Team Number:	apmcm2203757
Problem Chosen:	C

2022 APMCM summary sheet

In recent years, the rising trend of temperature in many countries has become more and more obvious. In this paper, we mainly use multiple statistics and neural networks to establish an explainable fitting model of the global temperature change trend and use some optimized correlation coefficient methods to dig out key factors affecting global warming. Finally, some scientific plans and suggestions are given based on the established global warming analysis and prediction model.

Aiming at problem 1, we first collect the global temperature data from 2013 to 2022 and conduct a multi-level difference analysis. It is found that 30-50 years after the industrial revolution (1793-1812) is the period of the most drastic temperature change. Secondly, we use **Transformer** model, **ARIMA** model, and optimized ARIMA **(ARIMA-LSTM)** model to effectively fit the historical data and they all conducted the **Interpretable Modeling** respectively. Then, based on the well-trained models, we inject the historical data and obtain the temperature prediction data for 2050 and 2100 respectively. The modeling results of ARIMA found that the global temperature will reach the threshold of 20 ° C in 2055, and the transformer model found that it will exceed 20 ° C for the first time in 2048.

Aiming at problem 2, we first add the collected CO2 dataset. Then Spearman Correlation Coefficient **(SCC)**, Pearson Correlation Coefficient**(PCC)**, and Multi-level Principal Component Analysis **(MPCA)** are then respectively applied to our model processing, and the results show that the correlation of spatial dimensional characteristics is the strongest. Then, we also collect other natural disaster data and they are then also reranked. It is found that the impact of temperature rise caused by CO2 is obvious.

Aiming at problem 3, based on the above conclusions, we summarized and reviewed the full text, and summarized the generation, development, influencing factors, and preventive measures of global warming. Finally, from the time dimension and space dimensions, we put forward some suggestions and initiatives to slow down global warming.

Keywords: Transformer neural network Interpretable modeling ARIMA-LSTM Multi-level principal component analysis Spearman correlation coefficient Pearson correlation coefficient

Contents

I. Problem background

Global warming is the most important environmental issue of concern to countries around the world today and is the most familiar environmental issue to the global public. It is a nature-related phenomenon that is caused by the accumulation of the greenhouse effect, which leads to an imbalance between the energy absorbed and emitted by the earth's air system, and the energy keeps accumulating in the earth's air system, thus leading to an increase in temperature and causing global warming. In recent years, more and more high-temperature records have been set, and almost all climatologists agree that the Earth has warmed in recent years.

II. Problem analysis

2.1 Question1:

In order to analyze the global temperature data and changes, we introduce ARIMA (Autoregressive Integrated Moving Average model) and TransFormer Model to analyze the predicted data, so as to get the conclusion of global temperature change

- a) First, we preprocess the data, and then global temperature data from 2013 to 2022 is collected. Here, all data are statistically analyzed, and the temperature increase per decade is plotted in decadal increments to determine whether the temperature increase in 2012-2022 is greater than in any previous decade
- b) Based on the historical data, statistical model ARIMA and neural network model Transformer is built to describe the past and predict the future global temperature level respectively, and the data are sliced and diced at the key point of the industrial revolution
- c) Using the two mathematical models in b to predict the data for the next 37 years and 87 years respectively, determine whether the global average temperature reaches 20°C and if not, then find the year when it reaches 20°C by setting the corresponding temperature value
- d) The optimal model is selected by two model evaluation indexes, rmse and mae, and its robustness is tested by applying sensitivity analysis

2.2 Question 2:

In order to explore the influencing factors and main causes of temperature change, we introduce correlation analysis model and principal component analysis to make targeted suggestions for mitigating global warming

- a) In order to analyze the correlation between temperature, time, longitude, and dimensions, we analyze the relationship between variables through the Spearman correlation coefficient
- b) We first collected and integrated data on natural disasters (volcanic eruptions, earthquakes, typhoons) and carbon dioxide, and then based on the characteristics of the data, based on the model in A, we improved and chose to use the Spearman coefficient for correlation analysis
- c) Combined with the above questions a and b, we then use the method of principal component analysis to reduce the dimensionality of the characteristic factors that are known to have an influence, so as to find the main reasons affecting global temperature changes
- d) Based on the results of the analysis in c, we make targeted recommendations based on the main factors affecting temperature change

III. Symbol description

Table 1 Symbol description

Symbol	Description	Symbol	Description
p	order of the autoregressive model	R	Correlation coefficient matrix
q	order of the moving average model	W_i	the weight of neural network
d	order of the difference	Z_i	Principal component of PCA
η	the dimension of PE	λ_i	The contribution rate of each ingredient
ρ_p	Person correlation coefficien	b_i	the bias of neural network
ρ_s	Spearman correlation coefficient	r_{ij}	correlation coefficient of PCA

IV. General Assumptions

- **Assumption 1:** It is assumed that the Earth's ecosystem will remain stable for the next 20 years.
- **Assumption 2:** It is assumed that there will be no great breakthroughs in human technology and that existing energy sources will remain the mainstay.
- **Assumption 3:** It is assumed that the data collected in this paper are authentic and reliable and can accurately reflect the basic patterns of global climate change.
- **Assumption 4:** It is assumed that no new factors affecting the Earth's climate will emerge in the next 20 years

V. Model and solution of problem one

5.1 Data processing and analysis

Here we preprocess the data and then perform statistical analysis to solve problem a. Note that, problems b, c, and d can be all processed by the following transformer and ARIMA model. Figure 1 shows the whole working process of models.

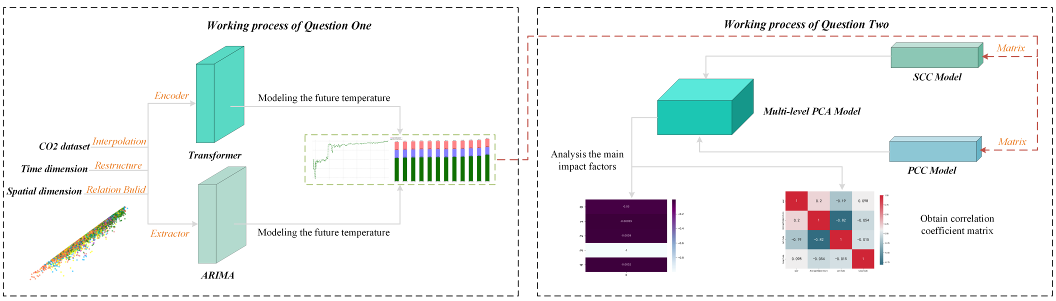


Figure 1 The pipeline of our proposed model, which can handle multiple types of data and give the main influencing factors.

5.1.1 *Pre-processing*

Here we first start by preprocessing the data, we find that many values have lost in the original dataset. Therefore, we performed filling in the missing parts of the average temperature and temperature deviation (uncertainty) of the data by the three-time insertion method.

There are various filling methods, such as linear interpolation, cubic spline interpolation.

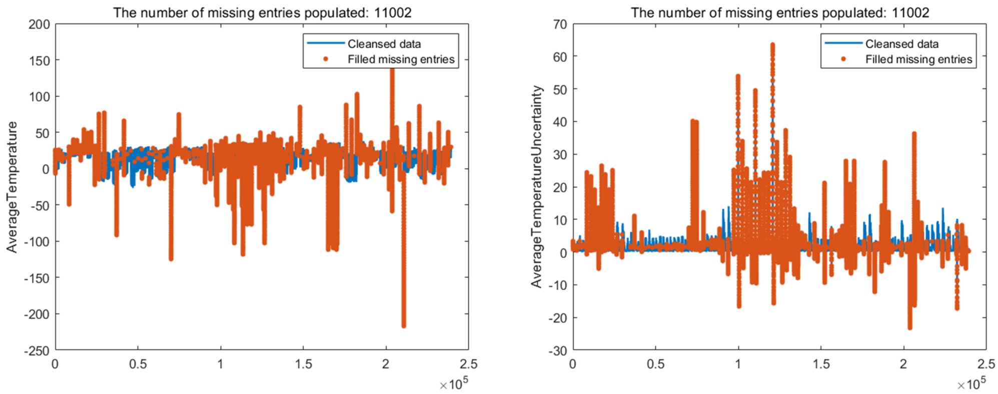
- Linear interpolation equation:

$$y = y_0 + (x - x_0)\frac{y_1 - y_0}{x_1 - x_0} \tag{1}$$

- Three-time spline interpolation formula:

$$y = a_i + b_i x + c_i x^2 + d_i x^3 \tag{2}$$

Since linear interpolation can only ensure the continuity of each small curve at the connection point, and it can not guarantee the smoothness of the entire curve. Here, we use the cubic segmented spline interpolation method, which can perfectly solve the problem and the segmentation avoids the higher-order Longer phenomenon.



5.1.2 *Statistical analysis*

We search the monthly average temperature data set of each region on the network and make a statistical analysis based on the official data set.

