ISSN 2616-5872 Vol.5, Issue 3: 21-29, DOI: 10.25236/AJEE.2023.050304

Research on whether global temperature warms and its influencing factors based on grey predictions

Gaopeng Xiang[1], Kaijun Guo[2], Yuhu Li[3], Hongnan Jin[3], Rumeng Lin[4]

[1]*School of Science, Heilongjiang University of Science and Technology, Harbin, 150022, China*
[2]*School of Electrical and Control Engineering, Heilongjiang University of Science and Technology, Harbin, 150022, China*
[3]*School of Management, Heilongjiang University of Science and Technology, Harbin, 150022, China*
[4]*School of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin, 150022, China*

Abstract: *In order to study whether the global temperature warms and the biggest factors affecting temperature change, this paper collects the global average temperature from 1900 to 2022, sets up an experimental group and a control group to train three types of grey prediction models to select the optimal model, selects the metabolic type and analyzes the error, and uses the least squares method to fill in the difference to achieve future temperature prediction, and judges whether the future temperature warms by comparing the century temperature growth rate. Then, the least squares method, multiple linear regression and comprehensive evaluation were used to analyze the influence between natural and regional factors and temperature change, and the influence factors of natural and human factors on temperature change were calculated by grey correlation to determine the maximum influencing factors. The results show that the temperature growth rate in the 20th century is 1.0412, and the temperature growth rate in the 21st century is 1.0958 higher than that of 0.0546 in the 20th century, so the global temperature will warm in the future; the biggest factor affecting global warming is "forest area", with an impact coefficient of 0.9376.*

Keywords: *temperature change, grey prediction, least squares, grey correlation, comprehensive evaluation system*

1. Introduction

'Global warming' is a phrase that refers to the effect on the climate of human activities, in particular the burning of fossil fuels (coal, oil and gas) and large-scale deforestation [1].

The rainfall distribution might change and the frequency of severe weather events, such as hurricanes and typhoons, might increase [2]. A common view is that the current global warming rate will continue or accelerate [3]. In this paper, mathematical models such as grey predictions are used to analyze whether the climate is warming and the biggest factors contributing to warming.

2. Materials and methods

2.1 Data

Through to the Berkeley earth data page (http://berkeleyearth.org/data/) search, it can know that the average global temperature data from 2019 to 2022 and the factors influencing the temperature of the data. The statistical table of natural and man-made disasters (only earthquakes, volcanoes, storms, floods and forest areas are used in this paper).

The statistical histogram is shown in Figure 1 and Figure 2 below:

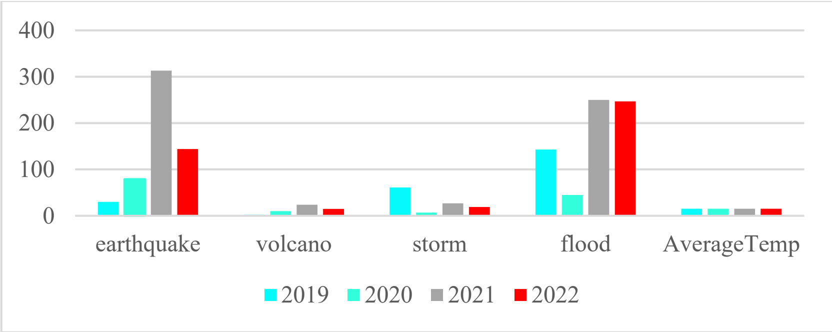

Figure 1: Natural Disaster and Average Temperature Visual Chart 2019-2022

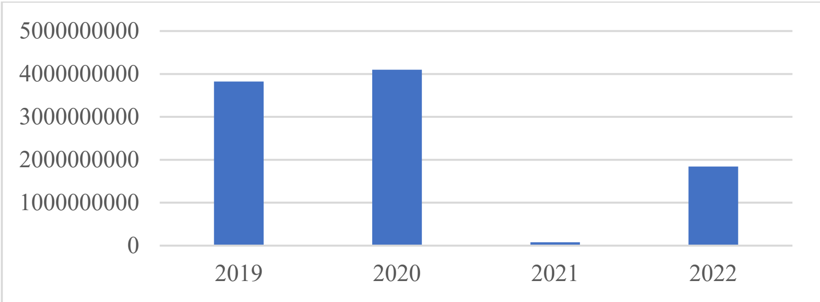

Figure 2: Human Factors Statistics Chart 2019-2022

The natural disasters covered in this article are all from the *Global Assessment of Natural Disasters from 2019 to 2022*.

