Team Number :	apmcm2202163
Problem Chosen :	C

2022 APMCM summary sheet

The dreaded heat is hurting people, and whether temperatures will continue to rise or not, the change of the global average temperature is closely bound up with everyone. It was reported that the highest temperature record has been broken in many places this year.

According to the requirements of the topic, our group designed several mathematical models to predict the change of global temperature. We used the **Independent sample t test** and the **independent sample manu test (specifically combined with normality test)**, Dividing the temperature data into *March 2010 to March 2022* and *March to October 2022*, we find that March 2022 does not show the largest temperature increase in the past 12 years, and there was no significant difference between the two sets of data. Our team still uses the **grey prediction model**, **Lasso prediction model** and **Value prediction** model to forecast the global average temperature, and we find that the global average temperature will not rise to 20℃ in 2050 or 2100, but the average temperature of the earth will rise to 20℃ in 3767. Our team uses **grey relation analysis** to find that the time and geographical location are related to temperature, in which latitude has the greatest influence on temperature, followed by time, city, and finally longitude. Through sensitivity analysis, we adjust the value of ρ from 0.6 to 0.7, and we find that the absolute result will be larger. But the relative error is less than 1% and the relative influence size is unchanged. Through the **linear prediction model**, we found that there was a negative correlation between volcanic eruption and temperature, and volcanic eruption would make the temperature drop. Taking the volcanic eruption in Tonga as an example, we found that the temperature dropped about 0.0315°C - 0.1118°C within 1-2 years after the eruption. The greenhouse gas effect is the main reason for the influence on temperature, because the atmosphere is mainly composed of greenhouse gases, and the emission of greenhouse gases increases exponentially. In recent years, the greenhouse gas emission index also increases linearly, increasing by about 0.1 every ten years, which is not optimistic.

Finally, our group analyzed the possible plan to reduce greenhouse gas emissions according to the existing data and the data found and wrote a letter to the organizing committee, which roughly described our findings and made a sincere appeal to human beings to protect the earth.

Keywords: Independent sample t test,
Independent Sample MANU test, grey prediction model,
Lasso prediction model,
Value prediction, grey relation analysis, linear prediction

model,
Temperature, Global Temperature, Prediction

Contents

1. Introduction

1.1 Background

It is hard to exaggerate how global warming effects the earth. According to China Daily, Shanghai heat wave sets new record. Over the past 150 years in total, Shanghai has only had 20 days in which temperatures exceeded 40 ℃.

Before 2009, this had only happened twice in 1934. China's national observatory on Thursday issued an orange alert for high temperatures as an intense heat wave lingers in multiple regions of the country. China has a four-tier, color-coded weather warning system, with red representing the most severe warning, followed by orange, yellow and blue. In Nanjing of Jiangsu province, the city's emergency center said there were 172 cases of heatstroke from Aug 1 to 9, with most cases happening around 2 pm. The center advised the public to avoid outdoor activities and suggested construction workers, couriers and nucleic acid testers shorten their exposure to high temperatures.

Circumstance looks terrible in Europe, At least 15,000 people have died in Europe because of hot weather in 2022 so far, the World Health Organization said Monday, with Spain and Germany among the worst-affected countries.

The three months from June-August were the hottest in Europe since records began, and the exceptionally high temperatures led to the worst drought the continent has witnessed since the Middle Ages." Based on country data submitted so far, it is estimated that at least 15,000 people died specifically due to the heat in 2022," the WHO's Regional Director for Europe Hans Kluge said in a statement.

"Nearly 4,000 deaths in Spain, more than 1,000 in Portugal, more than 3,200 in the United Kingdom, and around 4,500 deaths in Germany were reported by health authorities during the 3 months of summer," he added." This estimate is expected to increase as more countries report on excess deaths due to heat," it said, highlighting the UN climate summit in Egypt and its calls for rapid action.

The harm done by climate change is not visited on the people, or the generations, that have the best chance of acting against it. Those who suffer most harm are and will be predominantly poor and in poor countries. The people called on to pay the cost of reducing that harm are and will be mostly much better off.

So Global Warming or Not?

In order to solve the problems, our team collect data from NASA and other official websites to build models to predict the situation and solve the problems.

Figure 01: A woman walking on a hot city street in Shanghai tries to cool off with a mini-electric fan on Thursday. TANG YANJUN/CHINA NEWS SERVICE from China Daily, https://www.chinadaily.com.cn/a/202208/12/WS62f5afe6a310fd2b29e71cfa.html

1.2 Problem Statement and Analysis

Q1-a: Do you agree that the increase of global temperature in March 2022 resulted in a larger increase than observed over any previous 10-year period? Why or why not?

Our team mainly conducted a difference analysis by collecting the global average temperature data from March 2010 to November 2022, and then divided the data into two groups: from March 2010 to March, 2020 and from March to October, 2022. **Independent sample t test** or **independent sample manu test (specifically combined with normality test)** was used to verify whether there were differences, and the degree and effect of differences were analyzed.

Q1-b: Based on the historical data, please build two or more mathematical models to describe the past and predict the future global temperature level.

Our team uses three mathematical models to predict global mean temperature, which are the traditional time series prediction model **grey prediction**, the mainstream **Lasso model** in the industry and the deep learning time series prediction model **Value model**. We use **genetic algorithm** and **particle swarm optimization** for parameter tuning to obtain more accurate prediction.

Q1-c: Use each of your models in 1(b) to predict global temperatures in 2050 and 2100 respectively. Do any of your models agree with the prediction that the average global temperature of observation points in 2050 or 2100 will reach 20.00℃? If not

in 2050or 2100, when will the average temperature of observation points in your prediction models reach 20.00℃?

Q1-d: Which model you built in 1(b) do you consider most accurate? Why?

We will explain the details in previous analysis.

Q2-a: Use the results of question 1 and the data in the attachment 2022 APMCM C Data.csv and other datasets collected by your team, build a mathematical model to analyze the relationship (if any) between global temperature, time and location, and explain the relationship or prove that there is no relationship between them.

Our team mainly carries out correlation analysis. Just like the first part of the first question, we need to collect global average temperature data, as well as its corresponding latitude and longitude and time. We will use **Grey Relation Analysis** for further digging.

Q2-b: Please collect relevant data and analyze the factors of natural disasters (such as volcanic eruptions, forest fires and the COVID-19). Is there any impact on global temperature?