1. **The tropical, temperate, and global temperatures all show an upward trend over time:** we analyze the temperature changes in the tropical, temperate and global regions from 1903-2022. In order to more clearly show the relationship between the tropics, temperate zones, and the globe, as well as their changes over time, a stacked histogram is made.
   In figure 2, we can find that each temperature shows an upward trend over time, and the difference between tropical temperature, global temperature, and temperate temperature is not obvious over time.
2. **Fluctuation of global temperature:** In order to better reflect the fluctuation of global temperature, we decide to use the temperature fluctuations of 48 different

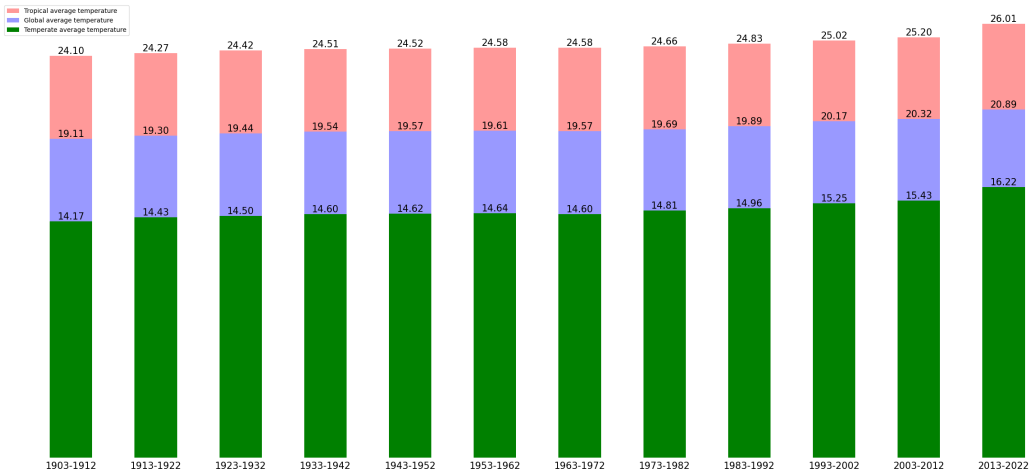


Figure 2 Comparison of annual average temperature in global, tropical, and temperate zones(℃)

countries to see the largest from the smallest. The abscissa of the image is the highest temperature T_{high} (Formula 3) of the country in a certain year and month. The ordinate is the lowest temperature T_{low}(Formula 4) in a certain year and month, randomly sampling 10000 points. We have set up 6 national collections, which are represented by different colors.

$$T_{high} = AverageTemperature + AverageTemperatureUncertainty \tag{3}$$

$$T_{low} = AverageTemperature - AverageTemperatureUncertainty \tag{4}$$

First, we can know the relationship between two points by their relative positions on the graph.

For example, if point A is directly above or to the left of point B, it means that the temperature change amplitude of point B country is greater than that of point A. That is, the more the point deviates from the linear line, the greater the temperature change of the country at that point in a specific time. Most of the points in the figure fall near the linear line, which indicates that the temperature changes in most countries are within a reasonable range in most of the time, but there are also large temperature fluctuations in some countries in some of the time.

Secondly, the maximum and minimum temperatures of most points are linear, and the density of points in the 10-40 Celsius range is the largest.

We speculate that this is because the temperature changes in different regions of the world are not different at the same time, indicating that global warming affects all

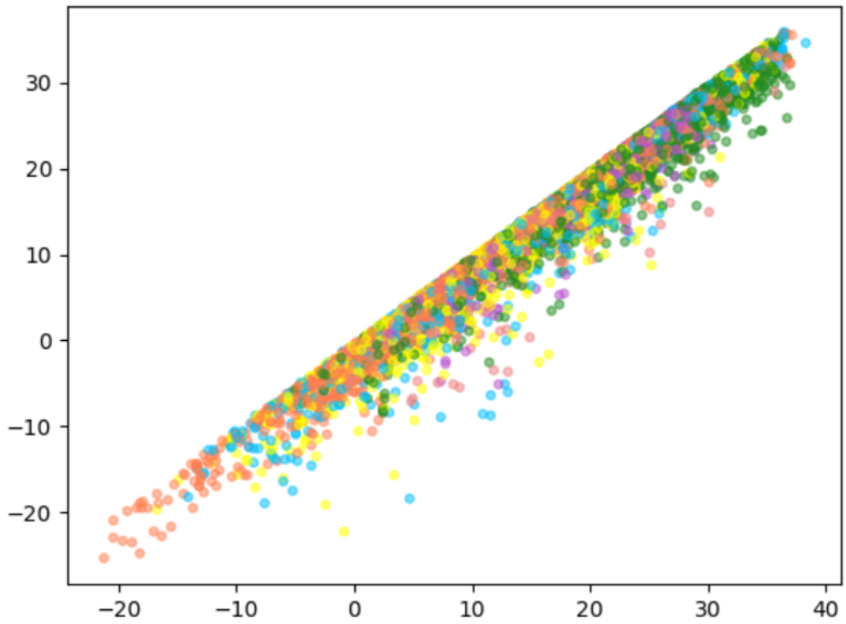


Figure 3 Correlation of temperature changes in different countries at different times(°C)

regions of the world evenly. In addition, it also shows that the world is a community of shared future, and temperature changes in several countries will simultaneously affect other countries in the world.

3. **Abnormal changes**

From 1743 to 2022, the global temperature change per decade was mostly within 1°C, but the global average temperature increased by 8.23°C from 1793 to 1802, and the global average temperature decreased by 4.57°C from 1803 to 1812. We plot Figure 4 separately. In order to better show the temperature changes during this time period, we have marked the global average temperature in the histogram for each year from 1793 to 1812. The histograms at the beginning and end are the changes in global temperature from 1793 to 1802 and 1803 to 1812; the pink and purple histograms represent the specific changes in global average temperature from 1793 to 1802 and 1803 to 1812, respectively.

In addition to the impact of climate anomalies, this huge change is also due to the fact that in the early days of the first industrial revolution, human beings did not properly handle the relationship between industrial development and the environment, and the emissions of carbon dioxide and other greenhouse gases increased greatly, resulting in their rapid growth.And the drastic drop in global temperature over the ensuing decade, which we speculate is due to the Earth's own ability to regulate and a better understanding of and remedial action on industrial development.

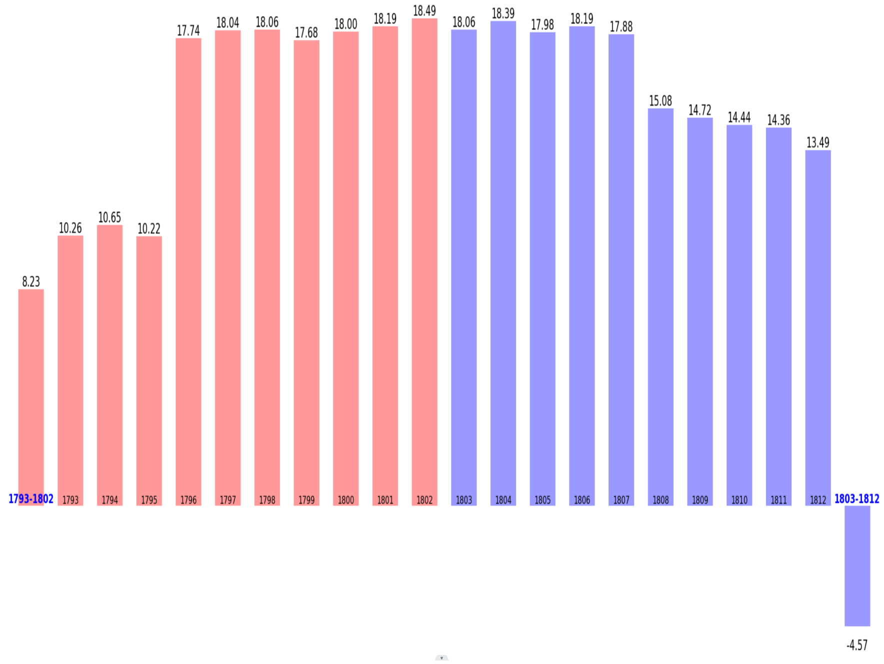


Figure 4 Temperature anomalies from 1793 to 1812(°C)

5.2 Change in temperature increase for problem a

We can conclude that **the 2013-2023 increase due to the March 2022 global temperature rise is not the largest increase of any previous 10-year period.** Specifically,its increase is larger than that during any decade in the last 150 years (1873 to 2022), but the temperature increase from 2013 to 2022 is not the largest when the data are counted up to 1793.

- **The 2013-2022 temperature change has the largest increase in any decade in the last 150 years(1873 to 2022).**The root cause of this phenomenon is global warming, and the Intergovernmental Panel on Climate Change (IPCC) report released earlier this year says that the dramatic increase in global average temperatures has had an irreversible impact on the vulnerability of global ecosystems, and that the heat wave that will sweep the world in 2022 could be considered one of many climate change phenomena caused by global warming.
- **But the temperature change from 2013-2022 is not the largest decadal increase in history.**It can be obtained from Figure 3,the first industrial revolution (1860s to mid-19th century), when human society entered the steam age, greatly increased

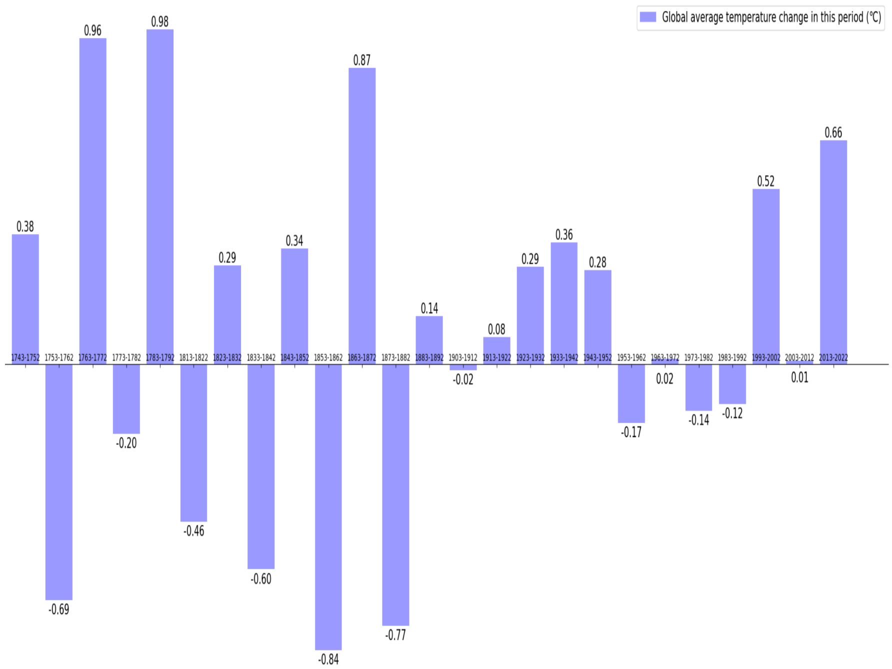


Figure 5 Decade change in temperature from 1743 to 2022(℃)

greenhouse gas emissions such as carbon dioxide, plus the impact of anomalous climate phenomena, which led to a global average temperature increase of 8.23°C from 1793 to 1802. And 178 parties around the world signed the Paris Agreement together in 2016, which also made the temperature increase slow down in recent years. Therefore, the increase from 2013-2022 is the largest in any decade in the last 150 years, but not the largest increase in any decade in history.

