coefficient, which is the parameter to be estimated in the model. d is the difference order.et is a white noise. In Eq. (10), the θ (L) is the moving average operator, and the expression is

$$\theta(L) = 1 + \theta_1 L + \ldots + \theta_p L^q \tag{12}$$

In equation (12), the e1, . . . , ep is the moving average coefficient, which is the parameter to be estimated in the model; both p and q are the orders of the model.

3. Modeling and Analysis Process

Analytical processing of data modeling with ARIMA models generally requires the following steps: stability testing, parameter estimation, hypothesis testing, and model prediction.

(i) Stability check

From the fitted curves of the sample time series of carbon dioxide concentrations and years, it is clear that carbon dioxide concentration cannot continue "inertially" in its current state for some time in the future. Therefore, the series needs to be differenced by 79 steps of order 1 [7-8]. And in this case, the time series is weakly stationary, i.e.:

a. E (x_t) is constant for all t ≥ 0.

b. Var (x_t) is constant for all t ≥ 0.

c. The covariance of xt and xt−h iis constant for all t ≥ 0.

In addition, the auto-correlation function (ACF) of the time series refers to the fact that for h=1, 2, 3, ..., which can be used in the time series {xt} t≥0. Under the assumption of weak smoothness, the ACF will be simplified to:

$$\mathrm{ACF}(X_t, X_{t-h}) = \frac{\mathrm{Covariance}(X_t, X_{t-h})}{\mathrm{Var}(X_t)} \tag{13}$$

(ii) Parameter estimation

The model was initially determined as ARIMA (p, 1, q) due to the first-order differencing of the original data. Based on the characteristics of the differenced PPM data, the order of the model can be determined based on the autocorrelation and then autocorrelation plots (ACF & PACF plots). The autocorrelation and partial autocorrelation plots are plotted with the help of the functions acf ( ) and pacf ( ) in MATLAB as follows figure 1.

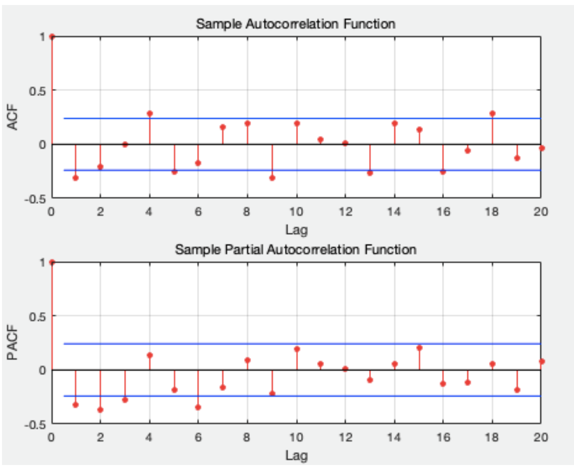

Figure 1. Autocorrelation and Partial Autocorrelation plots

According to the autocorrelation and partial autocorrelation plots, judging the stage of the model generally requires subjective guesses. From the information in the graphs, we can conclude that p=2 and q=1 in the ARIMA model.

(iii) Model diagnosis (residual test)

According to the graph of residual test results, standardized residuals are to see if the residuals are close to normal distribution. According to the QQ figure 2, the blue points are near the red line, so the residuals are close to the normal distribution. Therefore, the model ARIMA (2, 1, 1) is reasonable.

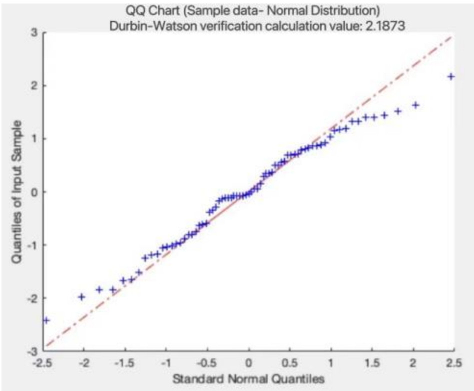

Figure 2. QQ Chart (Sample data-Standard Normal Distribution)

(iv) Model Prediction

Based on the ARIMA (2, 1, 1) model, the carbon dioxide concentration values can be predicted for 2022- 2100. The prediction can be done by MATLAB function, results are shown in the following figure 3 and table 1.