Based on the examples of the Tonga submarine volcanic eruption, our team analyzed whether they caused global warming, and if so, what was the cause. And as for volcanic eruption, On the basis of understanding the effects of past volcanic eruptions on climate, we use the climate simulation data of the past millennium based on the climate system model. It is found that there is a significant correlation between the magnitude of surface cooling and the intensity of the volcano in the next year after the volcanic eruption, and the **linear model** can reproduce the magnitude and spatial distribution of the cooling after the volcanic eruption.

Q2-c: What do you think is the main reason that affects the global temperature change?

Q2-d: Do you think there are some measures to curb or slow down global warming?

We will explain the details in previous analysis.

Q3: Prepare a non-technical article (1 page maximum) Please write a non-technical article (1 page at most) to the APMCM organizing committee, explaining your team's findings and suggestions for the future.

We will write a memo below.

2. General Assumptions and Justifications

To simplify the problem, we have the following basic assumptions, which are properly justified.

- We assume that the statistics we collected from the website are reliable and accurate.
- In the long run, the global temperature curve could be adjusted as a linear line.
- In the long run, there is no huge incident which will affect the temperature vastly.
- If there is some incident, human being will tackle it soon or human being chooses negative policy in order to guarantee economic growth, which means it will not affect environment deeply but the natural incident like volcanic eruption matters. So we will not pay much attention to COVID-19 effect but volcanic effect.
- The models we choose are suitable.

3. Variable Description

VARIABLES	DESCRIPTIONS
VALUE	The Global Average Temperature
YEAR	Current Year

Other Variables will be explained in each formula below.

4. Task 1

4.1 Building the Independent Sample Test and Independent Sample MANU Test

t test, also known as Student T-test, can be said to be a very common test method in statistical inference. It is used when statistics obey normal distribution but the variance is unknown. The formula is below,

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)}}$$

t is the value, and $x1$ and $x2$ are the means of the two groups being compared, is the pooled standard error of the two groups, andn_1 and n_2 are the number of observations in each of the groups.

Mann-Whitney U test is a nonparametric rank-sum hypothesis test. A t-test method for independent samples that does not require a normal distribution. It mainly tests whether

there are significant differences between two populations that are identical except for the population mean. When the sample size is larger than 20, the test results are the best.

Steps:

1. The two groups of independent samples tested are mixed first, arranged in ascending order according to the data size and arranged in rank rank. When encountering the same data, the rank values are equal, which is the average value before the arranged rank.
2. The grade and sum of the two samples were calculated respectively: R_1 R_2;
3. The calculation formula of Mann-Whitney U test statistics U_1 and U_2 is as follows:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$
$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

Among them, n_1 and $n_2$is the sample sizes, R_1 and $R_2$are sample grade. And the minimum values of U_1 and $U_2$are used for comparison with the significant test Uα (refer to Mann-Whitney Table for specific values). IfIndicating that there is a difference between the two samples.

4.2 Building Grey Prediction Model

Grey prediction model is a prediction method that builds mathematical model to make prediction through a small amount of incomplete information. It is based on the past and present development law of objective things, with the help of scientific methods to describe and analyze the future development trend and situation, and form scientific assumptions and judgments.

i. Data preprocessing:

Assuming that the original sequence $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), X^{(0)}(3), \dots, X^{(0)}(n)\}$ passes the stage ratio test,
Take the mean generating operator as an example:

$$Z^{(1)}(k) = \frac{1}{2}\left[X^{(1)}(k-1) + X^{(1)}(k)\right] \quad k = 2,3$$

New sequence: $Z^{(1)} = \{Z^{(1)}(1), Z^{(1)}(2), Z^{(1)}(3), \dots, Z^{(1)}(n)\}$

Similarly, if the grade ratio test fails, a positive number A can be added after it, so that all the grade ratios are within the acceptable range.

ii. Model building:

a) Original form $X^{(0)}(k)+aZ^{(1)}(k)=b$, We want to find a and b .

b) Use regression analysis to estimate the corresponding whitening differential equation and and their solutions as follows:

$$\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=b$$

$$x^{(1)}(t)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-a(t-1)}+\frac{b}{a}$$

Thus, the predicted value is obtained:

$$\widehat{x^{(1)}}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a},k=1,2,\cdots,n-1,$$

Original sequence predictive value:

$$\widehat{x^{(0)}}(k+1)=\widehat{x^{(1)}}(k+1)-\widehat{x^{(1)}}(k),k=1,2,\cdots,n-1$$

iii. Construct data matrix B and data vector Y :

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

$$Y=\begin{bmatrix}X^{(0)}(2)\\ X^{(0)}(3)\\ .\\ .\\ X^{(0)}(n)\end{bmatrix}$$

Let $u=\begin{bmatrix}a\\ b\end{bmatrix}$, solve the above differential equation $u=(B^T\cdot B)^{-1}B^TY$

to get a and b which is fitted to the predicted value function

iv. Precision inspection

Residual test: $\varepsilon(k)=\frac{x^{(0)}(k)-\widehat{x^{(0)}}(k)}{x^{(0)}(k)},k=1,2,\cdots,n,$

If the absolute value of all residuals is less than 0.1, it is considered to meet higher requirements. Less than 0.2 to meet the general requirements;

Test of deviation value of stage ratio: $\rho(k)=1-\frac{1-0.5a}{1+0_{05}5a}\lambda(k)$

If the absolute value of deviation of all stage ratios is less than 0.1, it is considered to meet higher requirements. Less than 0.2 to meet the general requirements;

4.3 Building LASSO Model

LASSO was first proposed by Robert Tibshirani in 1996. Its full name is Least absolute shrinkage and selection operator.

The method is a kind of compression estimation. It obtains a more refined model by constructing a penalty function, which makes it compress some regression coefficients, that is, the sum of the absolute values of the forced coefficients is less than a certain fixed value. Also set some regression coefficients to zero. Therefore, it retains the advantage of subset shrinkage and is a biased estimator that deals with data with complex collinearity.