5.3 ARIMA Model to describe and predict

ARIMA is a class of models that capture a set of different standard time structures in time series data. The lag of a stationary series in a prediction equation is called the ”autoregressive” term, the lag of the forecast error is called the ”moving average” term, and the time series that require differentiation to make it stationary is called the ”synthetic” version of the stationary sequence

The ARIMA model can be thought of as a ”filter” that tries to separate the signal from the noise and then extrapolate the signal into the future to make predictions. ARIMA models are particularly suitable for fitting data that show non-stationarity. ARIMA’s differential eliminates changes in the time series level, eliminating trends and seasonality, thereby stabilizing the average value of the time series.

5.3.1 *Data pre-porcessing*

We merged the data set 2022_$APMCM$_C_$Data.csv$ according to time, and obtained the global monthly average temperature from 1793 to 2013, which is convenient for model processing.

5.3.2 *Establishment of ARIMA model*

The identification problem and order problem of the model is mainly to determine the three parameters of P, D, Q, and the order d of the difference is generally by observing the diagram, 1st or 2nd order. Here we mainly introduce the determination of p and q. Let's start with two functions

- **Tocorrelation function:** ACF describes the linear correlation between a time series observation and its past observations. The calculation formula is as follows:

$$ACF(k) = \rho_k = \frac{Cov(x_t, x_{t-k})}{Var(x_t)} \tag{5}$$

  where k represents the number of lags, and if k=2, it represents y_t and y_{t-2}
- **Partial autocorrelation function:** The partial autocorrelation function PACF describes the linear correlation between expected past observations of time series observations given intermediate observations.
- **Tailing and truncation:** Tailing refers to the monotonic decreasing or oscillating decay of the sequence at an exponential rate, while truncation refers to the sequence becoming very small from a certain point in time.

  The following conditions are generally considered to bias, which will be truncated by order D of the autocorrelation coefficient:

  1) At the initial D order, it is significantly greater than 2 times the standard deviation range.

  2) Almost 95% of the (biased) autocorrelation coefficients then fall within the 2-fold standard deviation range.

  3) And the process of decaying from a non-zero autocorrelation coefficient to a small fluctuation around zero is very abrupt.

  The following conditions are generally considered to bias, which can be called as autocorrelation coefficient tailing:

  1) If more than 5% of the sample (biased) autocorrelation coefficients fall outside the 2x standard deviation range.

2) Or the process of decaying from a significant non-zero (biased) autocorrelation coefficient to small value fluctuations is relatively slow or very continuous.

- **Determination of the p, q order:**

Table 2 The principle of determining p,q

model (sequence)	AR(p)	MA(q)	ARMA(p,q)
ACF	trailing	The qth post truncation	trailing
PACF	the pth last truncation	trailing	trailing

- **Parameter estimation:** Since the above parameter selection has a large subjectivity, we can determine the order of the model according to the information criterion function method. The prediction error is usually expressed in terms of the squared error, which is the sum of squared residuals. There are generally two types of guidelines:
- **AIC:** full name is Akaike Information Criterion

$$AIC = 2 * (\lambda) - 2ln(L(\theta)) \tag{6}$$

where λ represent the number of parameter, $L(\theta)$ represent the maximum likelihood function of the model

- **BIC:** The AIC guidelines have certain deficiencies. When the sample size is large, the information provided by the fitting error in the AIC criterion is amplified by the sample size, and the penalty factor for the number of parameters has nothing to do with the sample size (it is always 2), so when the sample size is large, the model selected using the AIC criterion does not converge with the real model, which usually contains more unknown parameters than the real model. BIC (Bayesian Information Criterion) makes up for the lack of AIC, and the calculation formula is as follows:

$$BIC = ln(n) * (\lambda) - 2ln(L(\theta)) \tag{7}$$

where the n represents the capacity of the sample.

5.3.3 *Solution of ARIMA model*

1. **Parameter solving:**

- **Determination of p,q:** First, we use the built-in function of the stats models library, we get p=3, q=2. However, the results obtained with this parameter are not ideal. In order for the model to have a better fitting effect on the temperature curve, according to the characteristics of the data, we fine-tune the values of p and q on the original basis, conduct a comparative analysis, and finally determine that **p=5, q=2**. The results are verified by the model, which proves that our parameter selection is feasible.
- **Determination of d:**d represents the order of the difference. In order to make the data more stable, we choose the second-order difference. That is to say,**d=2**.

2. **Model results:**

- Autoregressive model AR:
  The autoregressive model describes the relationship between the current value and the historical value, and uses the historical time data of the variable itself to predict itself. The autoregressive model must meet the requirements of stationarity. The autoregressive model first needs to determine an order p, which means that the current value is predicted by the historical value of several periods. The formula of five-order autoregressive model is defined as:

$$g_t = \varphi + \sum_{i=1}^{5} \lambda_i g_{t-i} + \varepsilon_t \tag{8}$$

  In the above formula, g_t is the current value, φ is the constant term, 5 is the order,λ_i is the autocorrelation coefficient, and ε_t is the error.
- Moving average model MA:
  The moving average model is used to eliminate the random fluctuation error of the autoregressive model. After substituting q=2,the formula is as follows:

$$g_t = \varphi + \sum_{i=1}^{2} \alpha_i \varepsilon_{t-i} + \varepsilon_t \tag{9}$$

  Generally speaking, when q is constant, MA (q) model must be stable.So for ARIMA model, only the stationarity of AR model needs to be checked.
- Autoregressive moving average model ARMA:
  The autoregressive moving average model is to try to combine the autoregressive process AR with the moving average Average process MA is combined to jointly simulate the generation of existing time series sample data The

expression of the model of random process is as follows:

$$g_t = \varphi + \sum_{i=1}^{5} \lambda_i g_{t-i} + \sum_{i=1}^{2} \alpha_i \varepsilon_{t-i} + \varepsilon_t \tag{10}$$

- Differential autoregressive moving average model ARIMA:
  Combining autoregressive model, moving average model and difference method, we get ARIMA (5, 2, 2).The past and future global temperature levels can be solved by substituting the formula(10).

3. **The solution of problem b:**

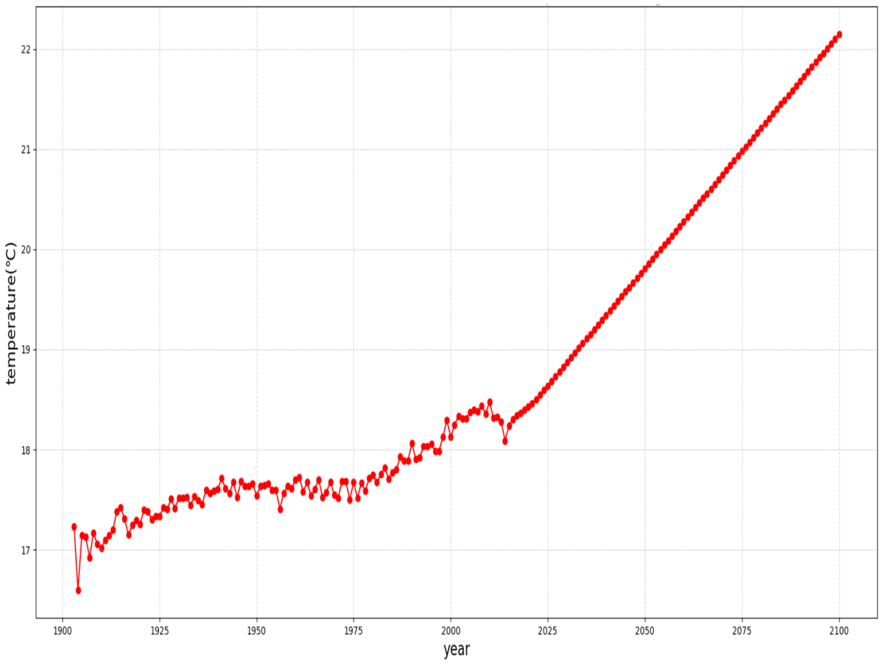


Figure 6 ARIMA's description of past and future global temperatures(°C)

- **The description of the past:**
  The years 1900-2022 in Figure 6 are descriptions of the past. It can be concluded that the global temperature fluctuation is rising, which is more in line with the actual situation.
- **Forecast for the future**
  ARIMA model's prediction of future temperature is close to linear. In order to more clearly reflect the prediction of future trends, we have made a separate picture (Figure 7). It can be concluded that the global temperature will increase linearly in the future.

