Team Number:	2205766
Problem Chosen:	C

2022 APMCM summary sheet

Global warming is a very serious problem facing human society at present. The research and analysis of global temperature and related factors have become an important topic.

In question 1, we first break down the data in the northern and southern hemispheres and divide the data with March 2022 as the boundary. Subsequently, normality analysis was carried out on the two groups of data respectively. On the premise of meeting normality, difference analysis was carried out and the **rank-sum test** was adopted to calculate the significance of the data. Finally, we decided that we could not agree.

When predicting the future temperature, we set up the **grey prediction model** and the **ARIMA prediction model** to predict the future temperature data respectively. Grey prediction models suggest that global temperatures will reach 20°C in 2215. The ARIMA prediction model predicts it will reach 20°C in 2187. Finally, the accuracy of the two models is tested respectively to ensure the accuracy of the model.

In question 2, in order to explore the relationship between global temperature and time and location, we divided the temperature zone by latitude and time by year, and analyzed the correlation between temperature and location, temperature and time respectively. In order to further study their relationship, we built a **linear regression model**, taking location and time as independent variables respectively, and obtained their linear regression equations of global temperature.

In view of the analysis of the impact of natural disasters on temperature, we selected the corresponding events for specific analysis. By building a grey prediction model to predict the theoretical value of global temperature during and after the occurrence of the event, and comparing it with the real value, it can be concluded that forest fires and volcanic eruptions will increase global temperature to a certain extent, while the COVID-19 pandemic has little impact on the temperature. Continuing the above analysis, we believe that the main factor affecting global temperature is carbon emission, and the **correlation analysis** of the two is strong correlation, which confirms the above view.

Finally, the comprehensive determination model has strong generality.

Keywords: grey prediction model ARIMA prediction model linear regression model

Contents

I. The Restatement of the Problem

1.1 Background of the Problem

The average global temperature between *1981* and *1990* was *0.48*℃ higher than *100* years ago. The main cause of global warming is that human beings have been using fossil fuels (such as coal and petroleum) in large quantities in the past century, which has emitted a large amount of CO_2 and other greenhouse gases. These greenhouse gases cause global climate warming. During the *20*th century, the world's average temperature rose by about *0.6* degrees Celsius. Spring thaws in the northern Hemisphere are starting nine days earlier than they did *150* years ago, while autumn frosts are starting about *10* days later.

There are many factors contributing to global climate warming, which can be divided into man-made factors and natural factors. Human factors include population increase, air pollution, forest resource reduction, and natural factors include volcanic activity.

1.2 Presentation of the Problem

In order to further explore the problem of global warming, the question asks to build a mathematical model based on relevant data to study the following problems:

- Analyze temperature data to determine whether the increase in global temperatures in March *2022* has resulted in a greater increase than in the previous *10* years; Build two prediction models to predict future global temperature levels; By predicting the temperature data in *2050* and *2100* through the prediction model, we can find out whether it will reach *20*°C, if not, when it will reach *20*°C. Check the accuracy of the prediction model to ensure its accuracy.
- Analyze the relationship between global temperature, time and location by combining corresponding data; Analyzing the impact of natural disasters (volcanic eruptions, forest fires, COVID-*19*) on temperature; Explore the main factors affecting global temperature; To provide measures to mitigate global warming.

II. Analysis of Problems

2.1 Analysis of Problem One

Problem One requires an analysis of whether the rise in global temperatures in March *2022* has led to a larger increase than in the previous decade. In this regard, we first preprocessed the data, divided the northern and southern hemispheres according to latitude, and analyzed the normality of the two groups of data respectively to ensure that the data set met the normal distribution curve. Then, we conducted the rank sum test to analyze the differences and reached the corresponding conclusions.

For the construction of prediction model. We have established the grey prediction model and the ARIMA prediction model successively, and completed the forecast of the weather data in *2050* and *2100*, and reached the corresponding conclusions.

Finally, the accuracy of the grey model was tested and the error analysis of the ARIMA model was carried out to ensure the accuracy of the prediction model.

2.2 Analysis of Problem Two

In Problem Two, we need to explore the relationship between global temperature and time and location. We further divide the data according to the temperature zone, and use correlation analysis to preliminarily test the relationship between the temperature and the two. In order to further study the specific functional relationship, we establish a linear regression model to obtain the specific relationship between the global temperature and the two.

When analyzing the impact of natural disasters on global temperature, we collected the data of specific volcanic eruptions, forest fires and COVID-*19*. The grey prediction model established in Problem One was used to predict the theoretical temperature at the time of the event, and then compared it with the real data to get the corresponding conclusions.

After the preliminary conclusion that carbon emissions may be the main reason affecting temperature, we collected the global carbon emissions data, conducted correlation analysis between them and the global temperature value, tested the degree of correlation between them, and then reached an accurate conclusion. Finally, we offer solutions to the main causes of global warming.

III. Model Assumptions

1: The global temperature is considered to be the average temperature of the northern and southern hemispheres, ignoring the influence of special factors.

2: In the analysis of location, it is considered that the south temperate zone, the tropical zone and the north temperate zone are the main factors, and the influence of temperature in the two cold zones is ignored.

3: Considers the global temperature data provided to be true and reliable.

IV. Terms, Definitions and Symbols

symbol	Definition
R^2	Degree of curve regression fitting
VIF	Degree of collinearity
df_1	Number of independent variables
df_2	Sample size - (Number of independent variables +1)
P	Value of significance
c	Posterior difference ratio
X_s	Mean annual temperature in the South Temperate zone
X_n	Mean annual temperature in the North Temperate Zone
X_t	Tropical annual mean temperature
X_y	Global average annual temperature

V. Data Processing

5.1 Data Screening

The attached data set contains temperature data for cities and countries from *1743* to *2013*.As the topic of this paper is global warming, various factors affecting global warming need to be taken into account. After the second Industrial Revolution, the emergence of electricity and the wide application of electrical appliances have made great progress in productivity, and it is because of this that the problem of global warming appears. In addition, due to problems such as serious missing and poor continuity of data dating back to a long time ago, based on the above considerations, our team chose the time of the second Industrial Revolution as the turning point, and only selected the temperature data of *1888* and later as the research object, excluding other data.

5.2 Missing value processing

There were some missing months in the data set.Considering that the temperature of adjacent months showed a certain trend of rising or falling, we used the average temperature of adjacent two months to complete the missing values.

VI. Model Building, Solving and Analyzing of Problem 1

6.1 Analysis of Temperature Rise

6.1.1 *Data Preprocessing*

In the relevant analysis of the temperature rise in this problem, considering that the temperature characteristics of different months in the northern and southern hemispheres are quite different, we divide the northern and southern hemispheres by latitude and classify the data for better discussion and analysis.

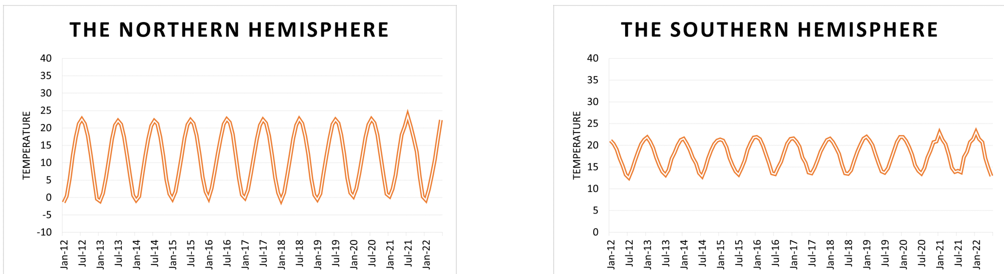


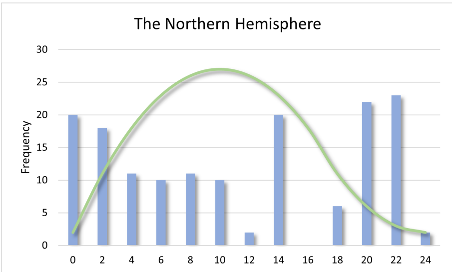
Figure 6.1 Temperature Trend Chart of the Northern and Southern Hemispheres

As can be seen from the temperature trend chart of the northern and southern hemispheres above, the global temperature rise in March *2022* shows no significant difference between the two hemispheres compared with the previous decade. Next, we will analyze the difference of the corresponding data to get a more accurate conclusion.