In considering the general linear regression problem, given n data sample points $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, where each $x_i \in R^d$ is a D-dimensional vector, that is, each observed data point is composed of d values of variables, and each $y_i \in R$ is a real value. What we need to do now is to find a mapping $f: R^d \to R$ according to the observed data points, which minimizes the sum of error squares. The optimization objective is:

$$\beta^*, \beta_0^* = argmin_{\beta,\beta_0} \frac{1}{n} \sum_{i=0}^{n} (y_i - \beta^T x_i - \beta_0)^2$$

Among them,$\beta \in R^d, \beta_0 \in R$ are the coefficients to be optimized. Generally speaking, β_0
can be regarded as a bias. Now let's look at how to deal with the bias term. Assuming the value of β is fixed now, then use the first derivative to find the optimal β_0. Then the derivative of the above loss with respect to β_0 is obtained:

$$\frac{1}{n} \sum_{i=0}^{n} (y_i - x_i^T \beta - \beta_0) = 0 \quad \Rightarrow \quad \beta_0 = \frac{1}{n} \sum_{i=0}^{n} y_i - \frac{1}{n} \sum_{i=0}^{n} x_i^T \beta = \overline{y} - \overline{x}^T \beta$$

The obtained results are substituted into the original optimization objective, we can get,

$$\beta^* = argmin_{\beta} \frac{1}{n} \sum_{i=0}^{n} \left((y_i - \overline{y}) - \beta^T (x_i - \overline{x}) \right)^2$$

As can be seen from the above formula, if we standardize (centralize) the data in advance, that is, subtract the mean value from each sample data to get the data sample with zero mean, then we can do linear regression without using bias. For the convenience of introduction below, we assume that the given n data sample points $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ are zero-mean, namely $\sum_{i=1}^{n} x_i = 0$, then the optimization objective of linear regression can be written as:

$$\beta^* = argmin_\beta \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta^T x_i)^2$$

The above can also be expressed in matrix form, denoted as $X = [x_1; x_2; \cdots; x_n]^T$, where each data point x_i is regarded as a column vector, then $X \in R^{n\times d}$, denoted as $y = (y_1, y_2, \cdots, y_n)^T$
, then the optimization objective of matrix form is:

$$\beta^* = argmin_\beta \frac{1}{n}\|y - X\beta\|_2^2$$

The Ridge Regression is optimized for:

$$\beta^* = argmin_\beta \frac{1}{n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2$$

The optimization objectives of Lasso are:

$$\beta^* = argmin_\beta \frac{1}{n}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

4.4 Building Value Model.

Value regression model uses the improved least square estimation method, by giving up the unbiased least square method, at the cost of losing part of the information and reducing the accuracy to get the regression coefficient more in line with the reality, more reliable regression method, the fit of ill-conditioned data is better than the least square method.

By adding a small perturbation λI to the original least square estimation of β, the generalized inverse can be obtained from the original situation, which makes the problem stable and solved.

Normally, Regression model can be written as $y = \sum_{j=1}^{p}\beta_j x_j + \beta_0$

$$\widehat{\beta^{\text{ridge}}} = argmin(\beta)\left\{\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right\}$$

Lambda here is also a parameter to be evaluated. In other words, ridge regression is the least squares regression with a di norm penalty.

4.5 Analyze the data and Draw the Conclusion

Q1-a)

Independent Sample t Test:

Output result 1: normality test result

variable name	sample capacity	median	average value	standard error	skewness	kurtosis	S-W checkout	K-S checkout
Value1	152	0.825	0.817	0.164	0.012	0.44	0.982(0.040**)	0.09(0.164)

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

The Value1 used the S-W test, with a significance P-value of 0.04, showing significance at the level, and rejecting the null hypothesis, so that the data did not satisfy the normal distribution. The absolute value of its kurtosis (0.44) is less than 10, and the absolute value of its skewness (0.012) is less than 3, which can be further analyzed by combining with the normal distribution histogram, PP diagram or QQ diagram.

Table of T-test analysis results of independent samples

variable name	variate-value	sample capacity	average value	standard error	t	P	Mean difference	The Cohen's d value
Value1	before	145	0.814	0.167	-	0.433	0.05	0.304
	now	7	0.864	0.04	0.786			
	amount to	152	0.817	0.164				

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

Previously, the mean on Value1 was now 0.814 / 0.864; the P value of F test result was 0.433> =0.05, so the statistical result was **not significant**, indicating that there was no significant difference on Value1; the magnitude Cohen's d value was 0.304, small (0.20,0.50 and 0.80 correspond to small, medium and large critical points respectively).

Output result 1: normality test result

variable name	sample capacity	average value	standard deviation	skewness	kurtosis	S-W checkout	K-S checkout
Value1	152	0.817	0.164	0.012	0.44	0.982(0.040**)	0.09(0.164)

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

Independent Sample MANU Test:

Output result 3: MannWhitney U test and analysis results table

variable name	variate-value	sample capacity	median	standard deviation	statistics	P	Median value difference	The Cohen's d value
Value1	1	145	0.82	0.167	407.5	0.379	0.06	0.304
	2	7	0.88	0.04				
	amount to	152	0.825	0.164				

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

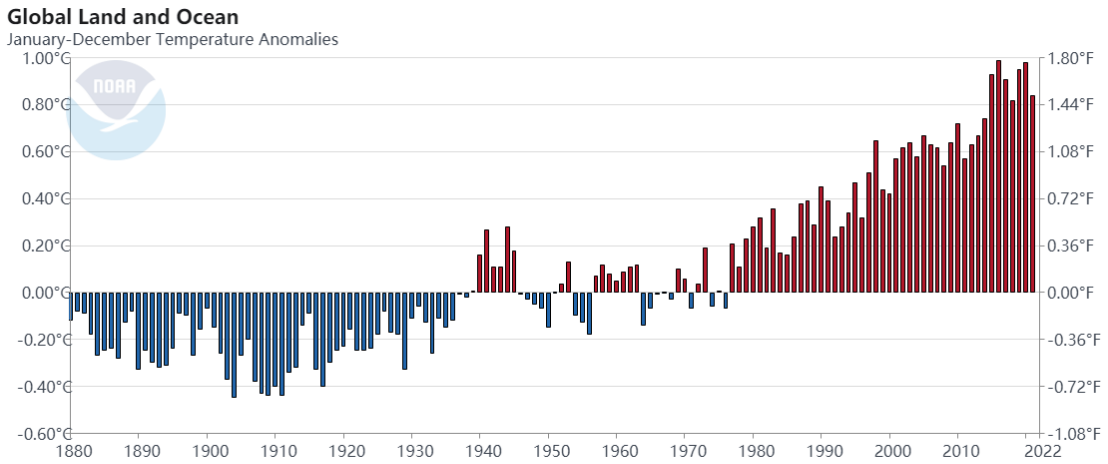
Note: * * *, * * and * represent the significance levels of 1%, 5% and 10%, respectively

The median of 1.0 and 2.0 on Value1 is respectively: 0.82 / 0.88; the P value of the test result is 0.379, so the statistical results are **not significant**, 1.0 and 2.0 in Value1; Cohen's d value is 0.304, the difference is small.