2.2 Method introduction

2.2.1 Grey forecast

The GM(1,1) model uses the original discrete non-negative data column, generates a new discrete data column that weakens randomness and is more regular by accumulation, and obtains the solution at the discrete point through differential equations, and accumulates to generate an approximation of the original data. This article is recorded $T^{(0)}=\left(T^{(0)}(1),T^{(0)}(2),\cdots,T^{(0)}(n)\right)$ as the original temperature data column; A new column of data generated $T^{(1)}=\left(T^{(1)}(1),T^{(1)}(2),\cdots,T^{(1)}(n)\right)$ for one accumulation; Generates a series for the $z^{(1)}$ immediate mean of the series $T^{(1)}$, for

$$z^{(1)}(m)=\delta T^{(1)}(m)+(1-\delta)T^{(1)}(m-1), m=2,3,\cdots,n;\ \delta=0.5$$

The basic morphology of the GM(1,1) model, $T^{(0)}(k)+az^{(1)}(k)=b$ where b represents the grey action and a represents the development coefficient, and then the matrix form:

$$u=(a,b)^{T}, Y=\begin{bmatrix}T^{(0)}(2)\\T^{(0)}(3)\\\vdots\\T^{(0)}(n)\end{bmatrix}, B=\begin{bmatrix}-z^{(1)}(2) & 1\\-z^{(1)}(3) & 1\\\vdots & \vdots\\-z^{(n)}(n) & 1\end{bmatrix} \qquad (1)$$

It can be expressed as using the least squares method to calculate the $T^{(0)}(k)+az^{(1)}(k)=b \quad Y=B\mu$ estimates of parameters a and b and use the whitening equation $\frac{dT^{(1)}(t)}{dt}=-\hat{a}T^{(1)}(t)+\hat{b}$ to obtain the final expression of the GM(1,1) model:

ISSN 2616-5872 Vol.5, Issue 3: 21-29, DOI: 10.25236/AJEE.2023.050304

$$\hat{T}^{(1)}(t)=\left[T^{(0)}(1)-\frac{\hat{b}}{\hat{a}}\right]e^{-\hat{a}(t-1)}+\frac{\hat{b}}{\hat{a}} \tag{2}$$

When making predictions on raw data, you only need to be satisfied $m \geq n$

2.2.2 Least squares method (OLS)

In least squares, the fitted value is denoted as $\widehat{y_i}=kx_i+b$. The following conditions k,b make the regression coefficients more precise:

$$\hat{k},\hat{b}=\arg_{k,b}\min(\sum_{i=1}^{n}\left(y_i-\widehat{y_i}\right)^2) \tag{3}$$

In order to ensure the minimum error, it is only necessary to ensure that the loss function value is small, and $\frac{\partial L}{\partial k}=0$ $\frac{\partial L}{\partial b}=0$ .The solution expression of the least squares coefficient can be completed.

2.2.3 Multiple linear regression

Multiple regression analysis (MR) is a highly flexible system for examining the relationship of a collection of independent variables (or predictors) to a single dependent variable (or criterion) [4]. Exploring the correlation between global temperature and various influencing factors is a multi-input and single output *problem*. Relevant characteristic indicators can include volcanic eruptions, COVID-19 epidemic, and so on. It is difficult to directly explore the correlation between them. Therefore, we use multiple linear regression to solve the problem.

The established model is:

$$\begin{cases} z=\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m+\varepsilon \\ \varepsilon\sim N(0,\sigma^2) \end{cases} \tag{4}$$

2.2.4 Grey Relation Analysis

Grey Relation Analysis (GRA) is a multivariate statistical analysis method. The GRA models developed from the models based on relation coefficients of each point in the sequences in early days to the generalized GRA models based on integral or overall perspective [5]. This concept is relative to the white system and the black system.