Table 1. Carbon Dioxide Concentration from 2022 to 2100 Predicted by Autoregressive Integrated Moving Average Model

year	carbon dioxide concentration/ ppm	year	carbon dioxide concentration/ ppm	year	carbon dioxide concentration/ ppm
2022	418.8348997	2049	491.079264	2076	580.6828196
2023	421. 1856161	2050	494.0885611	2077	584.3346193
2024	423.5823762	2051	497. 1216545	2078	588.0102154
2025	425.9943962	2052	500. 1785444	2079	591.7096079
2026	428.4334881	2053	503.2592307	2080	595.4327968
2027	430.8951195	2054	506.3637133	2081	599. 1797822
2028	433.3810297	2055	509.4919924	2082	602.9505639
2029	435.8905511	2056	512.6440678	2083	606.745142
2030	438.42394	2057	515.8199396	2084	610.5635165
2031	440.981098	2058	519.0196079	2085	614.4056873
2032	443.5620629	2059	522.2430725	2086	618.2716546
2033	446. 1668202	2060	525.4903335	2087	622. 1614183
2034	448.7953754	2061	528.7613909	2088	626.0749784
2035	451.4477263	2062	532.0562448	2089	630.0123348
2036	454. 123874	2063	535.374895	2090	633.9734877
2037	456.8238179	2064	538.7173416	2091	637.958437
2039	462.295095	2066	545.4736239	2093	645.9997247
2040	465.0664281	2067	548.8874597	2094	650.0560631
2041	467.8615576	2068	552.3250919	2095	654. 1361979
2042	470.6804835	2069	555.7865205	2096	658.2401292
2043	473.5232058	2070	559.2717455	2097	662.3678568
2044	476.3897245	2071	562.7807668	2098	666.5193808
2045	479.2800396	2072	566.3135846	2099	670.6947012
2046	482. 1941511	2073	569.8701987	2100	674.893818
2047	485. 132059	2074	573.4506093		
2048	488.0937633	2075	577.0548162		

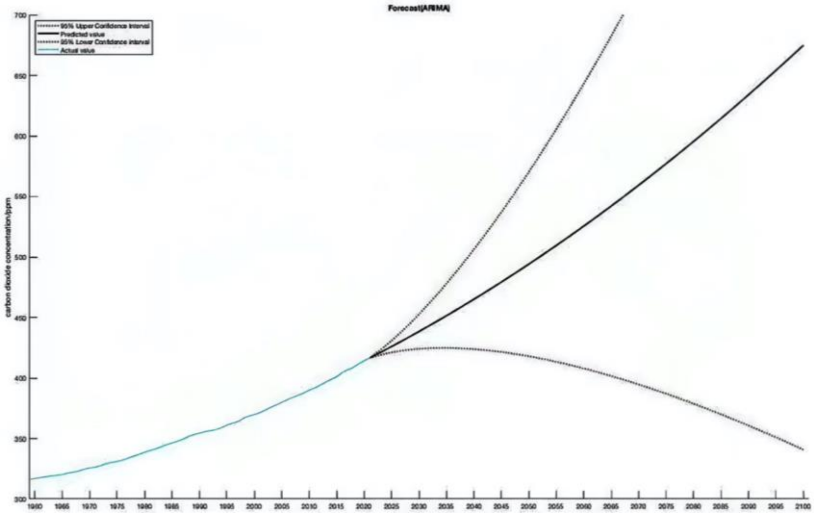

Figure 3. Measured and Predicted Carbon Dioxide Concentrations from 1959 to 2100, enerated by Autoregressive Integrated Moving Average Model

This paper can see that from the graph that the carbon dioxide concentration shows a relatively gradual upward trend, similar to the result predicted by polynomial fitting. It is predicted that in the year 2100, the carbon dioxide concentration is 674.893818 ppm in the atmosphere, a data which does not reach 685 ppm, but close. This paper then predict the level of carbon dioxide concentration for five more years. The result shows that in the year 2103, carbon dioxide concentration is 685. 1202 ppm, first exceeding 685 ppm.

4. Comparison

This paper use three models in total to describe the changes in carbon dioxide concentrations over time and predict future carbon dioxide concentrations from 2022 to 2100.

For the polynomial fitting, the advantage is that it works well for simple relationships with small data volumes. Furthermore, it provides faster access to carbon dioxide concentration data for the forecast years and gives an approximate trend of carbon dioxide concentration changes due to a simpler modeling process. One of the drawbacks is that polynomial fitting takes time if there is a large amount of given data. However, considering carbon dioxide concentration from 1959 to 2100 only contains 142 terms, using polynomial fitting may be considered a good choice.

Grey GM (1, 1) model is adept at dealing with a small amount of data. Even if the numbers of sample in data are not large enough, or the reliability is relatively low, it can still generate a more regular sequence from irregular original data. However, it makes inaccurate predictions for non-linear data samples. Since the carbon dioxide concentration levels over time are accurate and continuous with regularity, the advantage of grey GM (1, 1) model cannot be fully utilized here. In addition, this paper also assuming that the increase in carbon dioxide concentration is a first-order ordinary differential equation merely by objective looking at the graph, which may not exactly match the actual trend.

