ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

Multi-dimensional Analysis of Global Warming Based on Neural Network and Grey System

Mu Wenxuan[1,a,*], Meng Jiana[1], Dong Xin[1], Ma Tengfei[1]

[1]*Dalian Minzu University, Dalian, Liaoning, 116650, China*
[a]*mwxgod@qq.com*
[*]*Corresponding author*

Abstract: *The issue of global warming is becoming increasingly severe. To better understand its causes and take timely measures, this paper employs various models such as the Grey prediction model, ARIMA model, fourth-order polynomial regression model, and time series prediction model based on BP neural network to predict and analyze global temperature data. This research holds practical significance in enabling us to better comprehend the trend of global warming and improve our ability to prevent and control its effects.*

Keywords: *Global warming; Gray-scale prediction; ARIMA; Regression fitting; Neural network; Spearman correlation analysis*

1. Introduction

This paper examines the growth trend of the national average temperature between 1975 and 2022. Utilizing Matlab and SPSS tools, various methods are employed to analyze the data and predict the global average temperature in 2050 and 2100 under different trends. Furthermore, through correlation analysis, the impact of time, longitude and latitude, natural disasters, and other factors on global temperature are studied. This approach provides a more objective understanding of the main factors related to global warming, enabling us to take timely measures to protect the earth we inhabit.

2. Background

In recent years, there have been more frequent and alarming reports of rising temperatures across the globe. Many countries in the Middle East have seen temperatures exceed 50 degrees Celsius, while places like Italy, Canada, and California have experienced unprecedented high temperatures. There is no denying that the Earth is heating up at an alarming rate. Prior to the Industrial Revolution, the Earth did not experience such intense heat. At that time, the carbon dioxide concentration was approximately 280 ppm. However, as of March 2004, the carbon dioxide concentration had risen to 377.7 ppm, representing the largest average increase in 10 years. The root cause of this Earth burning is global temperature warming, which is a natural phenomenon resulting from an imbalance in the Earth's atmospheric system's energy absorption and emission, caused by the continuous accumulation of the greenhouse effect.

3. Data Source

The data utilized in this paper includes the national average temperature data from 1975 to 2022, average temperature, longitude, and latitude data at various times in different cities around the world, CO2 concentration data, and the number of people infected with the COVID-19 virus in recent years. The data source is the Berkeley Earth data page, available at http://berkeleyearth.org/data/.

4. Analysis and Modelling

4.1. Prediction model and analysis

4.1.1. Model1: Grey Model

The [6]grey model is a method used to predict systems that contain uncertain factors. [7] Its basic idea

ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

is to generate a new series by accumulating the observed values of the time series. The new series has the same dimensions as the original series.

The grey model can be expressed through the following mathematical model:

Suppose the time series $x^{(0)}$ has n observations [1] $x^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n),\right\}$

Generate new sequence by accumulation [1] $x^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n),\right\}$ [1], as follows:

$$x^{(1)} = \sum_{i}^{k}\left[x^{(0)}(i)\right] \qquad (k = 0, 1, \cdots) \tag{1}$$

The corresponding differential equation of the GM (1,1) model is: [2] $dx / dt + ax = b$

The grey prediction model is obtained by restoring the following equation:

$$x^{(0)}(k+1) = x^{(1)}(k+1) - x^{(1)}(k) = (1 - e^{a})(x^{(0)}(1) - \frac{b}{a})e^{-ak} \qquad (\mathrm{k} = 1, 2, \cdots n) \tag{2}$$

Result of Model1: Firstly, the annual average temperature data from the past is plotted and analyzed:

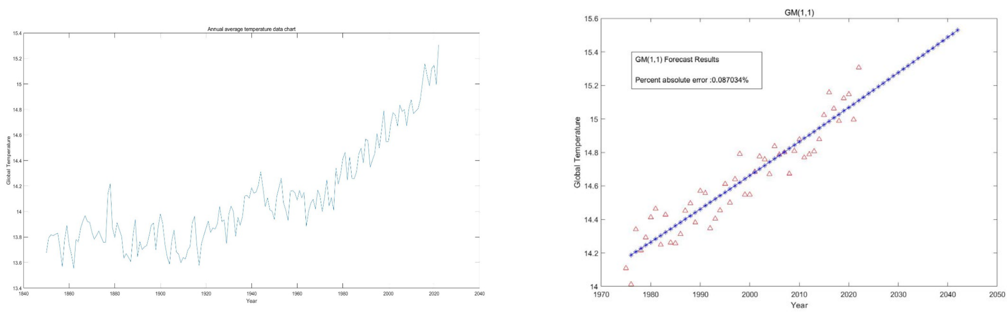

Figure 1: Annual average temperature data chart (left)& GM (1,1) Model(right).

The above figure 1 clearly shows that at the beginning of the 20th century, shortly after the end of the second industrial revolution, there was a short-term upward trend in the global average temperature. In the 1970s, there was a significant and sustained upward trend in the global average temperature, which continues to this day. Since historical data from hundreds of years ago may not be relevant to current impacts and the grey prediction model requires less data, we used data from 1975 to 2022 to establish and solve the model, as shown on the right side of Figure 1:

The percentage absolute error under this model is 0.087034%, indicating good prediction accuracy under the current trend. Furthermore, the model predicts the temperature for the next 20 years. Given the low prediction error, we believe that this model has high accuracy.

4.1.2. Model2:ARIMA

The [4]ARIMA model is an important method used for studying time series. [5]The basic idea of the ARIMA model is to treat the data series formed by the prediction object over time as a random series and describe the series approximately with a certain mathematical model.

1) [4]"AR" in ARIMA stands for "autoregression," which refers to the autoregressive model. The autoregressive model utilizes the historical data of the factor itself to predict the future values. The expression of the autoregressive model is as follows:

$$^{[4]}\, y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{3}$$

Where y_t is the value of time series y at time t, $\phi_1, \phi_2, \cdots, \phi_p$ is the autoregressive coefficient, p is the autoregressive order , and { ε_t } is a white noise sequence.

2) [4]"MA" in ARIMA stands for "moving average," which refers to the moving average model. The moving average model is used to address the issue of white noise in the autoregressive model. The expression of the moving average model is as follows:

ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

$$^{[4]}\ y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{4}$$

Where $\theta_1, \theta_2, \cdots, \theta_q$ is the moving average coefficient and q is the moving average order.

3) [3]"ARMA" in ARIMA stands for "autoregressive moving average," which refers to the autoregressive moving average model. The autoregressive moving average model is obtained by combining the autoregressive model with the moving average model, and its expression is as follows:

$$^{[4]}\ y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{5}$$

Result of Model 2:

In Model 1, it can be observed that the data has shown a clear upward trend since the 1970s. Therefore, we have performed second-order difference processing on the data, as follows:

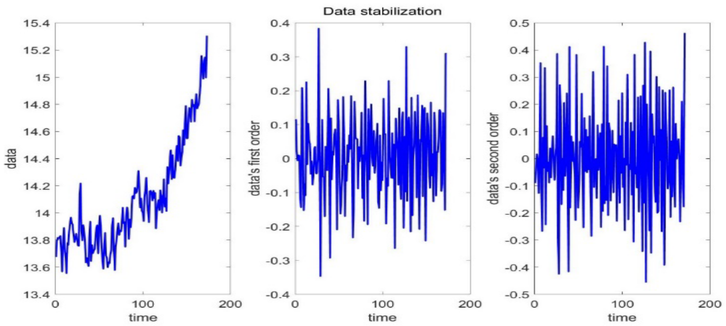

Figure 2: Data stabilization.