6.1.2 *Difference Analysis*

In order to more intuitively observe the difference of temperature rise, we separated two sets of data from the northern and southern hemispheres using March *2022* temperature as the boundary, and further analyzed the difference using Mann-Whitney U-tests.

Before conducting Mann-Whitney U-tests, normality analysis of the data set is required to ensure that it basically meets the normal distribution before proceeding to the next step.



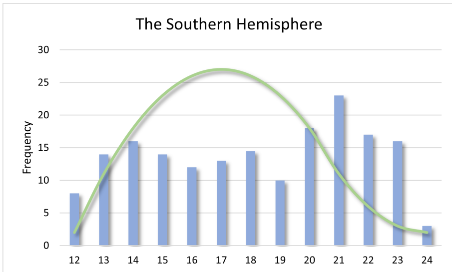


Figure 6.2 Normal distribution curve

The figure above shows the results of data normality test. The normal graph basically presents a bell shape (high in the middle and low at both ends), indicating that although the data is not absolutely normal, it can be basically accepted as normal distribution, so the following steps can be carried out.

In the temperature samples of the Southern Hemisphere, let the data quantity of the two samples be n_1 and n_2 respectively, and let R_1 represent the rank sum of sample *1* and R_2 represent the rank sum of sample *2*[1].

$$U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1 \tag{1}$$

$$U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - R_2$$

At the time of inspection, let:$U = min\,(u_1, u_2)$,when the sample is large

$$\mu_u = mn/2,$$

$$\sigma_\mu = \sqrt{n_1 n_2 (n_1 + n_2 + 1)/12} \tag{2}$$

The sampling distribution of U rapidly approaches the normal distribution:

$$Z = \frac{\left|U - \frac{n_1 n_2}{2}\right|}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{1}}} \tag{3}$$

After obtaining the *Z* value, the significance of mean difference between the two samples can be judged by looking up the table.

Variable name	Variable value	Sample size	Median number	Standard deviation	P	Cohen's d value
Southern Hemisphere Temperature	2012-2022.3	123	18.346	2.969	0.099*	1.014
	2022.3-2022.6	3	14.818	2.066		
	Summation	126	18.206	2.981		
Northern Hemisphere Temperature	2012-2022.3	123	11.987	8.165	0.362	0.662
	2022.3-2022.6	3	16.773	5.694		
	Summation	126	12.01	8.14		

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

Table 6.3 MannWhitney U Test Analysis Results

In the above table, the P values of both hemispheres are greater than *0.05*, so there is no significant difference between the temperature values of March *2022* in the southern Hemisphere and the Northern Hemisphere.

6.1.3 *Conclusion*

According to the above difference analysis, compared with the past decade, the global temperature rise in March *2022* shows no significant difference, so we do not agree with the statement of temperature rise.

6.2 Temperature Forecasting Problem

6.2.1 *Data Preprocessing*

For the temperature Forecasting problem, instead of using monthly data, we integrate the temperature of the northern and southern hemispheres separately and take the average value to obtain the annual average data of the year for subsequent prediction model construction and calculation.

6.2.2 *Construction and Analysis of Grey Prediction Model*

The $GM\ (1,1)$ model is the most widely used in grey prediction, and the prediction of data can often get more accurate results.Before this, it is necessary to conduct grade-ratio test for time series, and only the series that pass the grade-ratio test can be predicted and analyzed by grey model. If it does not pass the stage ratio test, "translation conversion" is performed on the sequence, so that the new sequence meets the stage ratio test. Finally, all the stage ratios must be in the interval $\left(e^{\frac{-2}{n+1}}, e^{\frac{-2}{n+1}}\right)$ to pass the test. Due to the complexity of the data, the results of the initial stage ratio test and the test results after translation and conversion will be shown in the appendix. The grey prediction model will be established below.

With the original data sequence :$X^{(0)}(1)$,$X^{(0)}(2)$,...,$X^{(0)}(n)$, they satisfy $X^{(0)}(k) \geq 0, k = 1, 2, \ldots, n$ The general steps for establishing a $GM\ (1,1)$ model with this data series are:

1) Make first-order accumulation to generate data series

$$X^{(1)}(k) = \sum_{m=1}^{k} X^{(0)}(m), \quad k = 1, 2, \ldots, n \tag{1}$$

2) Determine the data matrix B,Y_n

$$B = \begin{pmatrix} -\frac{1}{2}\left[X^{(1)}(1) + X^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[X^{(1)}(2) + X^{(1)}(3)\right] & 1 \\ \cdots & \cdots \\ \frac{1}{2}\left[X^{(1)}(n-1) + X^{(1)}(n)\right] & 1 \end{pmatrix}, \quad Y_n = \begin{pmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \cdots \\ X^{(0)}(n) \end{pmatrix} \tag{2}$$

3) Find the parameter column

$$[\widehat{au}]^T = \left(B^T B\right)^{-1} B^T Y_n \tag{3}$$

4) Build the model of generating data series

$$\widehat{X^{(1)}}(k+1) = \left(X^{(0)}(1) - \frac{u}{a}\right) \mathrm{e}^{-\widehat{a}k} + \frac{\widehat{u}}{\widehat{a}} \quad k = 1, 2, \ldots$$

5) Establish the original data series model

$$\widehat{X^{(0)}}(1) = X^{(0)}(1)$$

$$\widehat{X^{(0)}}(k) = \widehat{X}^{(1)}(k) - \widehat{X}^{(1)}(k-1)$$

$$= \left(1 - \mathrm{e}^{\widehat{a}}\right)\left(X^{(0)}(1) - \frac{u}{a}\right)\mathrm{e}^{-\widehat{a(k-1)}}, \quad k = 2, 3, \ldots \tag{4}$$

Here,$\widehat{X^{(0)}}\ (k)\ (k = 1, 2, \ldots, n)$ is the original data sequence $\widehat{X^{(0)}}(k)(k = 1, 2, \ldots, n)$ fitting values,$\widehat{X^{(0)}}\ (k)\ (k > n)$ is the predictive value of the original data sequence[2].

Using the grey prediction model to predict the future data, we obtained the future temperature model fitting forecast maps for the northern and southern hemispheres respectively.

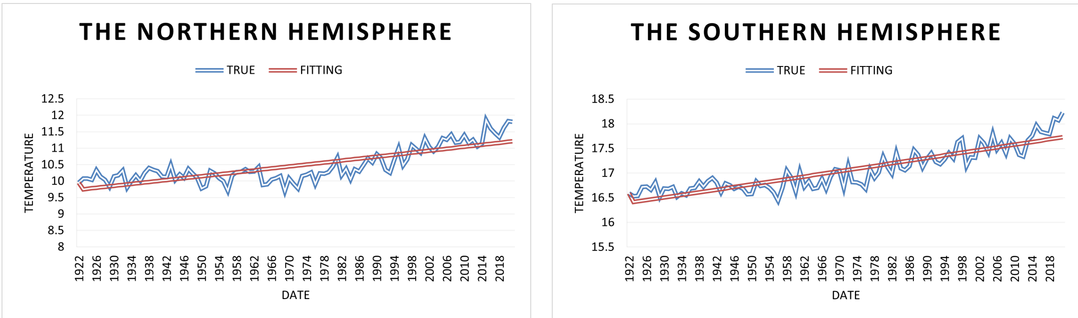


Figure 6.4 Grey prediction model fitting prediction graph

6.2.3 *ARIMA Model Prediction*

The following process is required to predict time series using the ARIMA model.