The last model this paper use is the auto-regressive integrated moving average model, which can work well for most time series sequences. Nevertheless, one of the key limitations of the model is that the parameters (p, d, q) need to be manually defined; therefore, finding the most accurate fit can be a long trial-and-error process. In this paragraph, this paper just subjectively define p=2 and q=1. More reliable results may be gained from multiple trials, which are not covered in this paper. Also, since ARIMA can be limited in forecasting extreme values, outliers are difficult to forecast for ARIMA for the very reason that they lie outside of the general trend as captured by the model. Although this limitation does not largely affect the predicted carbon dioxide concentration from 2022-2100, the model may not able to adapt to future dramatic changes in carbon dioxide concentration.

R2 is a measure of the goodness of fit of a model. This paper calculate the compare the R2 value of carbon dioxide concentrations from 1959 to 2100 for each model, it shows that the R2 values of polynomial fitting, grey GM (1, 1) model and autoregressive integrated moving average model are 0.99998, 0.99989, and 0.99998, respectively. It is found that all the three models are relatively accurate, in which polynomial fitting and auto-regressive integrated moving average models are more accurate

5. Conclusion

Carbon dioxide concentration has begun soaring up after Industrial Revolution. Carbon dioxide emission have direct connection to production, transportation and other daily human activities. When we are worring about the effect of carbon dioxide on global climate, predicting future carbon dioxide level is essential for balancing between economic prosperity and substainable development.

First of all, the team is tasked with developing a model to fit the past and predict the future carbon dioxide concentration levels. This paper first build a second model, grey GM (1, 1) model. This paper begin by confirming the feasibility of using the GM (1, 1) model. The original sequence is accumulated into a new sequence, and then this paper generate the mean sequence, grey differential equation, white differential equation and matrix. This paper finally use the least-squares method to obtain the general solution of the new sequence. Grey GM (1, 1) model predicts that carbon dioxide concentration level will reach 685 ppm in 2132. The third model is autoregressive integrated moving average model. This paper first use differencing to make the data more stable, then found the three orders p, d, and q based on the PACF and ACF plots. The prediction result indicates that in 2103, carbon dioxide concentration level will reach 685 ppm.

References

[1] Li Qian, Fu Wei, Li Mengke. Analysis of factors influencing agricultural carbon emissions and scenario prediction in Yunnan Province [J]. Journal of Southwest Forestry University (Social Science), 2023, 7(03): 75-81.

[2] Bi W-L, Ren X-Y. Factors influencing carbon emissions and peak prediction of construction industry in Shandong Province [J]. Journal of Shandong University of Technology (Natural Science Edition), 2023, 37(04): 20-26.

[3] Cheng H.S., Yu Ran, Wang Q., Ye Yun, Wei Lu. Evolution and classification of urban construction land under the peak carbon emission constraint [J]. Journal of Anhui Agricultural University, 2023, 50(02): 310-318. DOI: 10.13610/j.cnki.1672-352x.20230511.005.

[4] Gao Shan, Cao Mingxia, Li Dan. Analysis of carbon emission characteristics and peak forecast in Jiangsu Province [J/OL]. Resources and Industries: 1-9 [2023-05-18].

[5] FU Xiaohua, WANG He, YANG Xinxin. Research on carbon emission scenario prediction based on STIRPAT model--a case study of Hunan Province [J]. Journal of Shaoyang College (Social Science Edition), 2023, 22(02): 34-42.

[6] Yao M. X., Wang M. V., Lei Y. D.. A study of carbon peak prediction in Shanghai based on STIRPAT model [J]. Fudan Journal (Natural Science Edition), 2023, 62(02): 226-237.

[7] Zhang Yayu, Yang J, Tian Zhiwu. Study on decomposition and prediction of transportation carbon emission factors based on LMDI [J]. Journal of Wuhan University of Technology (Information and Management Engineering Edition), 2023, 45(02): 252-257.

[8] Chen Huan, Mou Ying. Research on the factors influencing carbon emission and its prediction in Chongqing under the "double carbon" target [J]. Journal of Chongqing University of Commerce and Industry (Natural Science Edition), 2023, 40(02): 7-14.

[9] Xinyi Huang, Chengqiang Wang, Jie Su, Xiangqing Zhao. Carbon emission prediction based on GRA-LSTM model--Suqian city as an example [J]. Heilongjiang Science, 2023, 14(05): 25-27+32.

[10] Jiang Chunran, Xu Qingjuan, Su Ting. Prediction of carbon emission influencing factors and carbon peaking combination scenarios in China's coal producing areas based on LMDI and threshold-STIRPAT models [J]. Journal of Nanning Normal University (Natural Science Edition), 2023, 40(01): 32-42.