As shown in the above figure 2, the global temperature has shown a certain upward trend during this period, and the ARIMA model cannot be directly used for modeling. Second-order difference can eliminate the trend of data growth. After second-order difference, the predicted data series is graphically displayed, which basically eliminates the influence of the long-term trend and tends to be stable, meeting the basic requirements of ARIMA model modeling. The autocorrelation and partial correlation data graphs are obtained as follows:

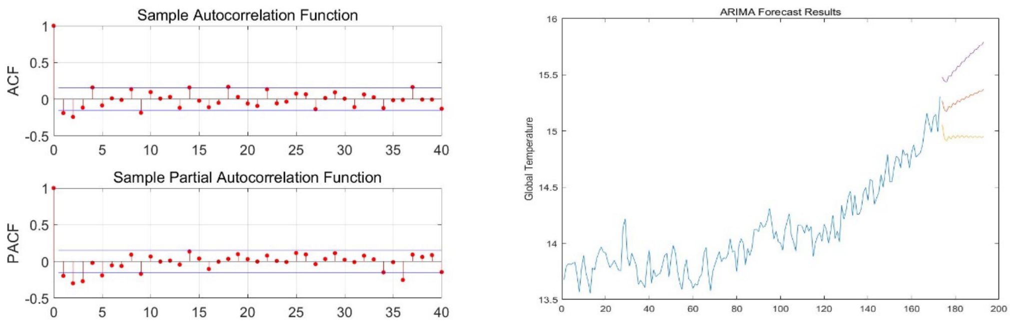

Figure 3: Sample Autocorrelation Function&Sample Partial Autocorrelation Function (left) & ARIMA Forecase Result (right).

The analysis results of stationarity, periodicity, autocorrelation, and partial correlation of annual data provide guidance for the construction of the ARIMA model and parameter optimization. After analysis and multiple parameter optimizations, the ARIMA (3, 1, 0) model was used to fit and forecast the annual data, and the prediction effect of the model was found to be the best. The prediction results are shown on the right side of Figure 3.

4.1.3. Model3: Regression fitting

In the regression fitting, we used linear regression fitting, second-order polynomial fitting, third-order polynomial fitting, and fourth-order polynomial fitting to fit the average temperature data over the years. The mathematical models are as follows:

$$y(x,w) = w_0 + w_1 x + w_0 + w_1 x^2 + \cdots + w_M x^M = \sum_{(j=0)}^{M} w_j x^{j(9)} \tag{6}$$

ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

For the different fitting methods, we have drawn the residual graph and residual norm as follows:

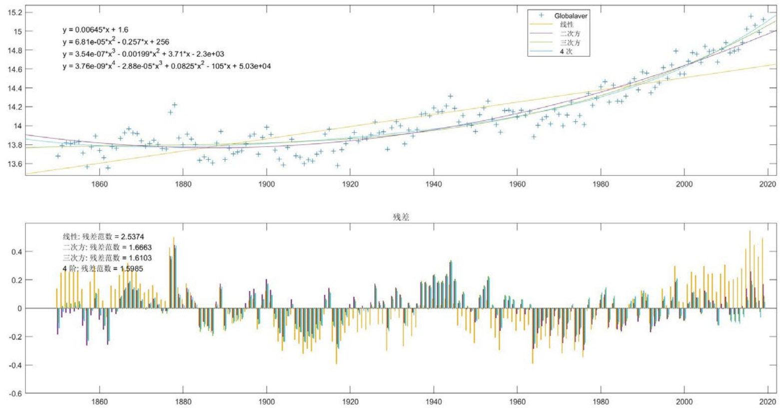

Figure 4: Fitting regression diagram.

As observed from the figures 4, the fourth-order polynomial fitting is good, and the residual norm is small, indicating a good prediction effect of this model. The residual graph shows that the residuals are randomly scattered around zero, while the residual norm indicates that the residuals follow a normal distribution. Therefore, it can be concluded that the fourth-order polynomial model is the best fit for the data and has a good prediction effect.

4.1.4. Model4: Prediction of Time Series Based on BP Neural Network

Model introduction

The artificial neural network simulates a series of operational mechanisms of the biological neural network. The basic structure of the model is illustrated in the figure 5.