Set four years of earthquakes, volcanoes, storms and floods as independent variables $x_1,\ldots,x_n$,and be recorded as X, and set the maximum temperature for four years to $t_1,\ldots,t_n$,recorded as T. From this, it can be concluded that:

$$X=\left[x_1,x_2,x_3,x_4,x_5\right] \tag{5}$$

$$T=\left[t_1,t_2,t_3,t_4,t_5\right] \tag{6}$$

After that, preprocess the data, forward it and calculate the average value:

If the forward matrix is ω, then:

$$\omega=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ x_{n1} & \cdots & \cdots & x_{nm} \end{bmatrix} \tag{7}$$

Where χ is the minimum difference between two poles, τ is the maximum difference between two poles, and ρ is the resolution coefficient

$$\chi = \min_i \min_k \left| x_0(k) - x_i(k) \right| \quad (8)$$

$$\tau = \max_i \max_k \left| x_0(k) - x_i(k) \right| \quad (9)$$

If $y\left(x_0, x_i\right)$ is defined as grey correlation degree, then:

$$y\left(x_0, x_i\right) = \frac{1}{n}\sum_{k=1}^{n} y\left(x_0(k), x_i(k)\right) \quad (10)$$

3. Model building and solving (results and analysis process)

3.1 Quasi-exponential law test on the data

The data that pass the quasi-exponential law test can be applied to the grey prediction model, so it is necessary to perform the quasi-exponential law test on the data first, and draw the global annual average temperature sequence plot as shown in Figure 3:

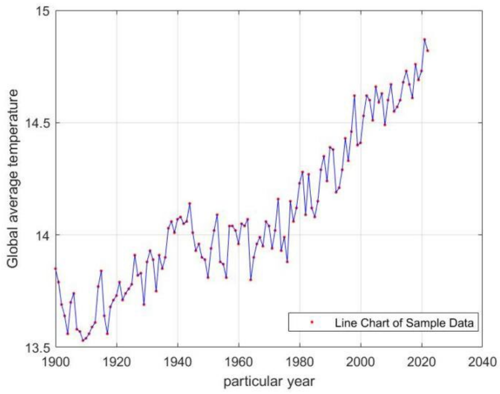

Figure 3: Time series plot of global mean temperature and year from 1900 to 2022

Data series that accumulate r times $T^{(1)}$ is $\sigma\left(k\right) = \frac{T^{(1)}\left(k\right)}{T^{(1)}\left(k-1\right)}$ and the smooth ratio of the original series is $\rho\left(k\right) = \frac{T^{(0)}\left(k\right)}{T^{(1)}\left(k-1\right)}$, and the higher the proportion, the better the quasi-exponential law of the model. Using this principle, the relationship between the smoothness and year of the raw data is generated with the help of MATLAB as shown in Figure 4 below.

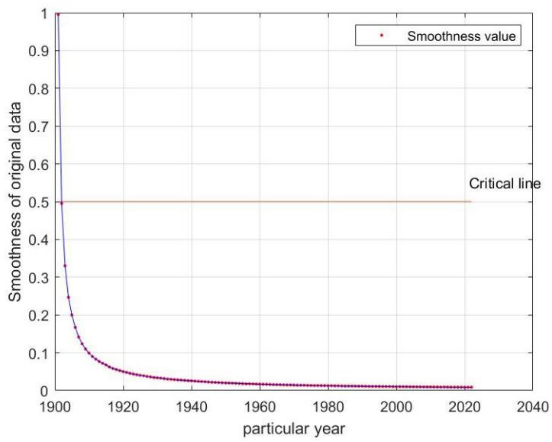

Figure 4: Relationship between the year and the smoothness ratio of the original data

This model can assume that when the proportion of data with a smoothness ratio of less than 0.5 of the original data is greater than 90%, the data passes the quasi-exponential law test. The proportion of data with a smoothness ratio less than 0.5 in the temperature data column was 99.15%, which could be considered to pass the quasi-exponential law test and conform to GM(1,1). Model usage conditions.