Firstly, the original time series data is preprocessed, and all of them are zero-valued.

Let the original sequence be the mean of $\{y(t)\}$,the mean of$\{y(t)\}$ $\overline{y} = \frac{1}{n}$.Let the sample value of the new sequence$x\ (t\ )= y\ (t) - \overline{y}$.

The number of differential processing is determined by the parameter d in the $ARIMA\ (p, d, 0)$ model structure:$d = 1$,one differential processing,$d = 2$, two differential processing, and so on. If $d = 0$, that is, no differential processing is carried out, and the model structure is $AR\ (p)$.

If the model of the new sequence obtained after d difference of the original sequence is $AR\ (p)$ model, then the original sequence is $ARIMA\ (p, d, 0)$ model. The parameter estimation and prediction described below are carried out for the $AR\ (p)$ model after differential processing, and the prediction value of the original sequence can be obtained by inverse transformation after the prediction result is obtained.

Then the parameter estimation work is carried out.

The recursive least squares method with forgetting factor was used for parameter estimation. The role of forgetting factor is to strengthen the effect of current observation data on parameter estimation, while weaken the influence of previous observation data. The forgetting factor mainly considers the time variation of model parameters[3].

The mathematical expression of $AR\ (p)\ model of sequence \{y\ (t)\}$ is as follows:

$$A\ (B)\ y\ (t) = e\ (t) \tag{1}$$

Among them, the $e\ (t)$ is the white noise of zero mean,B is the backward operator, and $A\ (B)$ is the same as equation (1). Then equation (2) can be written as:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \ldots + a_y y(t-p) + c(t)$$

Written in vector form:

$$y(t) = \varphi^T(t)\theta + e(t)$$

Among them

$$\varphi^T(t) = [y(t-1), y(t-2), \cdots, y(t-p)]$$

$$\theta = \left[a_1, a_2, \ldots, a_p\right]^T \tag{2}$$

Substitute$\varphi^T(t)$,θinto the formula of recursive least squares method with forgetting factor; With initial values, the online recursive parameter estimation can be carried out.

Finally, to complete the prediction of the model.

Astrom prediction method based on the linear least variance prediction principle is used to predict. This algorithm can solve the problem of large randomness in the prediction. Description form as follows:

For a stationary reversible $ARMA\ (p, q)$ process:

$$A\ (B)\ y\ (t) = C\ (B)\ e\ (t)$$

Where B is backward shift operator,$A\ (B)$ and $C\ (B)$ are the same as equation (1), and the minimum variance predictor is:

$$\hat{y}(t+k \mid t) = (G(B)/C(B))y(t)$$

Its recursive form:

$$C(B)\hat{y}(t+k \mid t) = G(B)y(t)$$

Where G(B) and F(B) can be obtained by the Diophantine equation as follows:

$$C(B) = A(B)F(B) + B^k G(B)$$

In the equation

$$F(B) = \mathrm{f}_0 + \mathrm{f}_1 B + \mathrm{f}_2 B^2 + \ldots + \mathrm{f}_{nj} B^{nf}$$

$$G(B) = \mathrm{g}_0 + \mathrm{g}_1 B + \mathrm{g}_2 B^2 + \ldots + \mathrm{g}_{\mathrm{kg}} B^{ng} \tag{3}$$

$nf = k - 1, ng = max\,(p - 1, q - k)$.This paper is a one-step prediction algorithm. It only needs to substitute $C\,(B) = 1, q = 0$ and the number of prediction steps $k = 1$ into the above formula. After simple derivation and calculation, the one-step prediction formula can be obtained.

Using the ARIMA model, we get future temperature projections for the northern and southern hemispheres.

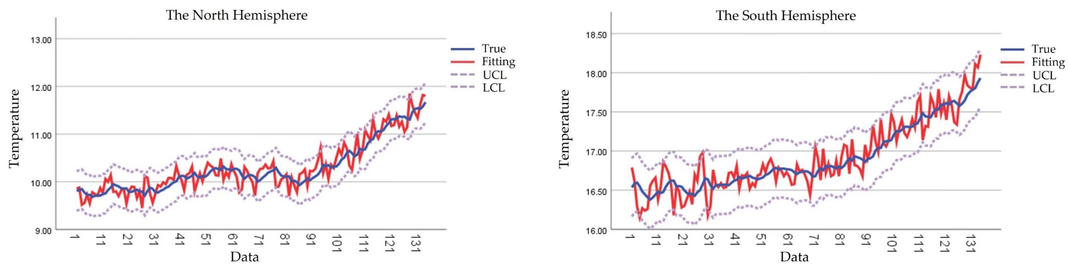


Figure 6.5 ARIMA prediction model fitting prediction diagram

6.3 Future Prediction and Result Analysis

The grey prediction model and ARIMA prediction model are respectively applied, and the following prediction data can be obtained after *2050* and *2100* are substituted into the model.

Model	Grey prediction				ARIMA prediction			
Region	Southern Hemisphere		Northern Hemisphere		Southern Hemisphere		Northern Hemisphere	
Year	2050	2100	2050	2100	2050	2100	2050	2100
Predicted Temperature/° C	18.54	19.64	15.21	17.29	18.71	19.95	15.74	17.3

Table 6.6 Future temperature forecast results

As can be seen from the above table, the annual mean temperature will not reach *20*°C in either *2050* or *2100*, nor in the Southern Hemisphere or the Northern Hemisphere. Therefore, we disagree with this statement.

In order to calculate when the global annual mean temperature will reach *20*°C, we average the northern and southern hemispheres of the two models respectively to better observe the change in global temperature values. Each model predicts the annual mean temperature for the next *500* years. Due to the complexity of the data, the detailed forecast results are presented in the appendix.

Through the prediction of the two models, the grey prediction model shows that the global temperature in *2215* is *20.04*°C, reaching *20*°C for the first time. The ARIMA forecast model says the global temperature will be *20.00*°C in 2187, reaching *20*°C for the first time.

6.4 Accuracy Test

6.4.1 *Accuracy Test of Grey Prediction*

This paper uses the posterior error method to test the accuracy of the grey prediction model $GM\ (1, 1)$. According to the classification standard of the accuracy grade of the grey prediction model, the following classification table is obtained[4].

Model accuracy level	P	c
Level 1 (Excellent)	≥ 0.95	≤ 0.35
Level 2 (Qualified)	$0.8 \leq P \leq 0.95$	$0.35 \leq c \leq 0.5$
Level 3 (barely)	$0.7 \leq P \leq 0.8$	$0.5 \leq c \leq 0.65$
Level 4 (Rejected)	< 0.7	> 0.65

Note:c is the posterior difference ratio,$c = \frac{S_2}{S_1}$,P is the small error probability.

Table 6.7 Standard for Determining Accuracy Level of Model

Then we need to calculate the posterior difference ratio and the small error probability.

First of all, we need to calculate the residual$\varepsilon(k)$.

$$\varepsilon(k) = X^{(0)}(k) - x^{(0)}(k), k = 2, 3, \cdots, n \tag{1}$$

Then calculate the relative error $\Delta(k)$.

$$\Delta(k) = \frac{|\varepsilon(k)|}{X(0)(k)}, k = 2, 3, \cdots, n \tag{2}$$

And then calculate the standard deviation of the original sequence S_1 and the standard deviation of the residual sequence S_2

$$S_1 = \sqrt{\frac{1}{n} \sum_{k=2}^{n} \left[X^{(0)}(k) - X\right]}$$

$$S_2 = \sqrt{\frac{1}{n}\sum_{k=2}^{n}\left[\Delta(k) - \Delta\right]} \tag{3}$$

Then variance ratio C

$$C = \frac{S_2}{S_1}$$

Small probability error P

$$P = P\left\{\left|\Delta(k) - \Delta\right| < 0.6745S_2\right\}, k = 2, 3, \cdots, n \tag{4}$$

The accuracy of model fitting is determined by C and P.