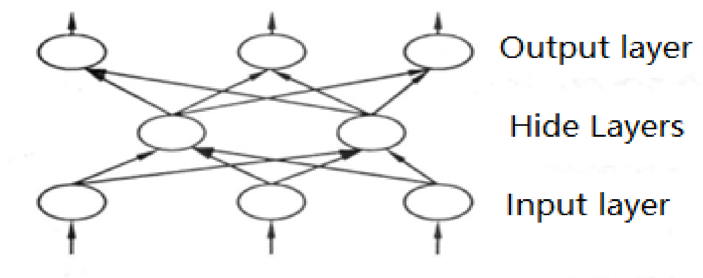

Figure 5: Structure of bp neural network.

Model establishment:

Structurally, a neural network can be divided into an input layer, an output layer, and hidden layers (see Figure 5). Each node in the input layer corresponds to a predictor variable, while each node in the output layer corresponds to a target variable, which can be multiple. Between the input and output layers lies the hidden layer(s), which are not visible to the user of the neural network. The number of hidden layers and the number of nodes in each layer determine the complexity of the neural network.

Except for the nodes in the input layer, each node in the neural network is connected to many of the nodes before it (called input nodes). Each connection corresponds to a weight W_{xy}, and the value of the node is the sum of the products of the input node values and their corresponding connection weights, which is then input into an activation or squeezing function.

Forward propagation: The process of data moving from input to output is a forward propagation process, where the value of each node is passed on from the nodes before it, then weighted by the connection weights and input into the activation function to obtain a new value, which is then propagated to the next node.

Feedback: When the output value of a node is different from the expected value, i.e., an error occurs, the neural network needs to "learn" from the error. We can think of the weights connecting the nodes as the "trust" level of the later node in the previous node (i.e., the output of the later node is more easily influenced by the input of the earlier node with a higher weight). The learning process involves punishment. If a node's output is in error, it looks at which input nodes (or node) caused the error and whether the nodes it trusts the most (i.e., nodes with the highest weights) betrayed it (causing the error). If so, the trust value (weight) of those nodes is reduced, punishing them, while the trust value of the nodes that made correct suggestions is increased. For the nodes that have been punished, they also need to punish the nodes before them in the same way. The punishment is propagated step by step until it reaches the input nodes.

Model solution:

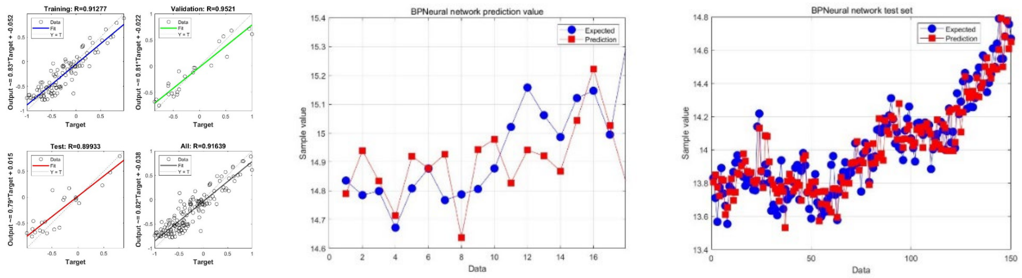

Figure 6: Neural network training results (left)&neural network test results (middle)& Region(right).

Based on the data spanning almost 168 years, we have used time series analysis to generate smooth training and testing data for the neural network. It is assumed that the temperature of a given year is influenced by the average temperature of the preceding five years, where the temperature of the previous five years serves as the input data, and the temperature of the current year serves as the actual data or output. We have used data samples from nearly 140 years for training the neural network, and the data of the last 18 years have been used for validation. The results are as follows:

In the figure 6, the left and middle figures represent the degree of fit of the training and test samples, respectively. And it is evident from the results that the R value of the training set is 0.91277, while the R value of the validation set is 0.9521, indicating that the current model has a good generalization performance. The overall R value is 0.91639, which is indicative of the model's high prediction accuracy.