4. **The solution of problem c:**

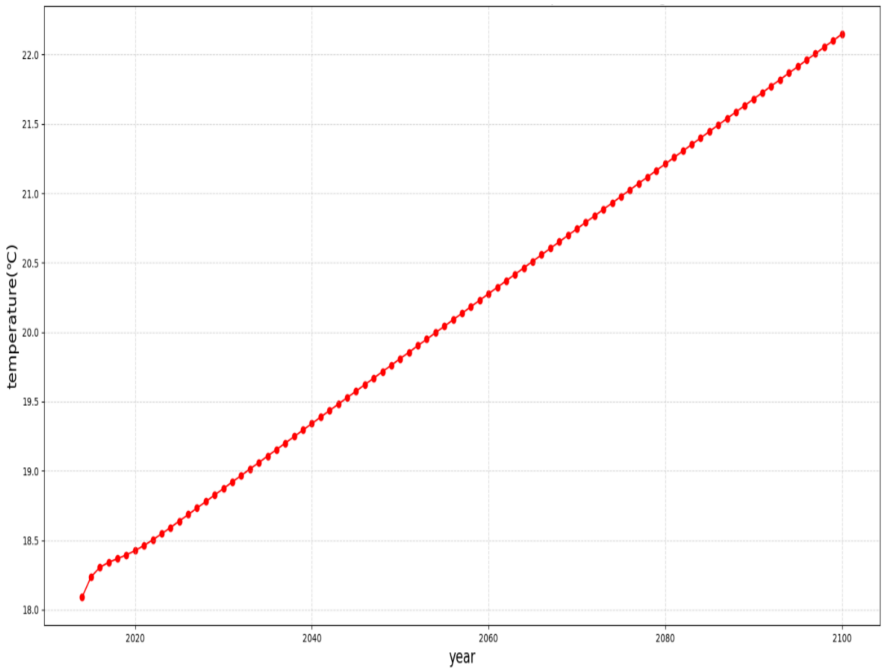


Figure 7 Prediction of global future temperature by ARIMA model(℃)

It can be seen from Figure 7 that ARIMA model predicts that the global temperature will reach 19.809 ℃ in 2050, and will reach 20 ℃ in 2055. The global temperature will reach 22.147 ℃ in 2100.

5.4 Transformer Model to describe and predict

The model will be used to solve problems B and C in the description and prediction.

For sequence models, the traditional neural network structure has problems such as difficulty in dealing with long-term dependence and low computational efficiency. Although researchers have proposed LSTM, attention mechanism, CNN binding RNN and other methods, they still cannot effectively solve these problems. Transformer is a new neural network structure, which is only based on the attention mechanism, and abandons the traditional loop or convolutional neural network structure. Besides, it has stronger long-term dependency modeling capabilities and works better on long sequences. On the one hand, it solves the problem that RNN cannot completely eliminate gradient vanishing and gradient explosion for long sequences, and on the other hand, it models long-term dependence and short-term dependence. Different heads in Multi-head Attention can focus on different patterns. Based on the above advantages, we chose TransForex to predict future temperature data.

5.4.1 *Data pre-processing*

Before being sent to the transformer network for prediction, we first normalized the data, which can effectively improve the accuracy of network prediction. Here we process the known time and the corresponding average temperature, and process them into data embedding. Embedding is actually a feature of the data. Since TransFormer itself cannot use the word order information, we introduce pos embedding here, and add it to the data itself to get the final data embedding.

The following is the positional coding formula

$$\begin{aligned} PE_{(pos,2i)} &= \sin\left(pos/10000^{2i/\eta}\right) \\ PE_{(pos,2i+1)} &= \cos\left(pos/10000^{2i/\eta}\right) \end{aligned} \tag{11}$$

η represents the dimension of PE.

5.4.2 *Establishment of TransFormer model*

The overall framework consists of 6 encoders and 6 decoders (as can be seen in figure 8(e)), and after the preprocessed data is entered into the encoder in parallel, the corresponding prediction data is output from behind the decoder after the model training prediction

First, the various structures are introduced inside the encoder and decoder and then effectively combined.

We will introduce **Self-Attention**, **Multi-Head Attention**, **Add Norm**, **Feed Forward**, and **Self-Attention** in turn

- **Self Attention:** Before introducing Multi-Head Attention, let's introduce basic self-attention(as can be seen in figure 8(b))
  To obtain Q, K, V, we take the data embedding matrix X as input and use the linear transformation matrix WQ, WK, WV, to calculate the specific Q, K, V After obtaining Q, K, V, you can find the output of Self-Attention according to the following formula

$$Attention(Q,K,V) = softmax(\frac{QK^T}{sqrt(d_k)})V \tag{12}$$

  d_k is the column of the matrix of Q and K, that is the vector dimension
- **Multi-Head Attention:** The essence is formed by multiple combinations of Self-Attention, first passing input X into h different Self-Attention respectively, and calculating the h output matrix Z. The figure below shows the situation when h=8,

at which point you get 8 output matrix Z. After getting another 8 output matrices Z1 to Z8, Multi-Head Attention stitches them together and passes them into a Linear layer to get the final output Z of Multi-Head Attention. You can see that the matrix Z output by Multi-Head Attention is the same dimension as the matrix X it inputs (figure 8(f)).

- **Add Norm:** The Add Norm layer consists of two parts, Add and Norm, and its calculation formula is as follows:

$$\begin{gathered} LayerNorm(X + MultiHeadAttention(X)) \\ LayerNorm(X + FeedForward(X)) \end{gathered} \tag{13}$$

where X represents the input of Multi-Head Attention or Feed Forward.
Add stands for X+Multi-HeadAttention, which is a residual connection that is usually used to solve the problem of multi-layer network training, allowing the network to focus only on the current differences, which is often used in ResNet(figure 8(c))
Norm stands for Layer Normalization, which is usually used for RNN structures, and Layer Normalization converts the input of each layer of neurons to the same mean-variance, which can speed up convergence.

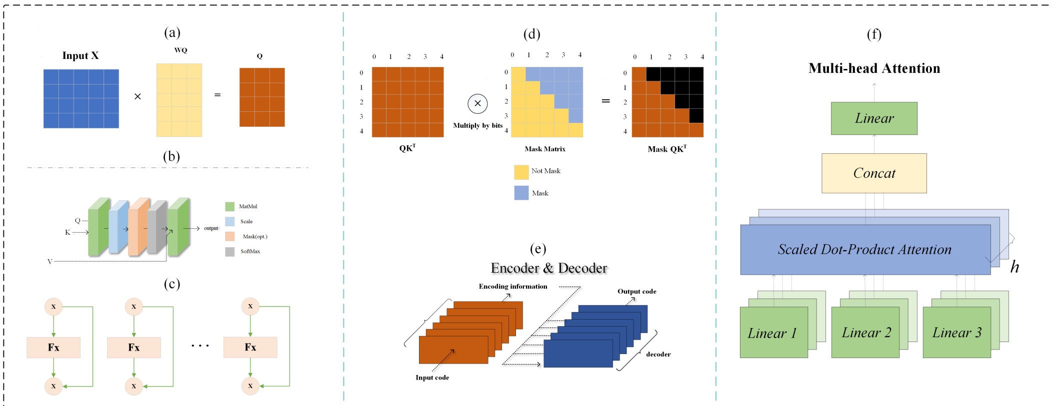


Figure 8 The whole workflow of each part of the transformer, which includes six effective blocks (from a to f).

- **Feed Forward:** The Feed Forward layer is relatively simple, it is a two-layer fully connected layer, the activation function of the first layer is Relu, and the second layer does not use the activation function, the corresponding formula is as follows.

$$max(0, XW_1 + b_1,)W_2 + b_2 \tag{14}$$

where X is the input, and the dimension of the output matrix obtained by Feed Forward is consistent with X.

- **Combine Encoder:** Finally, with the Multi-Head Attention, Feed Forward, Add Norm described above, we can construct an Encoder block that receives an input matrix of $X_{(n \times d)}$ and outputs a matrix $O_{(n \times d)}$. Multiple encoder blocks can be superimposed to form an encoder.