3.2 Selection and accuracy analysis of three types of grey prediction models

The model established by the original data series is the traditional all-data grey prediction model $GM(1,1)$. The model $T^{(0)}=\left(T^{(0)}(1),T^{(0)}(2),\cdots,T^{(0)}(n),T^{(0)}(n+1)\right)$ established by the data column is the new information grey prediction model; Data column the model $T^{(0)}=\left(T^{(0)}(2),\cdots,T^{(0)}(n),T^{(0)}(n+1)\right)$ established is a metabolic grey prediction model. The prediction effect of the three types of grey prediction models is different, so you need to set up indicators and choose the best prediction model.

Divide the temperature data into a training group and a test group. The training group is 1900-2013 temperature data, the test group is the predicted temperature data of the corresponding year, and the accuracy of the three models is compared through training, and the training result is shown in Figure 5 below.

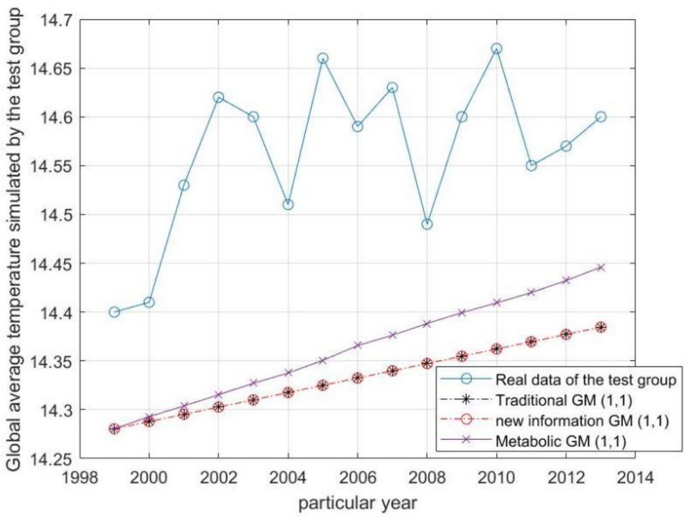

Figure 5: Fitted values of the test group of Class 3 GM(1,1) models

From the training results, it can be seen that the temperature data trained by the metabolic GM (1,1) model is more in line with the original data, and the sum of squared errors is 0.66472. This model is more appropriate for future temperature projections. Then, for the selected metabolic GM(1,1) model, the training results were analyzed by residual test and cascade error as shown in Figure 4 below, and the results were 0.84% and 0.59%, respectively , the mean relative residual $\overline{\varepsilon_r}$ and mean grade ratio deviation $\eta(k)$ are calculated as follows:

$$\overline{\varepsilon_r}=\frac{1}{n-1}\sum_{k=2}^{n}\left(\left|\frac{T^{(0)}(k)-\hat{T}^{(0)}(k)}{T^{(0)}(k)}\right|\times 100\%\right),k=2,3,\cdots,n \tag{11}$$

$$\eta(k)=\left|1-\frac{1-0.5\hat{a}}{1+0.5\hat{a}}\frac{1}{\sigma(k)}\right| \tag{12}$$

It is generally believed that when the mean relative residual $\overline{\varepsilon_r}$ and mean grade ratio deviation is less than 1 $\eta(k)$ 0%, the model fits the original data very well. Therefore, the results show that the training effect of the model on temperature data is very good. Figure 6 below shows:

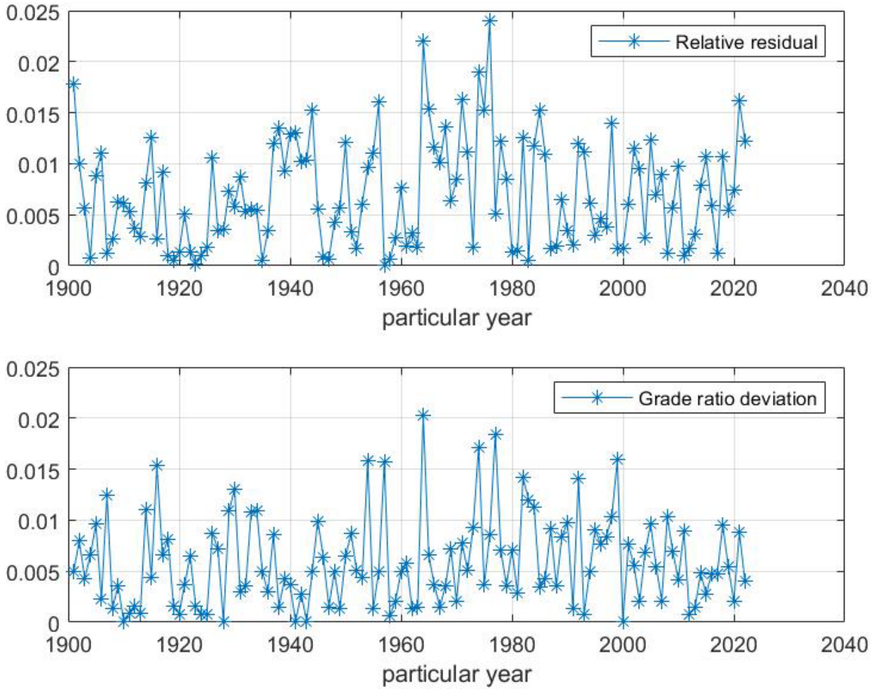

Figure 6: Residual test and error analysis of metabolic GM(1,1) model

3.3 Projections of future global temperatures

After analysis, the training value of the metabolic GM(1,1) model has a certain error, then the least squares method needs to be used to determine the error range, and the predicted temperature value plus the difference is the true temperature of the predicted year. Using least squares for the training group data yields k = 0.0072, b=0.0074, regression sum of squares SSR = 6.3245, and sum of error squares SSE =2.5791, sum of population squares SST= 11.2497, goodness-of-fit $R^2 = 0.5622$;

Analysis of Figure 7 shows that the average temperature difference between the fitted value and the training value is 0.2296, and the temperature value of 2100 years is predicted by this method to be 15.79 degrees Celsius, and the temperature growth rate in the 20th century is calculated as 1.0412; The temperature growth rate in the 2nd century was 1.0958, higher than 0.0546 in the 20th century. Therefore, the global temperature will warm in the future.

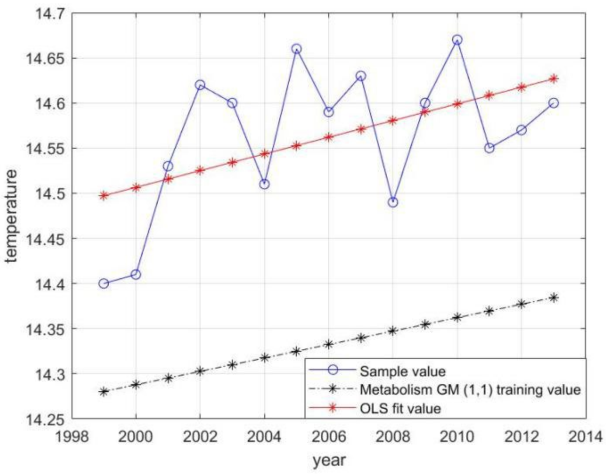

Figure 7: Comparison of fitted values with training data

ISSN 2616-5872 Vol.5, Issue 3: 21-29, DOI: 10.25236/AJEE.2023.050304

3.4 Study on the effect of regionalism on temperature

First of all, the data will be visually analyzed, the 100-year global average temperature as the research object, through the MATLAB software to establish a heat map to make the annual time and temperature change relationship more intuitive, to 10 years of analysis, blue represents the temperature, the darker the color represents the higher the temperature, the lighter the color represents the lower the temperature. Figure 8 and Figure 9 are shown below:

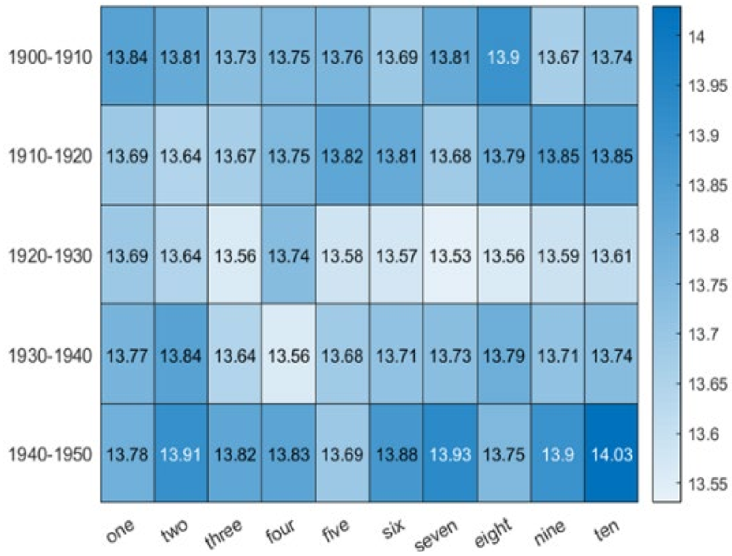

Figure 8: Visual plot of 1900-1950 time change with temperature

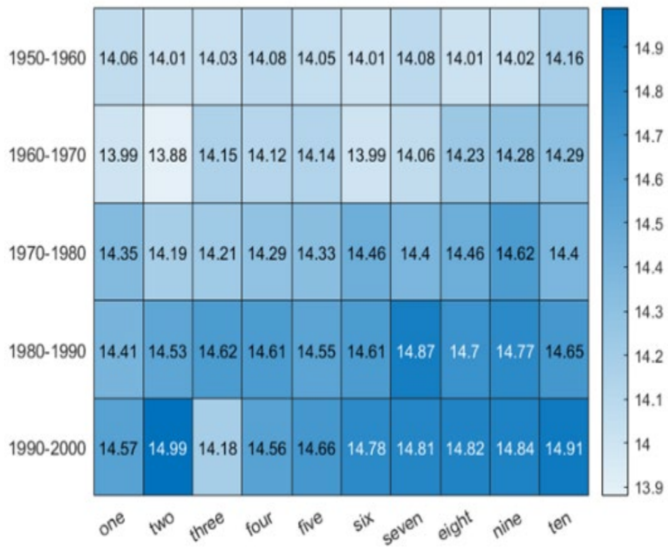

Figure 9: Visual plot of 1950-2000 time change with temperature

Our group conducts research on continents, which are divided into six regions: Asia, Europe, North America, South America, Africa and Oceania. We do not consider and analyze Antarctica. We establish a comprehensive evaluation model system between time, continent and temperature to explore the relationship between temperature and continent.

The comprehensive evaluation equation is established as follows:

- The temperature increase is recorded as the temperature addition value, which is recorded as T
- Time is recorded as DY in 10 years
- The time related change value of temperature in a continent is recorded as Z

The survey data can be obtained as shown in Table 1 of the following table:

Table 1: Data for each decade by geography

Every ten years: DY	Asia	Europe	North America	South America	Oceania	Africa
10	0.08	0.07	0.06	0.02	0.05	0.01
20	0.05	0.09	0.05	0.04	0.01	0.02
30	0.03	0.04	0.04	0.07	0.09	0.04
40	0.09	0.08	0.09	0.05	0.08	0.05
50	0.12	0.11	0.10	0.11	0.09	0.08
60	0.15	0.12	0.11	0.12	0.08	0.10
70	0.16	0.13	0.16	0.10	0.16	0.11
80	0.19	0.22	0.15	0.23	0.19	0.15
90	0.2	0.23	0.26	0.19	0.23	0.17
100	0.21	0.24	0.31	0.28	0.20	0.20

$$Z = DY * T \quad (13)$$

Calculate Z as shown in Table 2 of the following table:

Table 2: Calculation results for different regions

continent	Z
Asia	81.7
Europe	91
North America	95.6
South America	88.9
Oceania	83
Africa	68.7

It can be seen from the above that, regardless of population mobility and special circumstances, our group found that with the change of the time unit of 10 years, in the areas with denser population or more developed technology and industry, the temperature rises faster. Our group can think that the region and temperature show a correlation change in the unit of 10 years.