So, we conclude that the increase of global temperature in March 2022 does not result in a larger increase than observed over any previous 10-year period

Q1-b)

Global Temperature

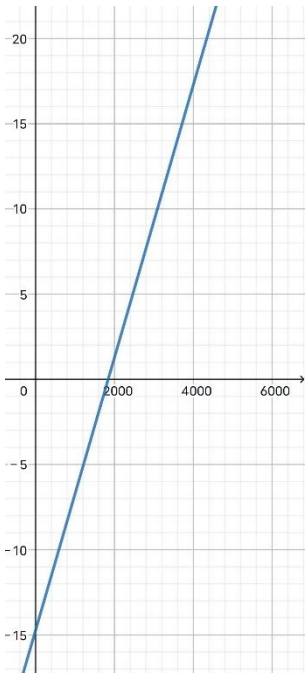


Lasso Regression Model:

The results of **Lasso regression** showed that: based on the variable intercept term and Year's standardized coefficient, the variable intercept term and Year were retained, and no variables were removed.

- Standardized formula of the model: y=-14.725+0.008 Year.
- Non-standardized formula of the model: y=-14.725+0.008 Year.

variable name	Standardization coefficient	Non-standardized coefficients	R^2
nodal increment	-14.725	-14.725	0.772
Year	0.008	0.008	



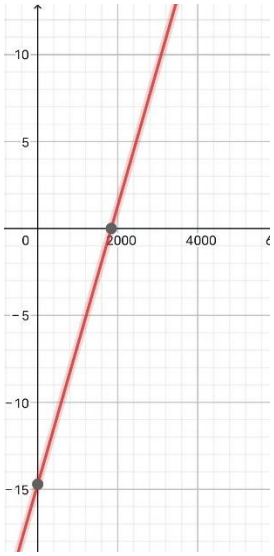
Value Regression Model:

The results of the **Value regression** showed that: based on the F test significance P-value of 0.000 * * *, the null hypothesis was rejected, indicating that there is a regression relationship between the independent variable and the dependent variable. Meanwhile, the goodness of fit R^2 is 0.772, and the model performance is relatively good.

- Formula for the model: Value=-14.71+0.008 Year

K=0.001	Non-standardized coefficients		Standardization coefficient	t	P	R^2	adjust R^2	F
	B	standard error	Beta					
constant	-14.71	0.679	-	-21.668	0.000***	0.772	0.771	474.731(0.000***)
Year	0.008	0	0.878	21.788	0.000***			
Dependent variable: Value								

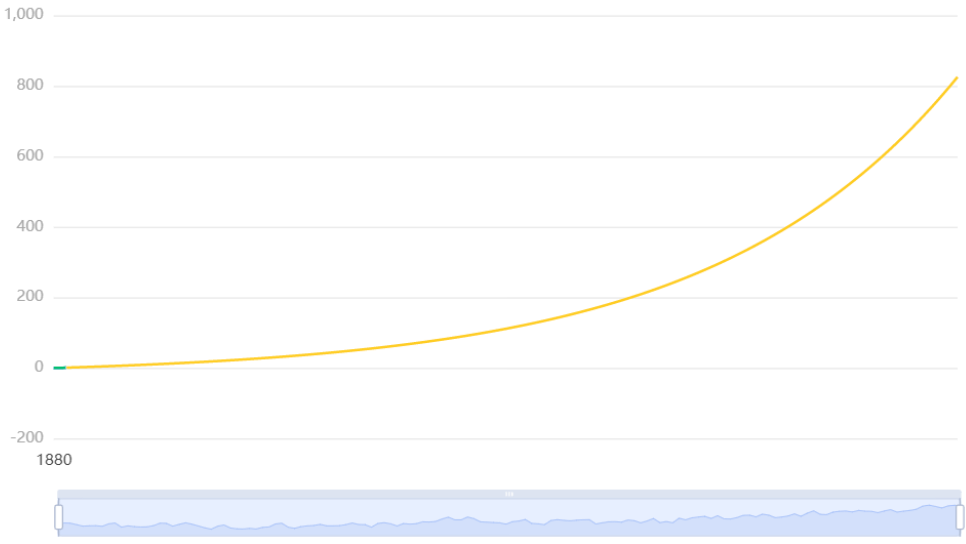
Note: * * *, * * and * represent the significance levels of 1%, 5% and 10%, respectively



Grey Prediction Model:

Development coefficient a	grey action b	posteriori ratio C
0	20.532	0.217

According to the analysis in the above table, the posterior difference ratio is 0.217, indicating a **high accuracy** of the model.



Q1-c)

From the calculation, we know that the global average temperature will not rise to 20℃

in 2050 or 2100, but the average temperature of the earth will rise to 20℃ in 3767.

Q1-d)

I think the grey prediction model is better because the main feature is that the model uses the generated data series rather than the original data series, and the core system is the grey model (GM), that is, the original data is generated by accumulation (reduction generation, weighted adjacent value generation) to obtain the approximate exponential rule and then modeling.

Pros:

- The irregular original data is generated to obtain the generated sequence with strong regularity.
- Large samples are not required.
- Samples do not need to be regularly distributed.
- the calculation work is small.
- The results of quantitative analysis will not be inconsistent with those of qualitative analysis.
- It can be used for recent, short and medium terms forecasts.
- High accuracy of grey prediction.

5. Task 2

5.1 Building Grey Relation Analysis Model

Grey Relation Analysis (GRA) is a multi-factor statistical analysis method. In a grey system, we want to know the relative strength of one of the projects we are concerned about relative strength of other factors, Grey Relation Analysis Model is a good model to analyze the relationship between global temperature, time and location.

Here are the steps to build the Grey Relation Analysis Model,

i. Establish the parent sequence,
ii. Normalization,
iii. Calculate the grey correlation coefficient,
iv. Calculated to form the correlation order to obtain the mean value of correlation coefficients.

In details,

There are two different calculational methods:

$$X_i = \left(\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \right)$$

Which is used for larger quality characteristic,
and

$$X_i = \left(\frac{x_{\max} - x_i}{x_{\max} - x_{\min}} \right)$$

Which is used for smaller quality characteristic

Where $x_{\max} - x_{\min}$ are the maximum and minimum values of the original sequence .