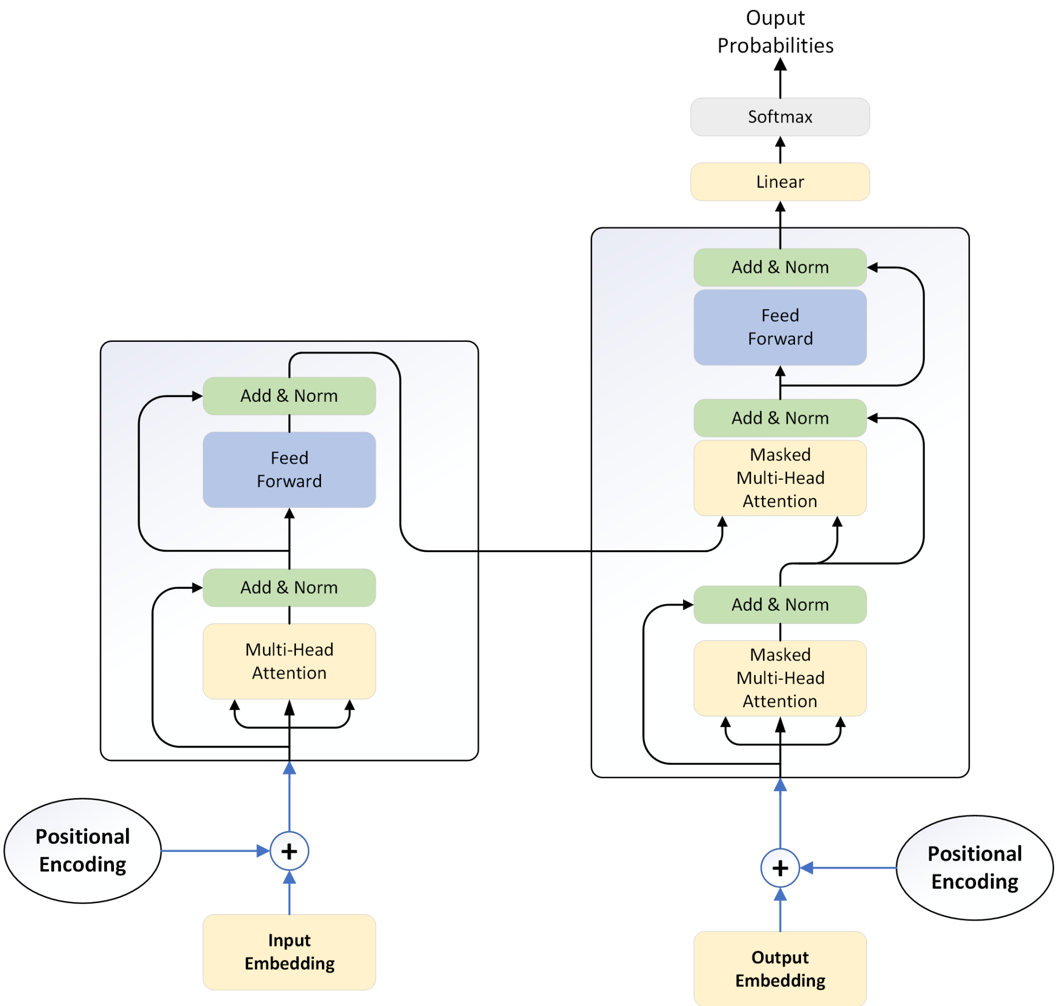


Figure 9 The detailed structure of the encoder and decoder.

- **decoder:** Regarding the decoder, the internal structure used is roughly the same as in the encoder, with the obvious difference being the Masked Multi-Head Attention
- **Masked Multi-Head Attention** The first Multi-Head Attention of the Decoder block uses the Masked operation because it is predicted sequentially in the prediction process, that is, the i+1st temperature can be predicted after the i-th temperature is predicted. The MASKED operation prevents the i-th temperature from knowing after the i+1 temperature. (mask matrix see figure d)

5.4.3 *Solution of TransFormer model*

The following part of the model is established for problem b

Describe the past:

In the process of transformer model training, we draw the annual global average temperature trend chart, as shown in the figure 10:

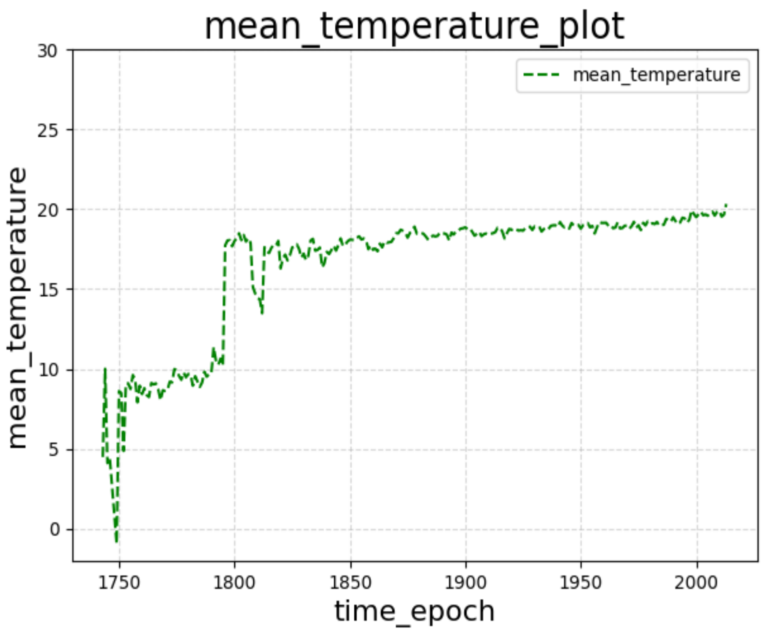


Figure 10 The global mean temperature from 1750 to 2010.

Through the analysis of the global average temperature map obtained during the training process, we can analyze: Between 1750 and 1830, due to the small amount of data provided and uneven regional distribution, the global average temperature described in the line chart during this period was inaccurate. Between 1830 and 2010, because the data were distributed evenly around the world, the icon can represent the global average temperature at that time. From the chart, we can draw a conclusion that between 1830-2010, the global average temperature grew slowly and remained below 20 ° C.

Describe the future:

We use the time series prediction of transformer model to predict the global average temperature of each month in the next ten years: 2023 2033. The prediction results are shown in the figure below:

We can see that in the prediction results of the transformer model, the annual average temperature from 2023 to 2032 is about 20°-24°C, and the average temperature in most months is about 20-22°C.

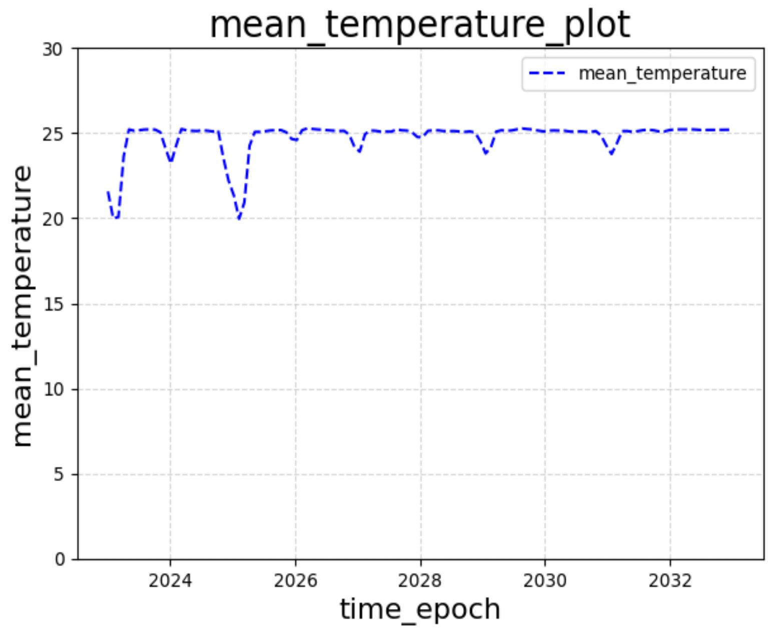


Figure 11 The global predicted mean temperature from 1750 to 2010.

The following part of the model is established for problem c:

Global temperature forecast from 2050.1-2050.12 and 2100.1-2100.12:

As shown in the figure below, we can see that the temperature in 2050 and 2100 will be about 25 ° C under the prediction of the transformer model, and there will be some slight fluctuations in the time series.

Through the time prediction of the transformer model between 2013-2050, we searched for the point where the global average temperature reached 20 ° C, and the results are as follows:

As shown in the figure, the result is about 2049, so we suggest that the average temperature can reach 20 ° C.

5.5 Evaluation of the Model

In this section, we use RMSE and MAE as metrics to evaluate model accuracy RMSE, full name Root Mean Square Error, represents the standard deviation of the sample for the difference between the predicted value and the observed value. The root means the square error is used to indicate how dispersed the sample is.

$$RMSE(X, h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(x_i) - y_i)^2} \tag{15}$$

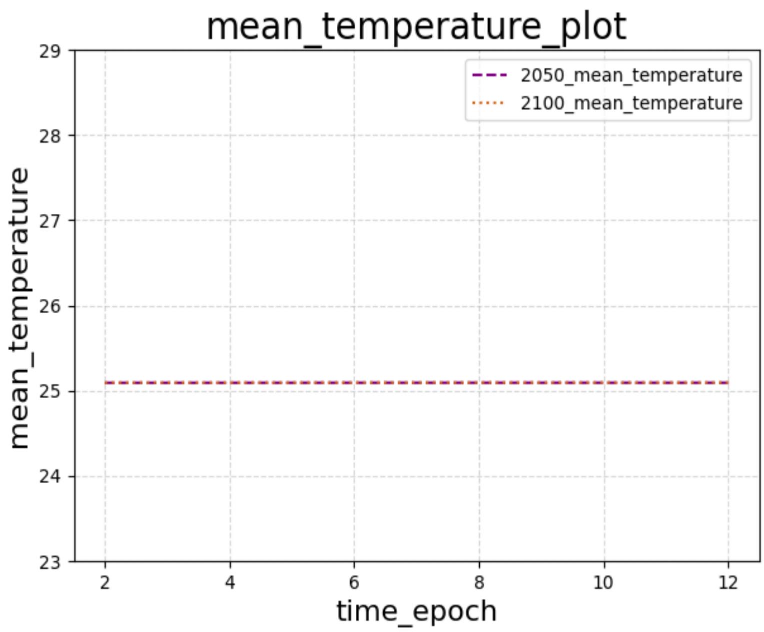


Figure 12 The global monthly predicted mean temperature in 2050 and 2100.