After the model accuracy test of the sample set, it is obtained that the mean residual $\overline{\varepsilon}$=*-19.92*, the variance of residual ε^2=*4333.97*, the posterior difference ratio *c=0.27<0.35*, and the small probability error *p=0.984≥0.95*. The prediction grade of the model is level 1, the prediction results are reliable and the model is relatively accurate.

6.4.2 *ARIMA error analysis*

In order to evaluate and compare the results predicted by the model, the following two indicators are used in this paper.

(1) Relative error, represented by RE, can describe the quality of prediction effect at a certain moment. The calculation formula is as follows:

$$RE\,(t) = \left[y(t) - \hat{y}(t)\right]/y\,(t) \tag{1}$$

Where $y\,(t)$ is the measured value,$\hat{y}(t)$ is the predicted value;

(2) MAPE(mean absolute percentage error), which is an indicator to comprehensively evaluate the prediction performance of the whole prediction process, is calculated as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|RE\,(t)\right| \tag{2}$$

Where $RE(t)$ is the relative error at time t and n is the number of samples.

With MAPE as the evaluation index, we randomly selected *14* samples from the original data set for error test and analysis.

Sample of year	1898	1906	1912	1925	1938	1947	1956	1968	1972	1980	1999	2017
MAPE	5.12%	4.69%	3.47%	4.56%	5.08%	2.87%	1.02%	4.75%	6.29%	3.33%	2.41%	4.98%

Table 6.8 Random sample error analysis

According to the error test results of random samples, the value of MAPE is basically stable below *5*%, indicating that the model prediction simulation results are excellent, has good adaptability, and the model is more accurate.

VII. Model Building, Solving and Analyzing of Problem 2

7.1 The Correlation of Global Temperature, Time and Location

7.1.1 *Data Preprocessing*

In order to better study the correlation between global temperature and regional location, the data provided in the annex are divided into specific regions. According to *23.5*°N and *23.5*°S, the temperature data in the annex are further classified into North temperate zone, tropical zone and South temperate zone. Due to the small number of countries in the northern and southern regions, and the temperature conditions are not suitable for the research objects, they are not considered here. Therefore, it is only divided into three temperature zones, and the annual average temperature of each temperature zone is taken for further analysis and processing.

Regarding the analysis of time, we further sorted and classified the data, and explored the relationship between the data and global temperature change with the year as the basis.

7.1.2 *Correlation Analysis*

The research on correlation analysis between regions and years and global temperature is divided into two studies, and the correlation analysis between regions and temperatures and years and temperatures is conducted twice respectively. Firstly, the Pearson correlation coefficient method is adopted to analyze the degree of pairwise correlation of each temperature zone.

The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and standard deviation between the two variables:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y} \tag{1}$$

The above equation defines the overall correlation coefficient, which is usually represented by the Greek lowercase letter ρ. To estimate the covariance and standard deviation of the sample, Pearson correlation coefficient can be obtained, commonly represented by the lowercase letter $r$in English:

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{X})^2}} \tag{2}$$

$r$can also be estimated by the mean value of the standard fraction of the sample point (X_i, Y_i), and the expression equivalent to the above equation can be obtained:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_X}\right)\left(\frac{Y_i-\bar{Y}}{\sigma_Y}\right) \tag{3}$$

Where $\frac{X_i-\bar{X}}{\sigma_X}$, $\bar{X}$and $\sigma_X$are the standard fraction, sample mean and sample standard deviation of the sample of X_i, respectively[5].

The correlation between air temperature and each temperature zone is calculated and expressed in the form of heat map. The process of correlation analysis between time and temperature is repeated.

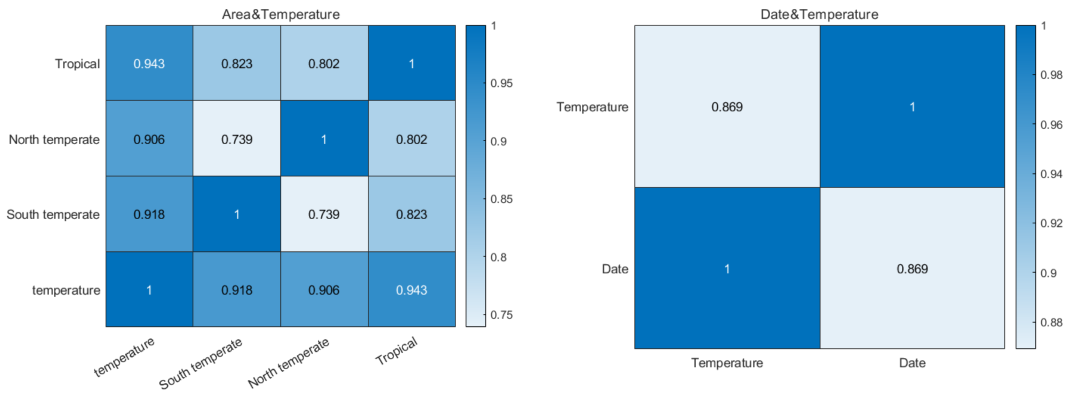


Figure 7.1 Correlation coefficient heat map

It can be intuitively observed from the correlation coefficient heat map that the global temperature is closely related to the region and time.

7.1.3 *Linear Regression Model*

According to the above correlation analysis, the correlation is very close. In order to further explore the specific relationship between temperature and region, temperature and time, linear regression is adopted to further solve the problem.

Let the sample point capacity be n, the dependent variable at time t be y^t, and the p independent variables be x_j^t $(j = 1, 2 \ldots p)$, then the form of the overall linear regression model is[6]:

$$y_i^t = \beta_0 + \beta_1^t x_{i1}^t + \cdots + \beta_p^t x_{ip}^t + \varepsilon_i^t \quad i = 1, 2, \cdots, n \tag{1}$$

$$\text{Let's call } \boldsymbol{y}^t = \begin{pmatrix} y_1^t \\ y_2^t \\ \vdots \\ y_n^t \end{pmatrix}_{n\times 1} \qquad \boldsymbol{\beta}^t = \begin{pmatrix} \beta_0^t \\ \beta_1^t \\ \vdots \\ \beta_p^t \end{pmatrix}_{(p+1)\times 1}$$

$$\boldsymbol{X}^t = \begin{pmatrix} 1 & x_{11}^t & x_{12}^t & \cdots & x_{1p}^t \\ 1 & x_{21}^t & x_{22}^t & \cdots & x_{2p}^t \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1}^t & x_{n2}^t & \cdots & x_{np}^t \end{pmatrix}_{n\times(p+1)}$$

Then the cross product V^t of the augmented matrix $\left(X^t, y^t\right)$ is

$$\boldsymbol{V}^t = \begin{pmatrix} \left(\boldsymbol{X}^t\right)' \boldsymbol{X}^t & \left(\boldsymbol{X}^t\right)' \boldsymbol{y}^t \\ \left(\boldsymbol{y}^t\right)' \boldsymbol{X}^t & \left(\boldsymbol{y}^t\right)' \boldsymbol{y}^t \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}_{11}^t & \boldsymbol{V}_{12}^t \\ \boldsymbol{V}_{21}^t & \boldsymbol{V}_{22}^t \end{pmatrix} \tag{2}$$

The least squares estimator of regression coefficient β^t is

$$\widehat{\boldsymbol{\beta}^t} = \left(\boldsymbol{V}_{11}^t\right)^{-1} \boldsymbol{V}_{12}^t \tag{3}$$

It can also be proved that the values of $\boldsymbol{S}_{SSE}$ and $\boldsymbol{S}_{SST}$ at time t are shown in Equation (11) and Equation (12) respectively.

$$\boldsymbol{S}_{SSE}^t = \boldsymbol{V}_{22}^t - \boldsymbol{V}_{21}^t \left(\boldsymbol{V}_{11}^t\right)^{-1} \boldsymbol{V}_{12}^t \tag{4}$$

$$\boldsymbol{S}_{SST}^t = \boldsymbol{V}_{22}^t - \frac{1}{n}\left(\boldsymbol{V}_{12,1}^t\right)^2 \tag{5}$$