Grey Coefficient, ζ_i is given by

$$\zeta_i(k) = \frac{\min\limits_i \min\limits_k |x_0(k) - x_i(k)| + \rho \cdot \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \cdot \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}$$

5.2 Building Linear Model

For two continuous variables, the statistical index describing the closeness of the linear relationship between the two variables and the direction of correlation is called the correlation coefficient, also known as Pearson product moment correlation coefficient statistically.

The linear correlation coefficient of samples is generally expressed by r, and the overall correlation coefficient is generally expressed by p. Correlation coefficient is no unit, its scope is $|r| < 1$.

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

5.3 Analyze the data and Draw the Conclusion

Q2-a)

Step of analysis

1. Conduct dimensionless processing (mean and initial value) for the data.
2. Solve the grey correlation value between the parent sequence (contrast sequence) and the feature sequence.
3. Solve the grey correlation degree value.

4. Rank the value of grey correlation degree and draw a conclusion.

Data analysis:

It can be seen from the above table that grey correlation degree analysis is carried out for 4 evaluation items (Latitude, Longitude, dt, City) and 228,175 data items, and Average Temperature is taken as the "reference value "(parent sequence). This paper studies the correlation between four evaluation items (Latitude, Longitude, dt, City and Average Temperature) and provides reference for analysis based on the correlation degree. When using grey correlation degree analysis, the resolution coefficient is set at 0.5 and the correlation value is calculated by combining the correlation coefficient calculation formula. According to the correlation value, the correlation degree value is calculated for evaluation and judgment.

Resolution coefficient $\rho \in (0, \infty)$, the smaller ρ, the greater the resolution, the general value of ρ interval is $(0, 1)$, the specific value depends on the situation. When $\rho \leq 0.5463$, the resolution is the best, usually take $\rho = 0.5$.

Correlation Coefficient Graph

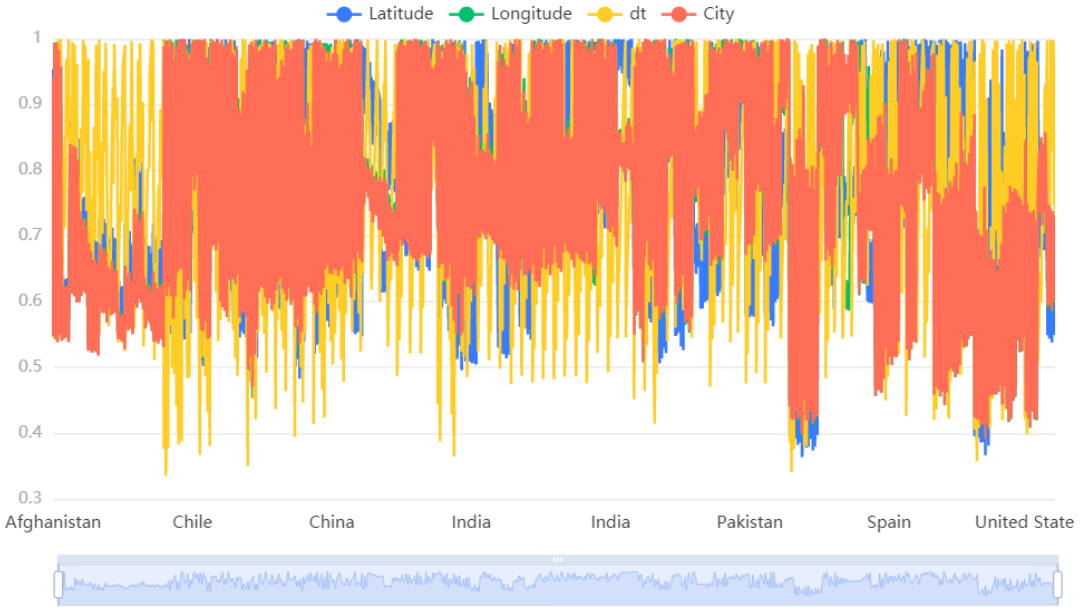


Chart Description:

The correlation coefficient represents the degree of connection between the dt, Latitude, Longitude pairs and the dimensions corresponding to the parent sequence (the larger the number, the stronger the connection).

Grey Relation Degree

Items	Relation	Ranking
Latitude	0.763	1
dt	0.755	2
City	0.75	3

Longitude	0.75	4

Chart Description:

Correlation degree represents the similar correlation degree between each evaluation item and the "reference value", which is obtained by calculating the average value of the correlation coefficient. The correlation degree value ranges from 0 to 1. The larger the value, the stronger the correlation between the evaluation item and the "reference value", and the higher the correlation degree, which means the closer the relationship between the evaluation item and the "reference value", so the higher the evaluation. Combined with the correlation degree value, all evaluation items were sorted to obtain the ranking of each evaluation item.

Data analysis:

According to the above correlation coefficient results, the weighted processing is carried out to obtain the correlation degree value, and the correlation degree value is used to evaluate and rank the four evaluation objects. The correlation degree value is between 0 and 1, and the larger the value is, the stronger the correlation is with the "reference value" (parent sequence), that is, the higher the evaluation is. As can be seen from the above table, among the four evaluation items, Latitude received the highest evaluation (correlation degree: 0.763), followed by dt(correlation degree: 0.755).

Taking analysis into consideration, we find that there is relationship between global temperature and time, location, and Latitude > dt > city=Longitude.

Q2-b)