3.5 The study of multiple factors affecting temperature by grey correlation and multiple linear regression

The goodness of fit is $R^2 = 0.9883$,There are graphs available, and the fitting results are not too different. And the regression model is available as a whole. Therefore, the final expression is as follows:

$$y = f(x_1, x_2, x_3, x_4, x_5) = 1.7029 - 0.4213x_1 - 0.9833x_3 - 0.1875x_4$$

That is, these natural disaster factors (earthquakes, volcanic eruptions, storms, floods) and human factors will have an impact on global warming.

And using MATLAB to solve, we can get:

1) The normalized image of the table (The final figure is shown in Figure 10 below):

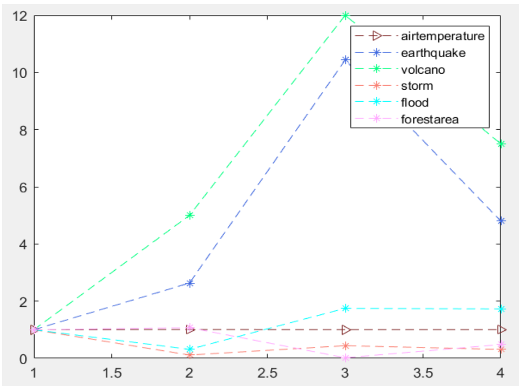

Figure 10: The normalized image

2) Grey incidence image (The final figure is shown in Figure 11 below):

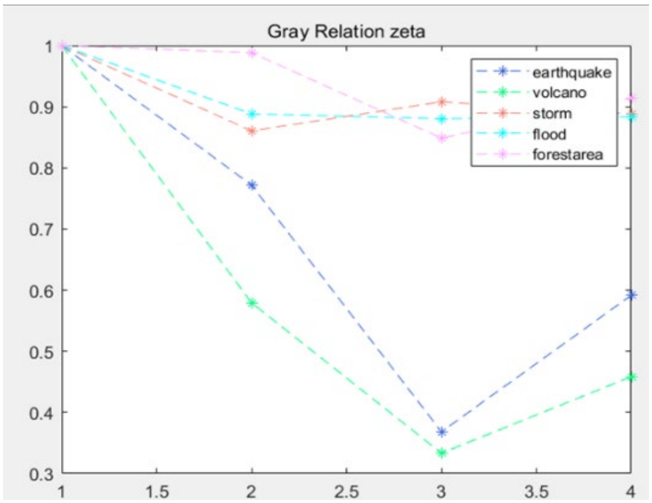

Figure 11: Grey incidence

Through MATLAB code calculation, we can easily find that Forest area is the biggest factor causing global warming, followed by storm, flood, earthquake and volcano. The influencing factor of forest area is 0.9376, and the impact factors of earthquakes, volcanoes, storms and floods are 0.6829, 0.5928, 0.9142 and 0.9131, respectively.

4. Conclusion

The forecast temperature in 2100 is 15.79 °C, an increase of 1.38 °C compared to 2000 temperatures; The temperature growth rate in the 21st century is higher than that of 0.0546 in the 20th century, and the global temperature will warm in the future. Both natural factors (including time, volcanic factors) and human factors (this article lists the forest area as a factor) can affect global warming, and the most influential is forest area.

Acknowledgements

Thanks to all those who contributed to this article; Thanks to all the authors who provided the data in this paper and to Heilongjiang University of Science and Technology for providing the academic forum.

References

[1] Houghton J. Global warming [J]. Reports on progress in physics, 2005, 68(6): 1343.
[2] Drake F. Global warming [M]. Routledge, 2014.
[3] Hansen J, Sato M, Ruedy R, et al. Global warming in the twenty-first century: An alternative scenario [J]. Proceedings of the National Academy of Sciences, 2000, 97(18): 9875-9880.
[4] Aiken L S, West S G, Pitts S C. Multiple linear regression [J]. Handbook of psychology, 2003: 481-507.
[5] Liu S, Yang Y, Cao Y, et al. A summary on the research of GRA models [J]. Grey Systems: Theory and Application, 2013.