4.1.5. Forecast of future global temperature

We have used the four prediction models established in 4.1.1-4.1.4 above to predict the global average temperature in the years 2050 and 2100 respectively. The prediction results are presented in table 1:

Table 1: Forecast of Global Average Temperature

Year	*gray prediction*	*ARIMA*	*Regression (fourth order)*	*BP*
2050	*15.70℃*	*15.64℃*	*16.35℃*	*15.81℃*
2100	*16.81℃*	*16.62℃*	*20.50℃*	*16.57℃*

4.1.6. Model accuracy analysis

Based on the results presented in Section 4.1.5, it is evident that the global average temperature predicted by each method in 2050 and 2100 is accurate. We believe that these four models have high accuracy in different situations.

First of all, the grey prediction model and ARIMA model show similar trends, and if the current growth rate of global average temperature persists, these models are likely to have the most accurate predictions. This is because both models rely on the current trend of the data to make predictions.

Secondly, the fourth-order polynomial regression model exhibits a clear upward gradient, and its predicted data also follow this trend. If the global average temperature continues to show an upward trend, the prediction accuracy of this model is likely to be higher. For instance, the model predicts that the global average temperature will reach 20 ℃ in 2100.

Finally, the time series prediction based on neural network shows a growth trend in its predicted results that is not as good as the other three models, as shown in Section 4.1.6. If humans make efforts to reduce environmental damage to the Earth, the global temperature increase may slow down. Eventually, it may reach a point where it stops growing and maintains a dynamic stability of the global average temperature.

ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

In summary, our four models have provided good prediction results under different conditions. However, given the current trend of social development, the prediction results of the fourth-order polynomial regression model are considered more realistic. This model predicts that the global average temperature will reach 20 ℃ in 2100.

4.2. Research and analysis on the causes of global temperature rise

4.2.1. Pearson correlation model

The Pearson correlation model is used for correlation analysis based on normal continuous distribution data. Its model expression is shown in the following equation:

$$\rho_{X,Y}=\frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \tag{7}$$

The two factors analyzed by the Pearson correlation model are represented by X and Y in the following equation, while *E* denotes the mathematical expectation.

4.2.2. Spearman correlation model

The Spearman correlation, also known as the Pearson correlation between rank variables, can handle variables with any distribution. Its equation is expressed as follows:

$$\rho_{x,y}=\frac{\sum_i(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_i(x_i-\overline{x})^2\sum_i(y_i-\overline{y})^2}} \tag{8}$$

The two factors analyzed by the Spearman correlation model are represented by x and y in the following equation, while $\overline{x}$ and $\overline{y}$ denote the mean values of the two factors:

4.2.3. The relationship between global temperature, time and location

To determine the appropriate correlation model for modeling, we first test the normality (in table 2) of the date, average temperature, longitude, and dimension data. If they conform to normality, the Pearson correlation model is used for modeling. Conversely, if the data does not conform to normality, the Spearman correlation model is used for modeling.

Table 2: Normality test of date, average temperature, longitude and latitude

	Null hypothesis	Test	Significance	Decision maker
1	The distribution of dates is normal	Single sample Kolmogorov Smirnov test	0.00[1]	Reject original assumption
2	The distribution of average temperature is normal	Single sample Kolmogorov Smirnov test	0.00[1]	Reject original assumption
3	Longitude distribution is normal distribution	Single sample Kolmogorov Smirnov test	0.00[1]	Reject original assumption
4	Latitude distribution is normal distribution	Single sample Kolmogorov Smirnov test	0.00[1]	Reject original assumption

4.2.4. Results and Analysis

a) Correlation results of average temperature and date

We utilized SPSS to test the correlation between date and global temperature, and the resulting test findings are presented in table 3.

The test results indicate a significant correlation between date and average temperature, with a correlation coefficient of 0.05 (2-tailed). The temperature change is consistent with the change in seasons.

ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

Table 3: Spearman Correlation Results of Date and Average Temperature

Spearman's Rho		Date	Average Temperature
Date	Correlation Coefficient	1.000	0.034
	Sig.(2-tailed) N	3764	0.036 3764
Average Temperature	Correlation Coefficient	0.034	1.000
	Sig.(2-tailed) N	0.036 3764	3764

b) Results of correlation between longitude and average temperature

In this model, east longitude is represented by positive values and west longitude by negative values. And SPSS was used to test the relationship between average air temperature and longitude. The test results are presented in table 4:

Table 4: Spearman Correlation Results of Longitude and Average Temperature

Spearman's Rho		Longitude	Average Temperature
Longitude	Correlation Coefficient	1.000	-0.031
	Sig.(2-tailed) N	3764	0.058 3764
Average Temperature	Correlation Coefficient	-0.031	1.000
	Sig.(2-tailed) N	0.058 3764	3764

The results of the test indicate that there is no significant correlation between longitude and average temperature. In other words, longitude does not have a direct impact on temperature changes.

c) Results of correlation between latitude and average temperature

In this model, positive values represent north latitude and negative values represent south latitude. To test the relationship between latitude and average temperature, SPSS was used and the test results are shown in table 5:

Table 5: Spearman Correlation Results of Latitude and Average Temperature

Spearman's Rho		Latitude	Average Temperature
Latitude	Correlation Coefficient	1.000	-0.468
	Sig.(2-tailed) N	3764	0.000 3764
Average Temperature	Correlation Coefficient	-0.468	1.000
	Sig.(2-tailed) N	0.000 3764	3764

The test results indicate a significant correlation between latitude and average temperature, with a correlation coefficient of 0.01 (two-tailed). Specifically, the temperature is higher closer to the equator.

To sum up, date and latitude are the two factors most closely related to the average temperature.

5. Model Evaluation

5.1. Strengths

Comprehensive consideration: We used four different models to predict the future temperature, taking into account different temperature growth trends. Different models have better prediction accuracy under different growth trends; Full utilization of information: We have made full use of all historical data in the fourth order polynomial regression and neural network prediction, so that the model involves more factors and the results are more reliable.

5.2. Weaknesses

Because the grey prediction model is used in the case of a small amount of data, the prediction effect is poor under a large amount of data in the past, so the data in recent decades is selected for prediction, and the prediction result is relatively good in local performance. However, if the impression factor of all historical data is taken into account, the prediction effect of this model is not outstanding.

ISSN 2616-5775 Vol. 6, Issue 4: 1-8, DOI: 10.25236/AJCIS.2023.060401

6. Conclusion

Through prediction and analysis, it has been discovered that there is a significant correlation between latitude and average temperature, as well as between date and average temperature. However, there is no significant correlation between longitude and average temperature. Without curbing the trend of global temperature increase, it is highly likely that the future will align with the fourth-order regression prediction model, with an average temperature of 20.50℃ and some areas possibly exceeding 50℃ year-round. This is unacceptable for human beings and could even be catastrophic. Therefore, we urge everyone to protect the environment without delay, for a better tomorrow.

References

[1] Chungang Li. Application of grey forecasting method in commercial chain enterprises [J]. Economist, 2001, (12)
[2] Hua-kai Chiou, Gwo-Hshiung. Grey Prediction Model Forthe Planning Material of Equipment Spare parts in Navy of Taiwan[C].Proceedings of World Automation Congress,2004(17):315-320
[3] George E.P. Box, Gwilym M. Jenkins. Time series analysis-forecasting and control [M]. Beijing: China Statistics Press, 1997:101-149
[4] Shujing Zhang, Lixin Qi. Concise tutorial on time series analysis [M]. Beijing:Northern Jiaotong University Press, 2003:24-59
[5] Weixin Wang, Jinguan Du, Yougui Wu .et al. Applied time series analysis [M]. Guilin:Guangxi Normal University Press, 1999:102-183
[6] Julong Deng. Basic methods of gray systems [M]. Wuhan: Huazhong University of Technology Press, 1987.
[7] Chunyan Lin. A new gray prediction model and its application in population management [J]. Population and Economy, 2002.10.