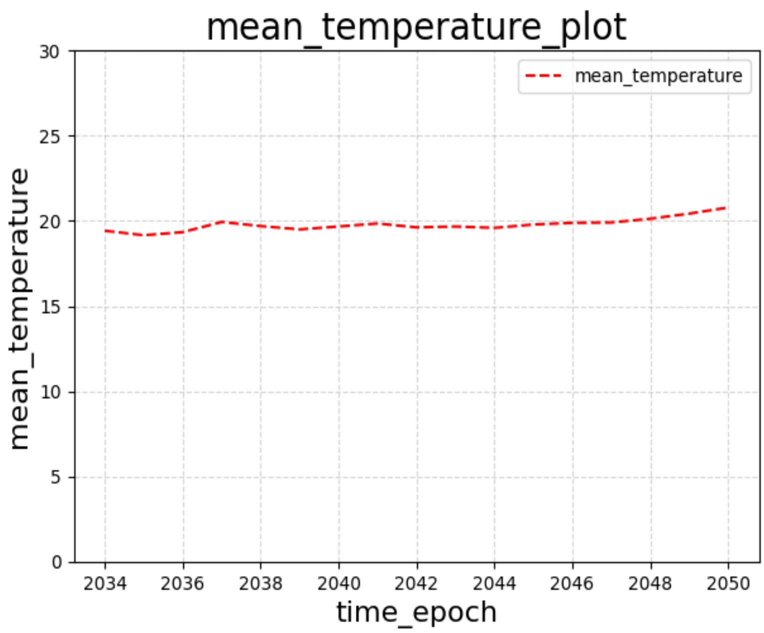


Figure 13 The global monthly predicted mean temperature from 2034 and 2050.

MAE, full name is Mean Absolute Error, which represents the average of the absolute error between the predicted value and the observed value

$$MAE(X, h) = \frac{1}{m} \sum_{i=1}^{m} |h(x_i) - y_i| \tag{16}$$

$h(x_i)$ stands for predicted value, y_i stands for true value

VI. Model and solution of problem two

6.1 Question a

In this question, we mainly use statistical correlation coefficient analysis to analyze the correlation between location, time, and temperature. Here we first introduce the three coefficients of correlation analysis.

6.1.1 *Data pre-processing*

In order to solve the relationship among the global temperature, time, and location-related variables, we collected the data sets of the temperature changes in the time dimension of each latitude, the monthly average temperature of each region, carbon dioxide emissions, and other data sets on the basis of the data provided by the topic, so as to facilitate the subsequent correlation analysis.

6.1.2 *Establishment of Correlation analysis Model*

Three correlation coefficients are summarized as follows:

- **Person correlation coefficient:** The Pearson correlation coefficient is a measure of the linear correlation between two variables X and Y, with values between -1 and 1, and the larger the absolute value, the more correlated the two variables.
- **Formula derivation:**

$$\rho_{x,y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{17}$$

Here, the numerator CoV represents the covariance and the denominator represents the standard deviation (taking two variables as examples):

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1} \tag{18}$$

So, the Person correlation coefficient is:

$$\rho_p = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\left[\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2\right]^{\frac{1}{2}}} \tag{19}$$

- **Spearman correlation coefficient:** The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the hierarchical variables. For samples with a sample size of n, n raw data are converted to hierarchical data with a correlation coefficient ρ.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{20}$$

In practice, the connection between variables is irrelevant, so can be calculated in a simple step. The difference between the ranks of the two observed variables can be calculated by the formula (21).

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n\left(n^2 - 1\right)} \tag{21}$$

- **Kendall correlation coefficient:** It is based on the rank of the data object to evaluate the correlation (strength and direction) between two (random variables). The target object of the analysis should be an ordered categorical variable, such as rank, age group, obesity grade (severely obese, moderately obese, mildly obese, not obese), etc.
  Based on the relationship between sample data pairs, the strength of the correlation coefficient is analyzed, and the data pairs can be divided into Concordant and Discordant.
- **Model choose:** Since in the ARIMA model of question 1, we know that global temperature changes have a good linear correlation with factors such as time, the growth is stable, and there are few phenomena with large fluctuations, so we choose the person correlation coefficient to analyze, which can make the data analysis more efficient.

6.1.3 *Solution of Correlation analysis Model*

Since in the ARIMA model of question 1, we know that global temperature changes have a good linear correlation with factors such as time, and the growth is stable, and there are few phenomena with large fluctuations, so we choose the person correlation coefficient to analyze, which can make the data analysis more efficient

Adopt correlation analysis, draw correlation coefficient matrix and adopt Person correlation coefficient. The resulting correlation coefficient is shown in the figure 14:

Through the description of the correlation coefficient matrix, it can be found that the correlation coefficient between temperature and average temperature is as high as 0.84, which means that there is a strong negative correlation between the average temperature and location latitude, which is also in line with our common sense. The closer to the

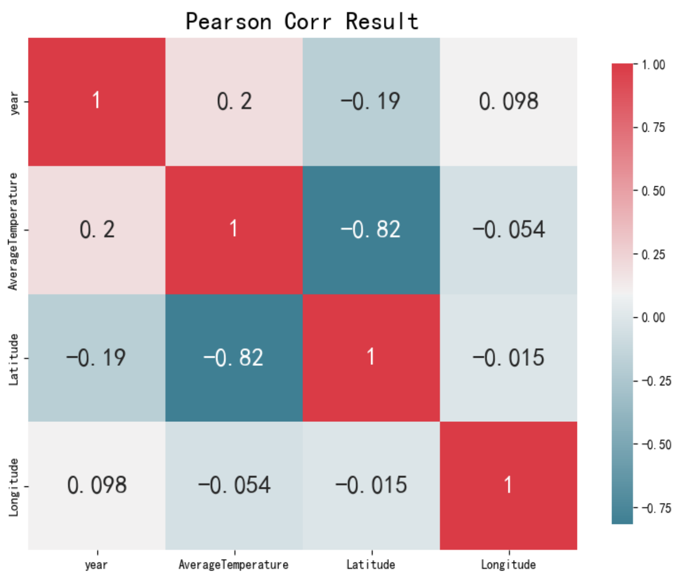


Figure 14 The Pearson correlation coefficient matrix result.

pole, the lower the average temperature. This also proves the correctness of our model analysis. At the same time, there is a weak positive correlation between time and average temperature. In fact, the average temperature will rise slowly due to global warming. At the same time, we conducted a second verification of the analysis results of the model through the additionally collected data sets of temperature changes in the time dimension of each latitude, as shown in figure 15.

We can find that the average temperature of the equator is the highest in the past 30 years, followed by the average temperature of the southern temperate zone and the northern temperate zone. This also verified the conclusion that latitude is negatively correlated with temperature and altitude. At the same time, we can also verify that between 1900 and 2020, the temperature increased slowly with time. In addition, we also found that the average air temperature in the southern hemisphere is slightly higher than that in the northern hemisphere through the experimental results. We think this is because the southern hemisphere has a larger ocean area, resulting in a smaller annual range in the southern hemisphere.

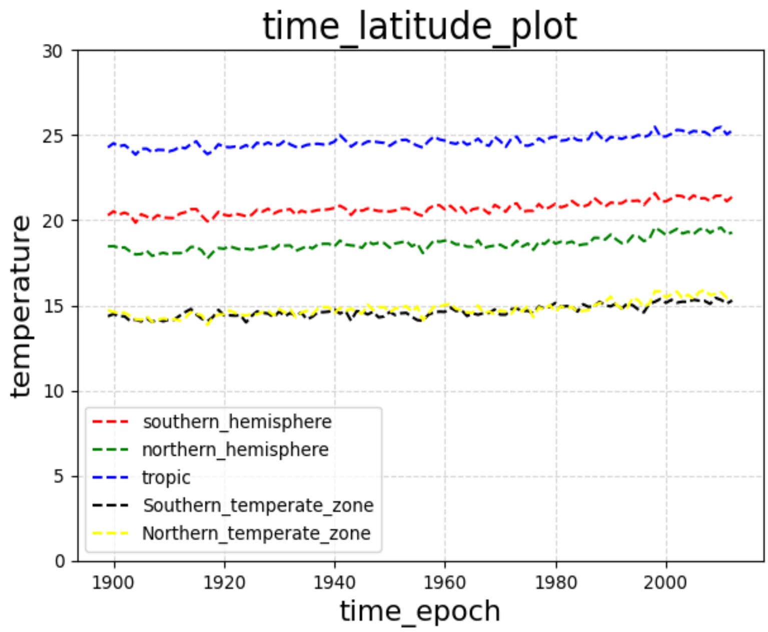


Figure 15 Average temperature trends of different temperature zones from 1900 to 2010.

6.2 Question b

6.2.1 *Data pre-porcessing*

In order to analyze the impact of natural disaster factors on global temperature, we collected data sets of annual average carbon dioxide emissions, annual volcanic eruptions, annual earthquake frequency, and annual typhoon landings.

6.2.2 *Establishment of model*

For this problem, we need to calculate the impact of the annual average carbon dioxide, the annual average number of volcanic eruptions, and other factors on the global temperature. Therefore, we also use correlation coefficients to evaluate the impact of global temperature. Because the data we collected may not be continuous and do not meet the normal distribution, the overall linear correlation is not high, and there is no relationship between grades and categories. Secondly, through data processing and model solving, we found that the Spearman correlation coefficient model was more in line with the actual situation, so we chose the Spearman correlation coefficient for correlation analysis.

6.2.3 *Solution of model*

Through the calculation of the Spearman correlation coefficient, we get the following results:

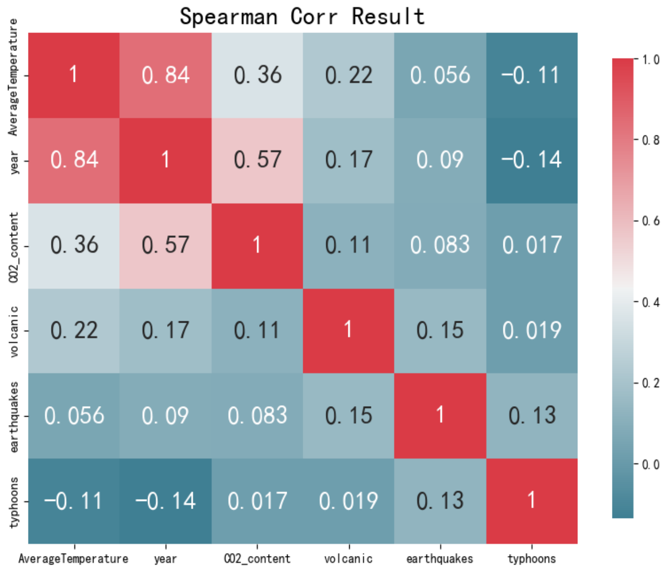


Figure 16 The Spearman correlation coefficient matrix result.