Where $V_{12,1}^t$ is the first component of V_{12}^t. According to equations (11) and (12), the goodness of fit (adjusted complex measurement coefficient) at time t is

$$\bar{R}_t^2 = 1 - \frac{S_{SSE'}^t(n-p-1)}{S_{SST}^t{}'(n-1)} \tag{6}$$

Meanwhile, the estimated standard error at time t is

$$S_e^t = \sqrt{\frac{1}{n-p-1} S_{SSE}^t} \tag{7}$$

The specific modeling method of linear regression model is as follows:

1) Calculate t $(t = 1, 2.....T)$ Cross product V^t of the augmented matrix $\left(X^t y^t\right))$ at time;

2) Calculate the eigenvalue of V^t $\lambda_1^t \geq \lambda_2^t \geq \ldots \geq \lambda_{p+2}^t \geq 0$ and the corresponding orthonormal eigenvector $u_1^t, u_2^t, \ldots, u_{p+2}^t$ $(t = 1, 2, \ldots, T)$;

3) According to the eigenvector matrix of the cross product matrix V^t at the time $1 \sim T$, the orthogonal matrix prediction method is used to predict the eigenvector matrix at the time $T + l$ $\left(u_1^{T+l} u_2^{T+l} {}_{p+2}^{T+l}\right)$;

4) According to the eigenvalues of the cross product matrixV^t at the time $1 \sim T$, the time series analysis method is applied to predict the eigenvalues at time $T+l$ respectively $\lambda_1^{T+l} \lambda_2^{T+l} \ldots \lambda_{p+2}^{T+l}$;

5) According to the calculation results of the last two steps, the cross product matrix V^{T+l} at time $T + l$ is calculated as

$$\boldsymbol{V}^{T+l} = \begin{pmatrix} \boldsymbol{u}_1^{T+l} & \boldsymbol{u}_2^{T+l} & \cdots & \boldsymbol{u}_{p+2}^{T+l} \end{pmatrix}$$
$$\begin{pmatrix} \lambda_1^{T+l} & 0 & 0 & 0 \\ 0 & \lambda_2^{T+l} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{p+2}^{T+l} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{u}_1^{T+l} & \boldsymbol{u}_2^{T+l} & \cdots & \boldsymbol{u}_{p+2}^{T+l} \end{pmatrix}'$$

6) According to Equation (3), the estimator $\widehat{\beta}^{T+l}$ of the regression parameter at time $T + l$ can be obtained;

7) According to equations (4) $\sim$ (7), the goodness of fit R_{T+l}^2 and the estimated standard error S_e^{T+l} of the regression model at time T+I can be calculated.

For the relationship between region and temperature, the following results are obtained according to the above process. The above table shows the analysis results of this model, including the standardized coefficient of the model, t value, VIF value, R^2, adjustmentR^2, etc., for the test of the model and the analysis of the formula of the model. F test is to judge whether there is a significant linear relationship, and R^2 is to judge whether the regression line fits the linear model. In linear regression, the main concern is whether the F test is passed, and in some cases, the size of R^2 is not necessarily related to the model explicity. According to the analysis of the results of F test, the significance

Linear regression analysis results n=114						
Non normalized coefficient		normalized coefficient		P	VIF	R^2
B	Standard error	Beta				
constant	0	0	-	0.321	-	1
Xs	0.333	0	0.332	0.000***	4.051	
Xn	0.333	0	0.408	0.000***	3.935	
Xt	0.333	0	0.32	0.000***	4.927	
Dependent variable: AVE						

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

Table 7.2 Linear regression results analysis table

P value is *0.000****, showing significance at the level, rejecting the null hypothesis that the regression coefficient is 0, so the model basically meets the requirements.

For the collinearity of variables, VIF is all less than *10*, so the model has no multicollinearity problem and the model is well constructed.

The formula of the model is as follows: $y = 0.0 + 0.333 \times X_s + 0.333 \times X_n + 0.333 \times X_t$. For the relationship between year and temperature, the following contents can be obtained by repeating the above process. The formula of the model is as follows:

Linear regression analysis results n=114						
Non normalized coefficient		normalized coefficient		P	VIF	R^2
B	Standard error	Beta				
constant	-0.36	1.011	-	0.723	-	0.747
Xy	0.009	0.001	0.864	0.000***	1	
Dependent variable: AVE						

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

Table 7.3 Linear regression results analysis table

$y = -0.36 + 0.009 \times X_y$.

7.1.4 *Conclusion and Analysis*

Through the above process, it can be seen that global temperature is strongly correlated with regions and years, and the specific relationship can be solved by linear regression model:

Global Temperature and Region: $y = 0.0 + 0.333 \times X_s + 0.333 \times X_n + 0.333 \times X_t$

Global temperature and Year: $y = -0.36 + 0.009 \times X_y$

7.2 The Influence of Other Factors on Temperature

7.2.1 *Data Preprocessing*

To further explore the impact of volcanic eruptions, forest fires and COVID-*19* on global temperatures, we need to separate these three periods into appropriate time nodes.

For volcanic eruptions, we selected the impact and the larger eruption of Mount Pinatubo in the Philippines to explore its impact on global temperatures. The eruption of Mount Pinatubo occurred in *1991*, and we use the *1991* and *1992* global temperature data as a boundary to explore further with the temperature data of the last *10* years.

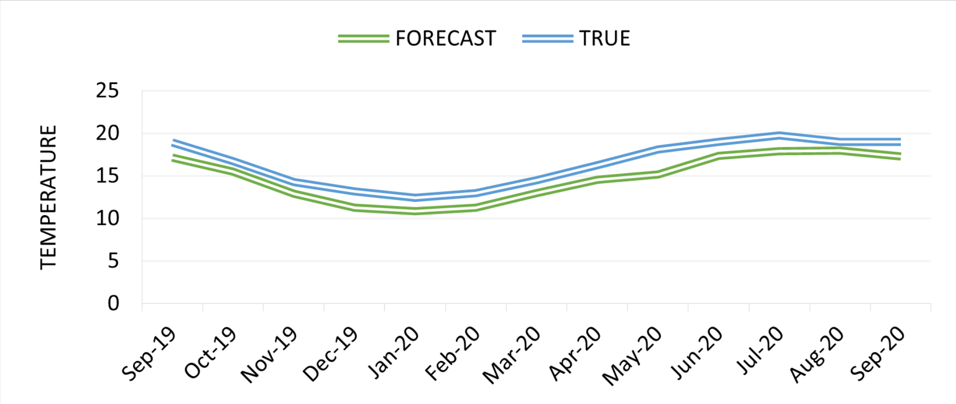
As for forest fires, the long duration of bushfires in Australia has caused a very bad impact on the ecology. The bushfires in Australia have been extended from September *2019* to January *2020*. Taking this as the time node, we select the global temperature data during the bushfires and the following *8* months, a total of *12* months, as well as the data of *12* months before the bushfires for further exploration.

For the COVID-*19* outbreak from December *2019* to now, we selected the global average temperature in *2020* and *2021*, as well as the temperature data from *2013* to *2019* before the outbreak.