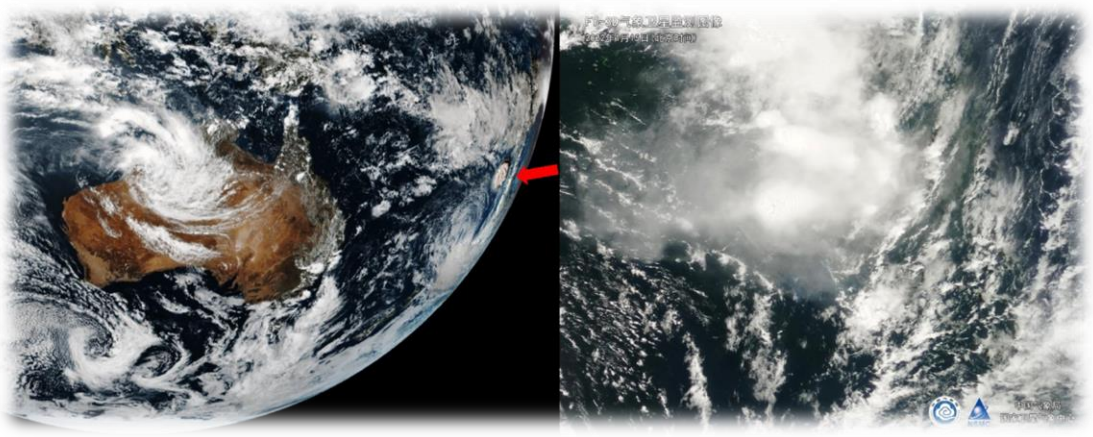
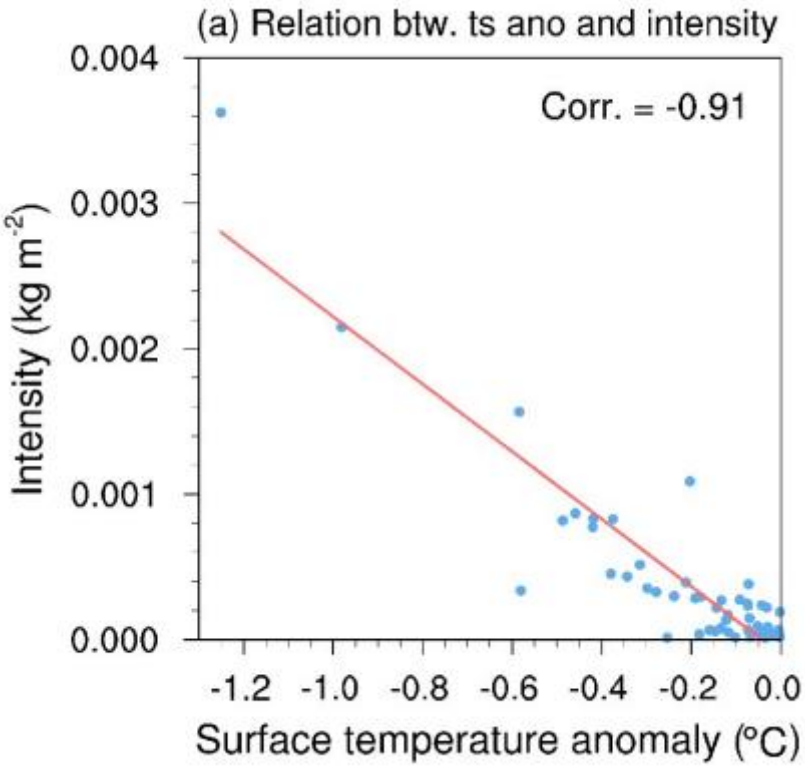


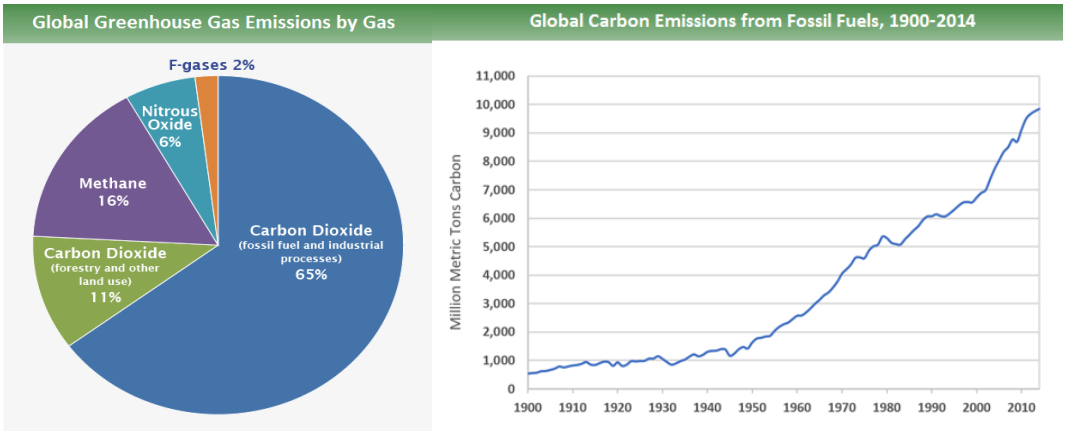
Figure 02 The FY-4B satellite captured the eruption of the Hunga Tonga-Hunga Ha'apai (HTHH) volcano (left):monitoring the diffusion of volcanic ash clouds by the FY-4B satellite from 1545 to 2100 LST (Local Standard TimeLST-UTC+13 hours) on 15 January 2022 (right) (Provided by the National Satellite Meteorological Center of China).



We estimated a reduction in global mean surface air temperature by approximately 0.0315°C −0.1118°C in the next 1−2 years after the eruption of HTHH volcano.

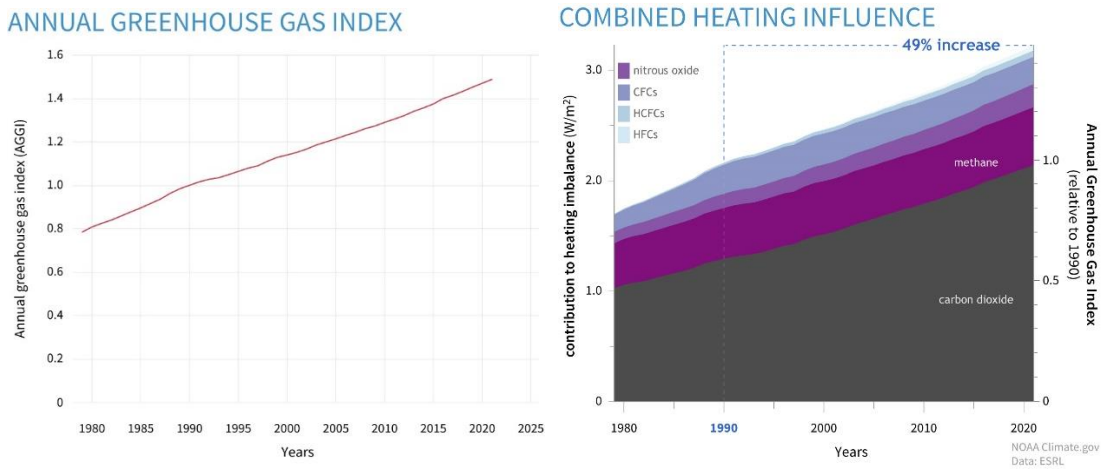
Q2-c)

The greenhouse effect is essential to life on Earth.