From the output results, we can see that the correlation coefficient between time and the global average temperature is 0.84, and the correlation coefficient between carbon dioxide and the global average temperature is 0.36. We can draw a conclusion that time and carbon dioxide are the factors that affect the annual average temperature. That is, the global average temperature rises with the annual average time, showing a strong positive correlation. At the same time, natural disasters, volcanic eruptions, and carbon dioxide will affect the global average temperature to some extent. This also conforms to the specific facts, which also verifies the correctness of our model. In addition, we can also conclude that the annual average earthquake frequency and typhoons have little impact on the global average temperature.

6.3 Question c

6.3.1 *Data pre-processing*

In order to find out the factors that have the greatest impact on global temperature change among the above factors, we use the relevant data sets of natural disasters collected in (b), combined with the relevant data sets provided in the topic. On the basis of the correlation coefficient matrix in (b), calculate their principal component analysis results on global temperature change, so as to obtain the contribution rate of each factor to global temperature change.

6.3.2 *Establishment of PCA Model*

PCA (Principal Component Analysis) is a multivariate statistical method, it is one of the most commonly used dimensionality reduction methods, through the orthogonal transformation of a set of potentially related variable data, into a set of linearly uncorrelated variables, the transformed variable is called principal component.

There are two methods that can be used for PCA, feature decomposition or singular value decomposition.

- **PCA principle implementation steps:** Solution methods: covariance matrix, correlation coefficient matrix, singular value decomposition, alternating least squares method
- **Step1: Standardize sample data**
  Because the dimensions of different indicators are usually not identical, in order to be comparable between indicators, the dimensions between indicators must be eliminated, and the standardized formula is:

$$X_{\mathrm{ij}} = \frac{Y_{\mathrm{ij}} - \bar{Y}_j}{S_j} = \frac{Y_{\mathrm{ij}} - \frac{1}{m}\sum_{i=1}^{m} Y_{\mathrm{ij}}}{\frac{1}{m-1}\sum_{i=1}^{m}\left(Y_{\mathrm{ij}} - \bar{Y}_j\right)^2}, j = 1, 2, 3, \ldots n \tag{22}$$

  Normalized matrix after normalization:

$$X = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{11} \\ \ldots & \ldots & \ldots & \ldots \\ X_{n1} & X_{\mathrm{m2}} & \ldots & X_{\mathrm{mn}} \end{bmatrix} \tag{23}$$

- **Step2: The correlation coefficient is calculated to obtain the correlation coefficient matrix**
  After standardization, calculate the correlation relationship between every two indicators of this matrix and obtain the correlation coefficient matrix R, that is, the

covariance matrix of n indicators:

$$R = \frac{1}{m-1} X'X = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{21} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{\mathrm{n}2} & \dots & r_{\mathrm{nn}} \end{bmatrix} \tag{24}$$

$$r_{\mathrm{ij}} = \frac{1}{m-1} \sum_{k=1}^{m} X_{\mathrm{ik}} X_{\mathrm{jk}} (i, j = 1, 2, 3, \dots, n)$$

- **Step3: Calculate the eigenroots of matrix R and the corresponding eigenvectors**
  It is possible to obtain n non-negative eigenroots and the corresponding n unitized eigenvectors to form an orthogonal matrix, denoted a:

$$a = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{21} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{\mathrm{n}2} & \dots & a_{\mathrm{nn}} \end{bmatrix} \tag{25}$$

- **Step4: Calculate principal components:**
  That is, as shown in the formula:

$$\begin{bmatrix} Z_1 & Z_2 & \dots & Z_n \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{21} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \tag{26}$$

- **Step5: Principal ingredient selection:** The proportion of the total variance belonging to the principal component Zi is:

$$\frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j} \tag{27}$$

  The contribution rate is called the principal component Zi. We selected several characteristics with a large contribution rate.

6.3.3 *Solution of PCA Model*

Through the principal component analysis and the corresponding factor analysis of the dataset, we achieved the following results (figure 17 and 18):

Through the diagram, we can get the analysis results of principal component contribution rate among variables and the analysis of influencing factors. Due to the slow rising trend of global temperature and the lack of data, the calculated impact factors

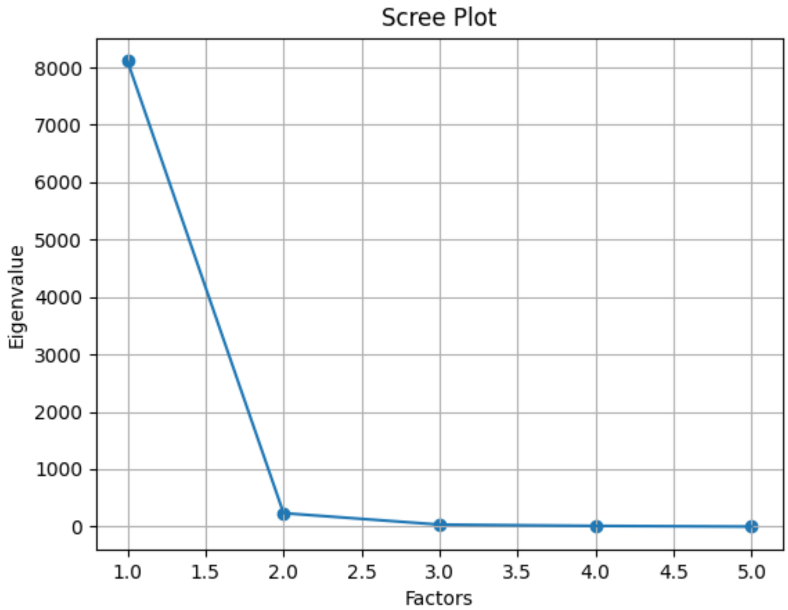


Figure 17 Results of principal component analysis.

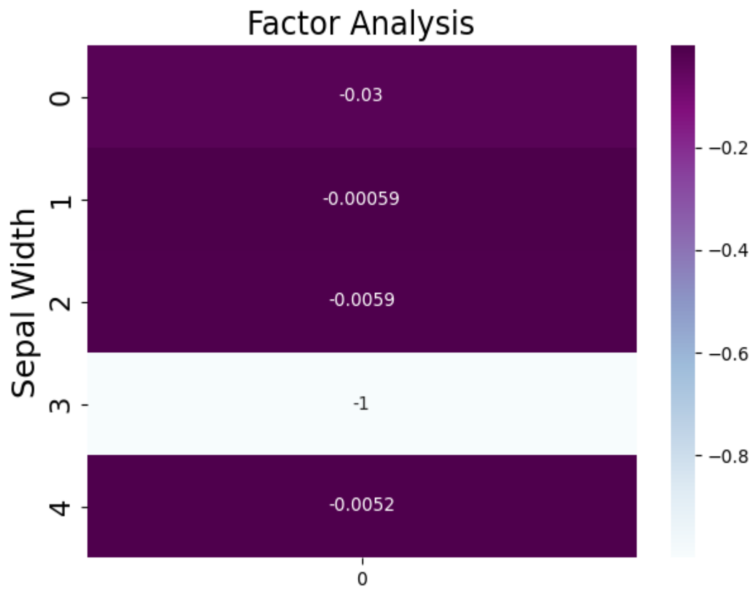


Figure 18 Influencing factor thermal diagram of principal component analysis.

are small. In addition, we can find that in addition to time, the index with the subscript of 0, namely carbon dioxide emissions, has the greatest impact on global temperature. The second is volcanic eruption. This shows that if we want to control global warming at its root, we must strictly control carbon emissions, and we will put forward detailed suggestions in the issue of d.

6.4 Question d

Global Warming, also called Climate Change, is one of the most difficult issues that the world is facing today. Global warming occurs due to the effect of greenhouse gases such as CO2, methane, nitrous oxide and fluorinated gases.

6.4.1 *Suggestion*

CO2 is generated by the combustion of fossil fuels such as coal, natural gas, and oil for power production and transportation needs. While electricity is clean at the point of use, its generation produces over 40% of all energy-related carbon emissions. The effective way to reduce CO2 emissions is to reduce fossil fuel consumption and shift to renewable energy sources like nuclear, solar, wind, hydroelectric, biomass, geothermal and tidal waves.

Nuclear power plants do not produce greenhouse gas emissions during their operation, and only very low emissions over their full life cycle. Nuclear power contributes 11% of the global electricity generated which amounts to one-third of the world's carbon-free electricity. Nuclear power can meet fluctuations in energy demand and provide stability to electric grids with a high share of variable renewable sources which otherwise have to depend on fossil fuels. Combined with smart power grids, nuclear energy can help the transition to low carbon electricity sources and ensure reliable, stable and sustainable energy supplies. Water Cooled Reactors (WCRs) have been the cornerstone of the nuclear industry in the 20th century, with 442 reactors in operation world-wide. FBRs represent a technological leap beyond WCRs, and are poised to become the mainstream. FBRs breed more fuel than they consume and allow more efficient use of uranium resources.

Fusion reactors are forecast to be the future of green energy and an unlimited resource. Nuclear fission and nuclear fusion energy do not emit carbon dioxide or other greenhouse gases into the atmosphere, and would be an abundant source of low carbon energy for energy security. Climate change is an urgent crisis that requires greater adoption and expansion of nuclear energy [1].