7.2.2 *Grey Prediction Model*

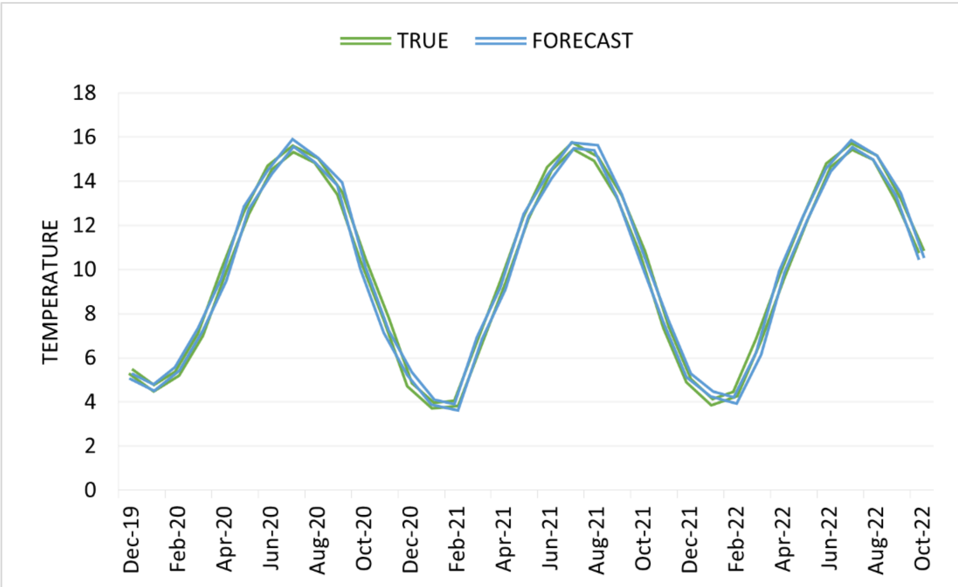
In order to test whether the above three events had an impact on the global temperature during the period, we used the grey prediction model in *6.2.2*, used the original data set before the event to predict the theoretical temperature at the time of the event, and then compared the real temperature data with it to observe the impact degree of the corresponding events.

The specific theoretical prediction results and the real values are compared as follows:



	Sep-19	Oct-19	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20
FORECAST	17.1	15.51	12.93	11.24	10.82	11.23	12.98	14.56	15.13	17.35	17.86	17.97	17.27
TRUE	18.93	16.75	14.29	13.19	12.43	12.99	14.46	16.27	18.07	18.99	19.73	19	19.01

Figure 7.4 Comparison of monthly bushfire temperatures in Australia



Note:Due to the large number of selected months for COVID-*19*, the comparison between the specific theoretical data and the real data will be shown in the appendix.

Figure 7.5 A comparison of monthly COVID-19 temperature data

The data selected for volcanic eruption are in the dimension of year, so the data amount is small. Here, the table is used to display the data intuitively.

7.2.3 *Results and Analysis*

From the comparison between the theoretical forecast data and the real data, it can be seen that the Australian bushfire and the eruption of Mount Pinatubo in the Philippine

Year	Prediction of theory	The actual
1991	14.02	14.05
1992	14.1	14.18

Table 7.6 Comparison table of annual eruption temperature data of Mount Pinatubo

Islands both had a certain impact on the temperature conditions within a certain time range, resulting in a small increase in global temperature. The outbreak of the novel coronavirus has lowered global temperatures within a certain observation range, but the decrease is negligible. Therefore, we believe that the factors that have a certain influence on global temperature include stronger volcanic eruptions and long duration forest fires.

7.3 Analysis of Main Causes of Temperature Change

According to the above analysis in *7.2*, both volcanic eruptions and forest fires can cause a small increase in global temperature to some extent, while both can cause a sharp increase in carbon emissions. Therefore, we believe that carbon emissions are the main cause affecting global temperature. To do this, we collected nearly *70* years of global annual carbon emissions, and again correlated them with global temperatures to observe the results.

	PPM	Temperature
PPM	1(0.000***)	0.947(0.000***)
Degrees C	0.947(0.000***)	1(0.000***)

Note: ***, ** and * represent the significance level of 1%, 5% and 10% respectively

Table 7.7 Significant value analysis table

As can be seen from the above table and the thermal map of correlation coefficient, the correlation between carbon emission and global temperature is very high, which just confirms the above conclusion. Therefore, we can infer that carbon emission is the main factor causing global temperature change.

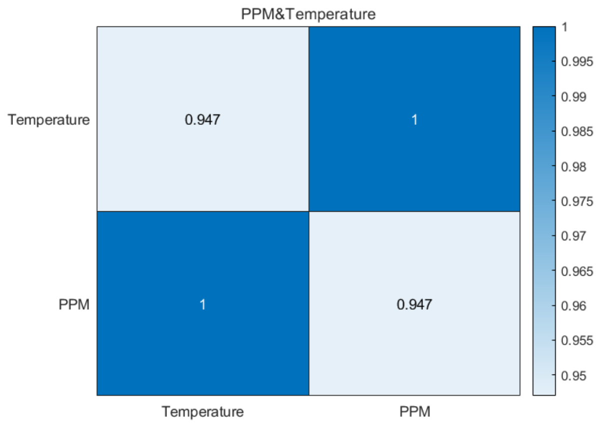


Figure 7.8 Correlation coefficient heat map

7.4 Solution

The above situation shows that carbon emission is the main factor affecting global warming. Therefore, we propose the following countermeasures:

1. Strictly strengthen the protection of ecological environment, and deal with natural disasters in time to avoid tragedies.

2. Reduce the use of fossil fuels, continue to promote the popularization of new energy, so that every country can reduce carbon emissions.

3. Through exchanges and cooperation between countries, research and development of advanced energy-saving technologies, reduce the emission of harmful gases, reduce the harm of pollution to the atmosphere environment. In accordance with the principle of "common but differentiated responsibilities" set out in the Convention, countries have earnestly strengthened cooperation among themselves on climate change mitigation and related issues[7].

VIII. Evaluation of Model

8.1 The Advantages of Model and Paper

Reasonable assumptions: Based on extensive literature reading, this paper establishes a series of scientific assumptions through in-depth research and analysis of the problem, ignores some factors that have little impact on the results, simplifies the model and algorithm, and improves the operational efficiency of the model.

Result visualization: In this paper, MATLAB software was used to visualize the operation and related change rules of the relevant model, making the operation results of the model more intuitive.

Model generality: The classification model and correspondence analysis model established in the process of problem solving can be changed with different sample sets to adapt to different actual situations, and the model has good generality.

8.2 Disadvantages and Deficiencies of the Model

1. For some temperature data divided according to different locations, the method is directly processed by adding and averaging, without further collecting real data for verification, so there are some loopholes in this method.

2. The established model has not been simulated and verified, which has certain limitations.

IX. Memorandum

Dear Sir:

The problem of global warming is becoming more and more serious. Our team has carried out specific analysis and research based on the data provided, and obtained some valuable conclusions.

i.First, we do not agree that the rise in global temperatures in March 2022 is responsible for a larger increase than in the previous decade, based on a differential analysis of the data. There was no significant difference in temperature increase between the months and years.

ii. Secondly, after building the prediction model, we predicted the global temperature in 2050 and 2100, and found that it could not reach 20°C. Therefore, we do not agree that the global temperature will reach 20°C in 2050 or 2100. After further calculation through the two prediction models we have established, we conclude that this time node should be 2187 or 2215 years. And our prediction model has excellent accuracy.

iii. Global temperature is strongly correlated with time and location. After corresponding linear regression analysis, we establish linear equations of temperature and location and temperature and time respectively, which is helpful to further judge their relationship.

iV. After studying the impact of natural disasters on global temperature, we concluded that both forest fires and volcanic eruptions can cause a small increase in global temperature, while the impact of COVID-19 on global temperature is not significant. Here we believe that the main reason behind it is carbon emissions, so we conducted correlation analysis to confirm our conclusions and provide some measures to mitigate global warming.

The above are the findings of our team after research. I hope the models, analysis and conclusions mentioned above will be helpful to your work or research, and I also hope that human society can solve the problem of global warming as soon as possible.

Sincerely,

Team apmcm2205766

X. References

[1] Yang Weikang, QIAO Jianfang, GAO Xingyi, ZHONG Wenqin. Application of rank-sum test in habitat selection analysis of birds [J]. Research in thousands of Arid Regions,2001(03):42-44.

[2] Deng Julong. Grey Prediction and Grey Decision making [M]. Wuhan: Huazhong University of Science and Technology Press, 2002.