Biologically produced greenhouse gases are a part of our atmosphere. Since its environment is exactly right for encouraging the growth of life, including our own, Earth is commonly referred to as the "Goldilocks" planet. The reason the Earth is so friendly is because of its natural greenhouse effect, which normally maintains the planet's surface temperature at a pleasant 15 °C (59 °F). However, over the past century

or so, human activity has changed the planet's energy balance, primarily through the burning of fossil fuels, which raises the atmospheric concentration of carbon dioxide. Since decades, the Earth's atmosphere has been steadily adding more carbon dioxide, which traps more heat near the surface and raises temperatures.



- The AGGI compares current circumstances to 1990 while combining the warming effects of the principal greenhouse gases created by humans.
- The AGGI reached 1.49 in 2021, which represents a 49 percent increase in the warming effect of greenhouse gases since 1990.
- It took more than 200 years from the start of the Industrial Revolution in 1750 for the AGGI to go from 0 to 1 (to reach 100%), but only another 30 years for it to hit nearly 1.5.
- Today's atmosphere absorbs an additional 3.22 watts of energy per square meter of Earth's surface compared to pre-industrial periods.

Q2-d)

We must prevent greenhouse gas concentrations, such as carbon dioxide, from rising in order to halt climate change. Carbon dioxide levels have been rising over the past 150 years as a result of burning fossil fuels and removing trees, which absorb carbon dioxide naturally. The two major approaches to prevent the atmospheric concentration of greenhouse gases from rising are to cease releasing them into the atmosphere and to improve Earth's capacity to absorb them.

Climate mitigation is what this is. There are several ways to reduce climate change. To stop the planet from warming, we must instead combine a variety of strategies. The key techniques we can employ are described in the sections below.

Numerous of these solutions are already in use in various parts of the world. Individuals may take on some of these challenges by using less energy, biking instead of driving, driving electric vehicles, and converting to renewable energy sources. Other measures to combat climate change entail cooperating amongst towns, regions, or countries to implement improvements, such as converting power plants from coal or gas to

renewable energy sources and expanding public transportation.

Here are what we can do to decrease of greenhouse effect

- We need to lessen the earth's greenhouse effect in order to prevent global warming.
- We must cut back on greenhouse gas emissions if we want to lessen the greenhouse effect.
- increasing tree planting and minimizing deforestation
- Pollution and the greenhouse effect can be reduced by reducing the usage of fossil fuels.
- Energy conservation is another way to lower emissions of greenhouse gases.

5. Task 3

With the continuous development of the level of productivity, the development of human production to a new level, the level of industrial progress. At the same time, the population is increasing day by day, the temperature of the atmosphere is rising, which leads to the melting of glaciers, the rise of the sea level, the sea water over the land, inundating land and cities, people are facing more and more serious temperature problems. Standing at the fork of fate, we cannot escape and must provide reliable forecasting solutions with a global strategic vision, so as to make further planning for human development. Based on this, our team first made the following findings:

Using data and algorithms, we know that the global average temperature will not rise to 20℃ in 2050 or 2100, but will rise to 20℃ in 3767. Latitude has the greatest effect on temperature, followed by time, then city, and finally longitude. Through the linear prediction model, we found that there is a negative correlation between the volcanic eruption and the temperature, and the volcanic eruption will make the temperature drop. The greenhouse gas effect is the main reason for the influence on the temperature, because the atmosphere is mainly composed of some greenhouse gases, and the emission of greenhouse gases increases exponentially. In recent years, the greenhouse gas emission index also increases linearly. The increase of 0.1 per decade or so is not encouraging.

In this regard, our team puts forward the following targeted calls and suggestions:

First of all, legislation should be made to ensure that energy can be guaranteed and carbon emissions can be controlled. Energy saving laws and necessary inducement measures should be adopted to strengthen the further energy saving work of factories that use a lot of energy, promote the energy saving work of factories, promote the development of innovative energy saving technology, promote the acceleration of introduction, and thoroughly promote the fuel conversion work.

Second, in this period, it should be recognized that the best new way of life in the 21st century is to save energy and resources. On the premise of deeply realizing that global warming is our own problem, we need to start from the little things around us and aim to build a society with small environmental load related to global warming. In order to improve the actual effect of national cooperation, it is not only necessary to

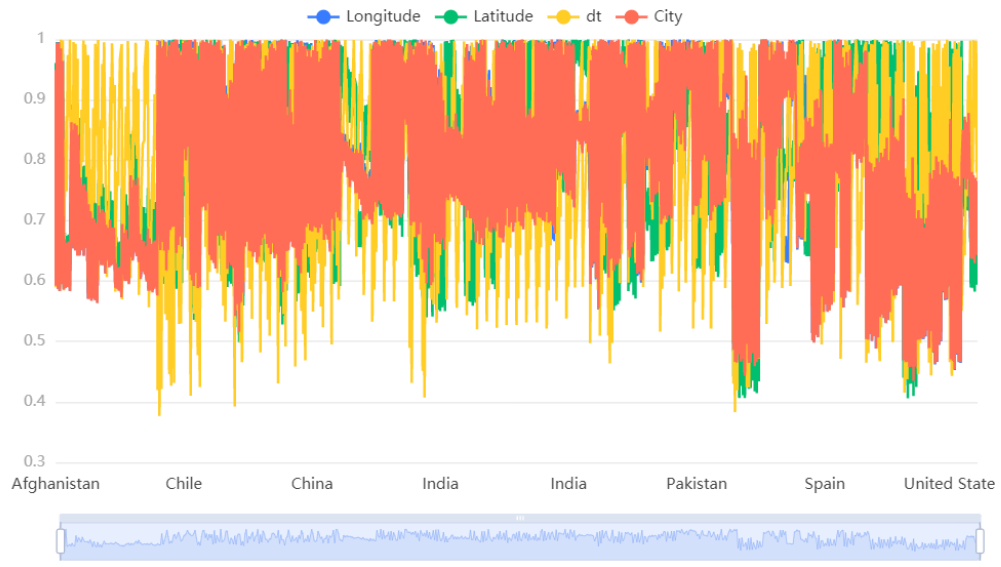
provide information on how to prevent global warming and how to cause further deterioration of warming, but also to create opportunities to promote national independent cooperation, and it is also necessary to actively commit to the development and sales of such products.