Table 3 Indicator system of Saihanba and Beijing

Impact indicators of Saihanba		Impact indicators of Beijing
wind and sand resistance	ecological environment	wind and sand resistance
Average visibility	Evaporation	Average visibility
Wind speed	Precipitation	Wind speed
Wind erosion and climate erosion	Soil temperature	Wind erosion and climate erosion
Net intensity of short-wave radiation	Soil humidity	Net intensity of short-wave radiation

VII. Non-technical report for the APMCM

Dear Organizeling Competition:

It's a great honor to convey our ideas to you!

Our survey results show that the main reasons for temperature change are time, location, carbon emissions and natural disasters. From 1743 to 2022, the global temperature generally showed an upward trend. With the increase of latitude, the region gradually moved from the tropics to temperate and cold zones, and the temperature showed a downward trend Under the influence of some natural disasters and human social activities, carbon emissions increased, and global temperature showed an upward trend.

1. **Analysis of the causes of global temperature rise from 1743 to 2022:** First, mankind discharged a large amount of greenhouse gases into the atmosphere during the industrial revolution. Second, human destruction of the earth's vegetation. This conclusion is also verified in the Solution of Correlation analysis Model (6.1.3) in 2 (a).

   Our suggestion: All countries should work together to improve energy efficiency, develop new energy, and save energy and emissions.
2. **Impact of latitude on air temperature:**

   The temperature of the earth decreases from the equator to the poles. This is because the higher the latitude, the less time the sun shines directly, the less heat it receives, and the colder it is. This conclusion is also verified in the Solution of Correlation analysis Model (6.1.3) in Question a.
3. **Impact of carbon emissions on temperature:**

   According to the correlation analysis, carbon emissions have a great impact on the earth temperature. This is because greenhouse gases such as carbon dioxide in the atmosphere will hinder long wave radiation, resulting in abnormal heat emission, which makes the earth's temperature rise. This conclusion is also verified in the Solution of Model (6.2.3) of 2 (b).

Our suggestions: 1.Reduce greenhouse gas emissions 2.Afforestation to increase global forest coverage

4. **Volcanic eruption makes the temperature rise:**
   At the beginning of volcanic eruption, sulfur dioxide, as a greenhouse gas, mainly led to the increase of surface temperature around the crater.

VIII. Sensitivity analysis

8.1 ARIMA Sensitivity analysis

In the ARIMA model, the most important parameters are p and q. As mentioned above, the parameters given by the built-in function are p=4, and q=3. However, because the effect is not ideal, we made several parameter selections based on the built-in function and finally determined the optimal scheme: p=5, q=2.

Here, we list the MSE and RMSE obtained from some parameters, and draw a parallel histogram for comparison.

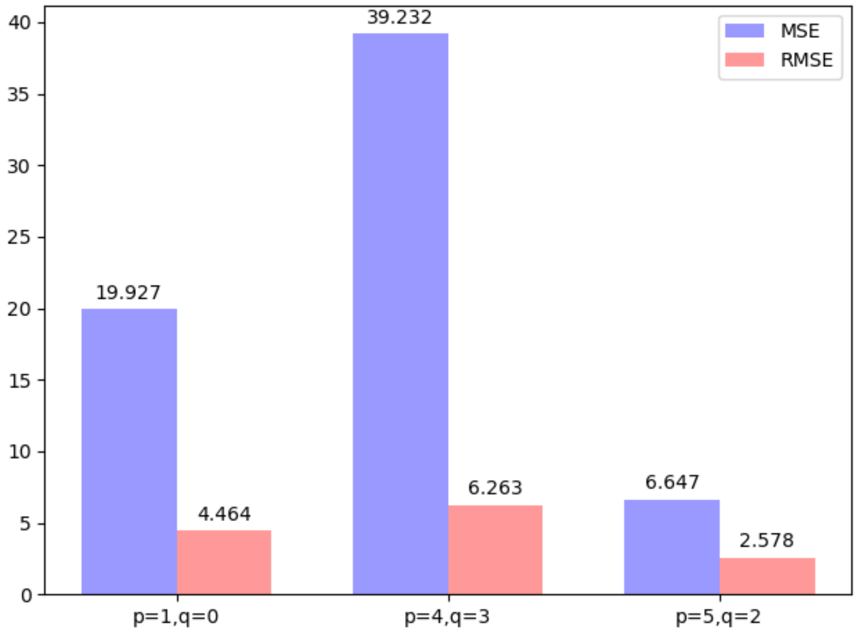


Figure 19 MSE, RMSE obtained by ARIMA different parameter selection

8.2 TransFormer Sensitivity analysis

For the sensitivity analysis of the transformer model, we adopt the method of parameter adjustment, and input windows output Window and batch size, and other parameters are adjusted and compared. The results are as follows:

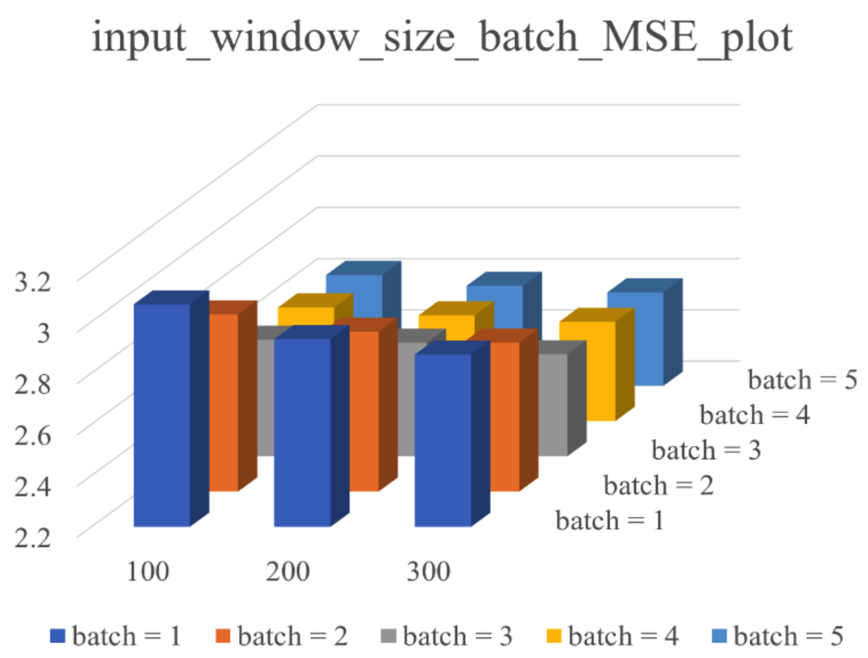


Figure 20 Ablation experiment results.

By comparing the relationship between various parameters and the MSE change trend, we can draw a conclusion that our losses will decrease with the increase in input window size, and at the same time, our losses will also decrease with the increase in batch size. From this, we can conclude that our model has good sensitivity and robustness. At the same time, we can also conclude that in order to make our model prediction results more accurate, we can increase the input window size and batch size to a certain extent.

In addition, we can find that an important way to improve the transformer model is to increase the batch size within the tolerance of the network structure.

IX. Evaluation, Improvement and Promotion of the Model

9.1 Evaluation of the Model

- **Advantage:** 1)The ARIMA model, using differential techniques, cleverly eliminates changes in the level of the time series, eliminates trends and seasonality, and removes the influence of seasonal components, thereby stabilizing the average value of the time series. At the same time, the predicted RMSE is 2.578 and the MAE is 6.64, and the overall prediction effect is outstanding. 2) TransFormer, good interpretability, eliminate the "black box properties" of general neural networks. Good generalization, the use of a multi-head attention mechanism enables the network to better capture context information; The network structure parameters are large, the prediction effect is good, MAE is only 0.77, and the difference between the prediction result and the real result is small, relatively small; the robustness is strong, RMSE is only 1.7, in the case of fine-tuning parameters, the prediction result will not appear large deviation. 3)We use both traditional statistical analysis prediction models and artificial intelligence neural networks to make the effect more convincing and reliable
- **Shortcoming:** 1)In reality, there are many indicators that affect global temperature change, and our model only uses a few more important indicators, which may not fully reflect the characteristics of global temperature change. 2)Although we have done the interpretability of the transformer model, there are still some areas that cannot be perfectly explained, which is also a big problem that the neural network has not solved

9.2 Improvement and promotion of the Model

(1) Based on the factors influencing the global temperature in the second question, since a single time series model cannot predict all climate factors, the improvement can be achieved by establishing LSTM long-term short-term memory

The network model predicts the future carbon dioxide and uses the prediction results as characteristic data that affect temperature changes to help predict future continental average temperature changes. This allows the predicted temperature to be more accurate

LSTM is a time-loop neural network, which can solve the long-term dependence problem of general RNN (recurrent neural network).

(2) Because after the industrial revolution, human civilization stepped into industrial civilization, and with the further destruction of the natural environment, the invention and mass use of various industrial machines led to a large amount of carbon dioxide emissions, making the rate of global warming greatly accelerated, and the global temperature increase before and after the industrial revolution will be different, so we will cut the data before and after the industrial revolution, and use the data before the industrial revolution as pre-training data to enhance model recognition, and at the same time input the data after the revolution to train the model again. This allows the model to fit global temperature changes more accurately

X. References

[1] Zhang Wentong, SPSS11 statistical analysis course [M]. Beijing Hope Electronic Publishing House, Beijing, 2002, 197-201.

[2] Daniel A. Lashof& Dilip R. Ahuja,Relative contributions of greenhouse gas emissions to global warming

XI. Appendix