[3] Han Chao, SONG Su, WANG Chenghongjizi Real-time Adaptive Prediction of Short-term Traffic Flow Based on ARIMA Model [J]. Journal of System Injury,2004,(07):1530-1532+1535.

[4] Lu Yu, Zhang Li, ZHANG Fengdeng. Fault Tolerant Clock Synchronization Algorithm Based on Grey Prediction [J]. Electronic Science and Technology :1-8.

[5] PiantadosiJ.; Howlett,P.;BolandJ. (2007)“Matching the grade correlation coef ficient using a copula with maximum disorder ",Journal of lndustrial andManagement Optimization3 (2),305-312PiantadosiJ.?;HowlettP.?; Boland,J.(2007)"Matching the grade correlation coefficient using a copula with maximum disorder "Journal of Industrial and Management Optimization3 (2)305-312

[6] Wang Huiwen, MENG Jie. Prediction modeling method of Multiple Linear Regression [J]. Journal of Beijing Journal of Aeronautics and Astronsutics, 2007, (04):500-504.

[7] Hui Jingqi. Global Warming and International Response [J]. Fujian Quality Management,2019(1):61. (in Chinese) DOI :10.3969/ j.issn.1673-9604.2019.o1.046.

I. Appendix

1.1 Global Arima

Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature
2022	14.872219	2059	16.0216545	2096	17.17108971	2133	18.32052493	2170	19.46996015
2023	14.903285	2060	16.05272031	2097	17.20215553	2134	18.35159075	2171	19.50102596
2024	14.934351	2061	16.08378613	2098	17.23322135	2135	18.38265656	2172	19.53209178
2025	14.965417	2062	16.11485195	2099	17.26428716	2136	18.41372238	2173	19.5631576
2026	14.996483	2063	16.14591776	2100	17.29535298	2137	18.4447882	2174	19.59422341
2027	15.027548	2064	16.17698358	2101	17.3264188	2138	18.47585401	2175	19.62528923
2028	15.058614	2065	16.2080494	2102	17.35748461	2139	18.50691983	2176	19.65635505
2029	15.089680	2066	16.23911521	2103	17.38855043	2140	18.53798565	2177	19.68742086
2030	15.120746	2067	16.27018103	2104	17.41961625	2141	18.56905146	2178	19.71848668
2031	15.151812	2068	16.30124685	2105	17.45068206	2142	18.60011728	2179	19.7495525
2032	15.182877	2069	16.33231266	2106	17.48174788	2143	18.6311831	2180	19.78061831
2033	15.213943	2070	16.36337848	2107	17.5128137	2144	18.66224891	2181	19.81168413
2034	15.245009	2071	16.3944443	2108	17.54387951	2145	18.69331473	2182	19.84274995
2035	15.276075	2072	16.42551011	2109	17.57494533	2146	18.72438055	2183	19.87381576
2036	15.307141	2073	16.45657593	2110	17.60601115	2147	18.75544636	2184	19.90488158
2037	15.338207	2074	16.48764175	2111	17.63707696	2148	18.78651218	2185	19.9359474
2038	15.369272	2075	16.51870756	2112	17.66814278	2149	18.817578	2186	19.96701321
2039	15.400338	2076	16.54977338	2113	17.6992086	2150	18.84864381	2187	20.00156728
2040	15.431404	2077	16.5808392	2114	17.73027441	2151	18.87970963	2188	20.02914485
2041	15.462470	2078	16.61190501	2115	17.76134023	2152	18.91077545	2189	20.06021066
2042	15.493536	2079	16.64297083	2116	17.79240605	2153	18.94184126	2190	20.09127648
2043	15.524601	2080	16.67403665	2117	17.82347186	2154	18.97290708	2191	20.1223423
2044	15.555667	2081	16.70510246	2118	17.85453768	2155	19.0039729	2192	20.15340811
2045	15.586733	2082	16.73616828	2119	17.8856035	2156	19.03503871	2193	20.18447393
2046	15.617799	2083	16.7672341	2120	17.91666931	2157	19.06610453	2194	20.21553975
2047	15.648865	2084	16.79829991	2121	17.94773513	2158	19.09717035	2195	20.24660556
2048	15.679931	2085	16.82936573	2122	17.97880095	2159	19.12823616	2196	20.27767138
2049	15.710996	2086	16.86043155	2123	18.00986676	2160	19.15930198	2197	20.3087372
2050	15.742062	2087	16.89149736	2124	18.04093258	2161	19.1903678	2198	20.33980301
2051	15.773128	2088	16.92256318	2125	18.0719984	2162	19.22143361	2199	20.37086883
2052	15.804194	2089	16.953629	2126	18.10306421	2163	19.25249943	2200	20.40193465
2053	15.835260	2090	16.98469481	2127	18.13413003	2164	19.28356525	2201	20.43300046
2054	15.866325	2091	17.01576063	2128	18.16519585	2165	19.31463106	2202	20.46406628
2055	15.897391	2092	17.04682645	2129	18.19626166	2166	19.34569688	2203	20.4951321
2056	15.928457	2093	17.07789226	2130	18.22732748	2167	19.3767627	2204	20.52619791
2057	15.959523	2094	17.10895808	2131	18.2583933	2168	19.40782851	2205	20.55726373
2058	15.990589	2095	17.1400239	2132	18.28945911	2169	19.43889433	2206	20.58832954

Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature
2207	20.61939536	2244	21.76883058	2281	22.91826579	2318	24.06770101	2355	25.21713623
2208	20.65046118	2245	21.79989639	2282	22.94933161	2319	24.09876683	2356	25.24820204
2209	20.68152699	2246	21.83096221	2283	22.98039743	2320	24.12983264	2357	25.27926786
2210	20.71259281	2247	21.86202803	2284	23.01146324	2321	24.16089846	2358	25.31033368
2211	20.74365863	2248	21.89309384	2285	23.04252906	2322	24.19196428	2359	25.34139949
2212	20.77472444	2249	21.92415966	2286	23.07359488	2323	24.22303009	2360	25.37246531
2213	20.80579026	2250	21.95522548	2287	23.10466069	2324	24.25409591	2361	25.40353113
2214	20.83685608	2251	21.98629129	2288	23.13572651	2325	24.28516173	2362	25.43459694
2215	20.86792189	2252	22.01735711	2289	23.16679233	2326	24.31622754	2363	25.46566276
2216	20.89898771	2253	22.04842293	2290	23.19785814	2327	24.34729336	2364	25.49672858
2217	20.93005353	2254	22.07948874	2291	23.22892396	2328	24.37835918	2365	25.52779439
2218	20.96111934	2255	22.11055456	2292	23.25998978	2329	24.40942499	2366	25.55886021
2219	20.99218516	2256	22.14162038	2293	23.29105559	2330	24.44049081	2367	25.58992603
2220	21.02325098	2257	22.17268619	2294	23.32212141	2331	24.47155663	2368	25.62099184
2221	21.05431679	2258	22.20375201	2295	23.35318723	2332	24.50262244	2369	25.65205766
2222	21.08538261	2259	22.23481783	2296	23.38425304	2333	24.53368826	2370	25.68312348
2223	21.11644843	2260	22.26588364	2297	23.41531886	2334	24.56475408	2371	25.71418929
2224	21.14751424	2261	22.29694946	2298	23.44638468	2335	24.59581989	2372	25.74525511
2225	21.17858006	2262	22.32801528	2299	23.47745049	2336	24.62688571	2373	25.77632093
2226	21.20964588	2263	22.35908109	2300	23.50851631	2337	24.65795153	2374	25.80738674
2227	21.24071169	2264	22.39014691	2301	23.53958213	2338	24.68901734	2375	25.83845256
2228	21.27177751	2265	22.42121273	2302	23.57064794	2339	24.72008316	2376	25.86951838
2229	21.30284333	2266	22.45227854	2303	23.60171376	2340	24.75114898	2377	25.90058419
2230	21.33390914	2267	22.48334436	2304	23.63277958	2341	24.78221479	2378	25.93165001
2231	21.36497496	2268	22.51441018	2305	23.66384539	2342	24.81328061	2379	25.96271583
2232	21.39604078	2269	22.54547599	2306	23.69491121	2343	24.84434643	2380	25.99378164
2233	21.42710659	2270	22.57654181	2307	23.72597703	2344	24.87541224	2381	26.02484746
2234	21.45817241	2271	22.60760763	2308	23.75704284	2345	24.90647806	2382	26.05591328
2235	21.48923823	2272	22.63867344	2309	23.78810866	2346	24.93754388	2383	26.08697909
2236	21.52030404	2273	22.66973926	2310	23.81917448	2347	24.96860969	2384	26.11804491
2237	21.55136986	2274	22.70080508	2311	23.85024029	2348	24.99967551	2385	26.14911073
2238	21.58243568	2275	22.73187089	2312	23.88130611	2349	25.03074133	2386	26.18017654
2239	21.61350149	2276	22.76293671	2313	23.91237193	2350	25.06180714	2387	26.21124236
2240	21.64456731	2277	22.79400253	2314	23.94343774	2351	25.09287296	2388	26.24230818
2241	21.67563313	2278	22.82506834	2315	23.97450356	2352	25.12393878	2389	26.27337399
2242	21.70669894	2279	22.85613416	2316	24.00556938	2353	25.15500459	2390	26.30443981
2243	21.73776476	2280	22.88719998	2317	24.03663519	2354	25.18607041	2391	26.33550563

Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature
2392	26.366571	2399	26.58403216	2406	26.80149288	2413	27.01895359	2419	27.20534849
2393	26.397637	2400	26.61509798	2407	26.83255869	2414	27.05001941	2420	27.23641431
2394	26.428703	2401	26.64616379	2408	26.86362451	2415	27.08108523	2421	27.26748013
2395	26.459769	2402	26.67722961	2409	26.89469033	2416	27.11215104	2422	27.29854594
2396	26.490835	2403	26.70829543	2410	26.92575614	2417	27.14321686	2423	27.32961176
2397	26.521901	2404	26.73936124	2411	26.95682196	2418	27.17428268	2424	27.36067758
2398	26.552966	2405	26.77042706	2412	26.98788778				

1.2 Global Grey Forecasting

Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature	Date	Temperature
2022	14.818210	2059	15.73212038	2096	16.77570991	2133	17.82976521	2170	18.79931006
2023	14.839490	2060	15.76559559	2097	16.79101885	2134	17.84702646	2171	18.83000642
2024	14.876312	2061	15.7859768	2098	16.82612881	2135	17.87271683	2172	18.84753156
2025	14.899249	2062	15.80674107	2099	16.85197099	2136	17.89752145	2173	18.87484126
2026	14.926935	2063	15.82794747	2100	16.8754438	2137	17.93631027	2174	18.91288372
2027	14.946740	2064	15.86483698	2101	16.9141877	2138	17.96895452	2175	18.93056434
2028	14.982580	2065	15.88125902	2102	16.9370367	2139	18.00018688	2176	18.94661068
2029	15.008811	2066	15.91312746	2103	16.96822849	2140	18.02916069	2177	18.96397943
2030	15.024678	2067	15.95406244	2104	17.00544197	2141	18.0593171	2178	18.97656277
2031	15.060330	2068	15.99365739	2105	17.03049461	2142	18.08669333	2179	19.00556285
2032	15.074284	2069	16.01494594	2106	17.05908108	2143	18.12090016	2180	19.03978708
2033	15.115462	2070	16.05586846	2107	17.07986046	2144	18.14558532	2181	19.07720904
2034	15.134112	2071	16.09435718	2108	17.11735589	2145	18.16084997	2182	19.09721771
2035	15.173147	2072	16.12663949	2109	17.14729222	2146	18.1912005	2183	19.11081976
2036	15.205011	2073	16.15380218	2110	17.1811757	2147	18.20647239	2184	19.1239922
2037	15.217153	2074	16.17503557	2111	17.20859206	2148	18.21926924	2185	19.15459936
2038	15.234261	2075	16.19385766	2112	17.22585338	2149	18.25228822	2186	19.18446677
2039	15.246358	2076	16.22378685	2113	17.25544662	2150	18.2684802	2187	19.2119424
2040	15.283531	2077	16.23604927	2114	17.27449944	2151	18.28177215	2188	19.23693363
2041	15.298993	2078	16.25531237	2115	17.3081237	2152	18.31498039	2189	19.27725766
2042	15.315066	2079	16.27723024	2116	17.34952145	2153	18.34575915	2190	19.3187775
2043	15.332718	2080	16.31833186	2117	17.38244896	2154	18.38250633	2191	19.34343398
2044	15.354189	2081	16.34399201	2118	17.41652837	2155	18.40081045	2192	19.3665654
2045	15.378984	2082	16.37863878	2119	17.45377365	2156	18.44089775	2193	19.38857491
2046	15.400462	2083	16.40212225	2120	17.46799788	2157	18.45826678	2194	19.40743372
2047	15.415143	2084	16.43635101	2121	17.49880506	2158	18.48230172	2195	19.44418648
2048	15.433399	2085	16.46521366	2122	17.51280667	2159	18.49461598	2196	19.47318225
2049	15.452294	2086	16.48762039	2123	17.53162503	2160	18.50986315	2197	19.49302951
2050	15.467463	2087	16.5017173	2124	17.56144394	2161	18.53451941	2198	19.52596276
2051	15.500004	2088	16.5346221	2125	17.57478445	2162	18.55820938	2199	19.54233117
2052	15.541283	2089	16.57142875	2126	17.6133129	2163	18.59061259	2200	19.57400491
2053	15.572919	2090	16.60765379	2127	17.64167267	2164	18.61361305	2201	19.59873728
2054	15.610009	2091	16.6403753	2128	17.67594243	2165	18.64867825	2202	19.63679319
2055	15.627691	2092	16.67219512	2129	17.71219061	2166	18.68734057	2203	19.67845017
2056	15.669687	2093	16.70497566	2130	17.72682539	2167	18.72349275	2204	19.71573953
2057	15.689038	2094	16.73170995	2131	17.76649212	2168	18.7357309	2205	19.73317246
2058	15.719604	2095	16.75342093	2132	17.80097032	2169	18.77674762	2206	19.76824616
2207	19.784298	2208	19.80551968	2209	19.84458809	2210	19.8611617	2211	19.88124088
2212	19.896980	2213	19.91486486	2214	19.93783419	2215	20.04521		

1.3 Epidemic prediction

DATE	TRUE	FORECAST	DATE	TRUE	FORECAST
Dec-19	5.393	5.181757	Jun-21	14.577	14.206730
Jan-20	4.619	4.653073	Jul-21	15.614	15.632419
Feb-20	5.291	5.502031	Aug-21	15.026	15.518338
Mar-20	7.045	7.281486	Sep-21	13.312	13.369185
Apr-20	9.963	9.533822	Oct-21	10.783	10.460047
May-20	12.575	12.818327	Nov-21	7.386	7.650799
Jun-20	14.622	14.419193	Dec-21	4.984	5.212365
Jul-20	15.46	15.736430	Jan-22	3.989	4.377662
Aug-20	14.934	14.967673	Feb-22	4.358	4.071753
Sep-20	13.464	13.885336	Mar-22	6.812	6.188469
Oct-20	10.449	10.061181	Apr-22	9.665	9.915674
Nov-20	7.763	7.183884	May-22	12.313	12.321522
Dec-20	4.785	5.301810	Jun-22	14.742	14.525337
Jan-21	3.843	3.995662	Jul-22	15.588	15.713146
Feb-21	3.953	3.769487	Aug-22	15.07	15.080825
Mar-21	6.684	6.906457	Sep-22	13.151	13.407939
Apr-21	9.419	9.136778	Oct-22	10.811	10.490445
May-21	12.346	12.451534			