Thirdly, in order to fundamentally solve the problem of global warming, we must significantly reduce greenhouse gases. In order to achieve fundamental energy conservation and reduce carbon dioxide emission sites, innovative technology development has become the key. In terms of developing innovative technologies, not only as an enterprise, the state should also play an active role. For developing countries, it is necessary to carry out comprehensive and planned strict follow-up actions on the basis of the current global warming countermeasures.

These are our group's findings and strategic views. Thank you.

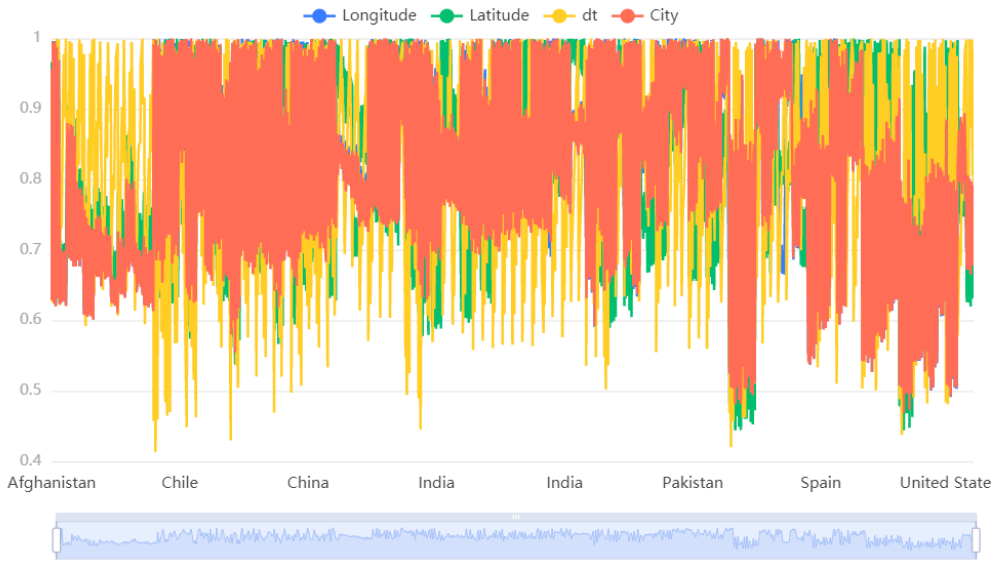
7. Sensitivity Analysis

Set $\rho = 0.6$ in Grey Relation Analysis Model, we can get,



Items	Relation(ρ =0.6)	Ranking	Relation(ρ =0.5)	Relative Error
Latitude	0.792	1	0.763	0.038007864
dt	0.784	2	0.755	0.038410596
City	0.78	3	0.75	0.04
Longitude	0.779	4	0.75	0.038666667

Set $\rho = 0.7$ in Grey Relation Analysis Model, we can get,



Items	Relation(ρ =0.7)	Ranking	Relation(ρ =0.5)	Relative Error
Latitude	0.814	1	0.763	0.066841415
dt	0.807	2	0.755	0.068874172
City	0.803	3	0.75	0.070666667
Longitude	0.803	4	0.75	0.070666667

The degree of absolute relation will rise but all are below 1%,and the degree of relative relation will not change. So, we believe that our model is stable.

8. Strengths and Weakness

8.1 Strengths

- By sensitivity analysis, our model has good robustness and accuracy.
- In our model, we transform many parameter normalization processes into dimensionless indexes, and the results are relatively correct and more convincing.
- At the same time, by qualifying the factors, the process of evaluation is clearer and more intuitive.
- The evaluation indexes are comprehensive and objective.
- After normalization, they are presented in various forms of charts, which are clear and intuitive. We linked the components in our model to the factors in social development, so our models are understandable.

8.2 Weakness

- Data we collect is from multiple websites, the difference in statistical standards

may result in bias conclusions.

- What's more, the lack of data for more indicators may lead to the errors in the evaluation system.
- Some parameters are based on common sense because few data are available. Some may be not credible.
- We do not use any advanced algorithm in order to simplify the problem, which may lead to the lack of robustness of the results. In our model, the explosive changes and huge human disruptive events are not taken into consideration.

9. References

[1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com.

[2] Azzeh M , Neagu D , Cowling P I . Fuzzy grey relational analysis for software effort estimation.[M]. Kluwer Academic Publish Zuo, M., Zhou, T., Man, W. et al. Volcanoes and Climate: Sizing up the Impact of the Recent Hunga Tonga-Hunga Ha'apai Volcanic Eruption from a Historical Perspective. Adv. Atmos. Sci. 39, 1986–1993 (2022). https://doi.org/10.1007/s00376-022-2034-1ers, 2010.

[3] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com.

[4] Fisher Box, Joan. Guinness, Gosset, Fisher, and Small Samples. Statistical Science. 1987, 2 (1): 45–52.

[5] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com.

[6] 杨维康,乔建芳,高行宜,钟文勤.秩和检验在鸟类栖息地选择分析中的运用[J].干旱区研究,2001(03):42-44.

[7] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com.

[8] 邓聚龙. 灰色预测与灰决策[M]. 武汉:华中科技大学出版社,2002.

Appendix

```
clc;clear;
syms a b;
c = [a b]';
A = data(C:\Users\Administrator\Desktop\1880-2022.xls);
T1=142
T2=10000;
t1=1:T1;
t2=1:T1+T2;
n = length(A);
B = cumsum(A);
```

```
for i = 2:n
    C(i) = (B(i) + B(i - 1))/2;
end
C(1) = [];
B = [-C;ones(1,n-1)];
Y = A; Y(1) = []; Y = Y';
c = inv(B*B')*B*Y;
c = c';
a = c(1); b = c(2);

F = []; F(1) = A(1);
for i = 2:(n+10)
    F(i) = (A(1)-b/a)/exp(a*(i-1))+ b/a;
end

G = []; G(1) = A(1);
for i = 2:(n+10)
    G(i) = F(i) - F(i-1);
end

disp('The Prediction is：');
G

H = G(1:10);
epsilon = A - H;

delta = abs(epsilon./A);
disp('Q Test：')
Q = mean(delta)

C = std(epsilon, 1)/std(A, 1)

%
S1 = std(A, 1);
tmp = find(abs(epsilon - mean(epsilon))< 0.6745 * S1);
disp('P Test is：')
P = length(tmp)/n


plot(t1, A,'ro'); hold on;
plot(t2, G, 'g-');
grid on;
```