Problem Chosen	2022 APMCM Summary Sheet	Team Control Number
C		C2200789

Prediction and correlation analysis of global temperature conditions

Summary

We established that the current number of terrible temperatures around the world, global warming has become a public concern. The purpose of this paper is to establish the most accurate global temperature prediction model, and to analyze the development trend of future temperature and the key factors affecting global temperature, so as to put forward reasonable suggestions for curbing and mitigating global warming.

In sub-question a of TASK 1,We agree with the statement in the title, and conducted MK mutation test on the data, and found that the two branches of the curve intersected in March 2022, which proved that the mutation occurred at that time point, and carried out a visual processing for easy comparison.

In sub-question b of TASK 1,We built the LSTM neural network prediction model and grey system theory GMP model, respectively describing the previous temperature and predicting the future temperature.

In sub-question c of TASK 1,We use the model established by 1.b to make predictions respectively, and come to the conclusion that the global temperature will reach 20 degrees Celsius in 2050. Analyze when the temperature will reach 20 degrees Celsius and determine the range of years in which 20 degrees Celsius will occur.

In sub-question d of TASK 1,We have analyzed the prediction results and use conditions of the two models, and believe that LSTM neural network model has more stable prediction ability, and the prediction results are more accurate than GM grey prediction model.

In sub-question a of TASK 2, We use Pearson correlation coefficient analysis to analyze the correlation between temperature and altitude, carbon dioxide and location, to determine whether temperature is closely related to the above factors, and to further infer its impact on global temperature

In sub-question b of TASK 2, We analyze the global temperature change by analyzing the CO2 produced by forest fires, and also use the Pearson correlation coefficient analysis method to conclude that forest fires have an impact on global temperature change

In sub-question c of TASK 2,Through Pearson correlation analysis, we believe that the concentration of carbon dioxide is the main cause of global temperature change.

In sub-question d of TASK 2, We believe that the key to curbing global warming lies in the concentration of carbon dioxide, so some relevant measures should be taken to control the emission of carbon dioxide.

Finally, through the statistics and prediction of global temperature and the correlation analysis of various factors affecting it, we find that the most fundamental reason is the influence of the concentration of carbon dioxide. If we want to curb global warming, we must control carbon emissions.

Keywords:LSTM neural network , grey system theory GMP model,Pearson correlation coefficient

Contents

1 Introduction

1.1 Problem Background

Since this year, we have seen a large amount of amazing temperature reports. The reality that the earth is burning is beyond doubt. Following the terrible high temperature in these regions from the end of June to the beginning of July, Italy once again set a European temperature record, reaching an astonishing 48.8 řC, and many countries declared a state of emergency. Global climate warming is a phenomenon related to nature. It is due to the continuous accumulation of greenhouse effect, which leads to the imbalance of energy absorbed and emitted by the earth atmosphere system, and the continuous accumulation of energy in the earth atmosphere system, leading to temperature rise and global climate warming. The team are provided with the following tasks:

- 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?
- Based on the historical data, please build two or more mathematical models to describe the past and predict the future global temperature level.
- 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 řC? If not in 2050 or 2100, when will the average temperature of observation points in your prediction models reach 20.00 řC?
- Which model you built in 1(b) do you consider most accurate? Why?
- Build a model to analyze the relationship between global temperature, time and place, and explain the relationship.
- 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?
- What do you think is the main reason that affects the global temperature change?
- Do you think there are some measures to curb or slow down global warming?

1.2 Our work

1. We used python to process the data, and conducted MK mutation test on the data, and found that the two branches of the curve intersected in March 2022, which proved that the mutation occurred at that time point, and carried out a visual processing on it for easy comparison.

2. We set up the LSTM time series prediction model and the grey system theory GMP model respectively, according to their different characteristics for the future temperature forecast. At the same time, we analyze the advantages and disadvantages of the two models and compare them. We think that the former model is more accurate and has stronger practicability.

3. Pearson correlation coefficient analysis is used to analyze the temperature and other variables, and it is clear that the carbon dioxide concentration is the most critical influence on global temperature, so we give suggestions to reduce carbon emissions.

2 Assumptions and Notations

2.1 Assumptions

- Ignore the effect on the final M-K mutation test of removing one to two years of data residuals from the two hundred years of temperature change already collected.
- Assume that the Earth's ecosystem remains stable for the next 100 years without significant deviations from the present.
- Assume that there are no major breakthroughs in human science and technology, and that the energy sources currently used are still the main energy suppliers.
- Assume that the data collected in this paper are accurate and realistic, and that the errors are small enough to accurately reflect the real changes in global temperature.
- Assume that there are no huge natural disasters or other factors affecting global temperature change in the next 100 years.

2.2 Notations

Notations	**Definition**
S	Output trend check indicators
UFk	Sequence Indicates the order column of a time series
UBk	The order column of a reverse time series
r	Pearson correlation coefficient
t	Pearson correlation coefficient test value
p	The probability of no correlation between the two variables

Table 1: Notations Table

3 Task1.a:Data collection and pre-processing

3.1 Data collection

By looking at the available data, we found that there was a lack of global temperature change data from 2013 to the present. Our team used python software to collect data from the NOAA internet for the last few years. The data collected was interpolated and removed as appropriate so that the global temperature change data could be analysed easily and still maintain continuity. The data can then be filtered for outliers and errors based on the RClimDex software to meet the analysis requirements.

3.2 Data pre-processing

3.2.1 Mann-Kenddall mutation test

After the data were initially collected and counted, we used the Mann-Kenddall mutation test to first input the raw data into the spatio-temporal data set as input to the program's independent variables, with the following output trend test indicators

$$S=\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}\left(x_{j}-x_{i}\right) \tag{1}$$

The output mutation trend test is as follows

$$\mathrm{UF}_{k}=\frac{S_{k}-E\left(S_{k}\right)}{\sqrt{\operatorname{Var}\left(S_{k}\right)}} \quad(\mathrm{k}=1,2, \cdots, \mathrm{n}) \tag{2}$$

where E(Sk) Var(Sk) is the mean and variance of the cumulative count Sk at X1,X2... .Xn are independent of each other and have the same continuous distribution, they can be calculated by the following equation

$$\mathrm{E}\left(S_{k}\right)=n(n+1) / 4 \tag{3}$$

$$\operatorname{Var}\left(S_{k}\right)=n(n-1)(2 n+3) \tag{4}$$

We used Matlab software to plot images of the time series data for UBk and UFk for different years (1900 to 2022), as shown in

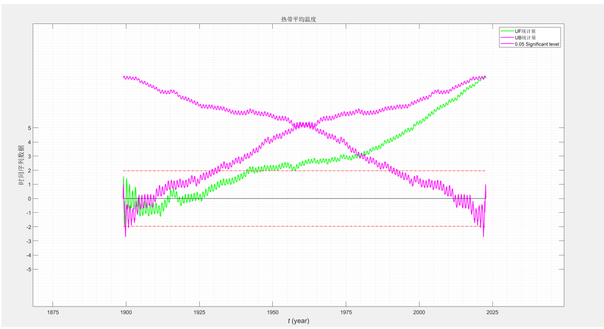


Figure 1: The MK test for mean tropical temperatures

From the image, it is clear that the two curves UBk and UFk intersect near March 2022 and the intersection point is near the critical line. The Mann-Kenddall mutation test shows that a mutation occurs near this point in time. We therefore conclude that the increase in temperature in March 2022 leads to a higher temperature increase than that observed during any previous 10-year period

4 Task1.b:The development of the temperature prediction model

4.1 GM(1,1) grey system theory model development

4.1.1 Preparation of the model

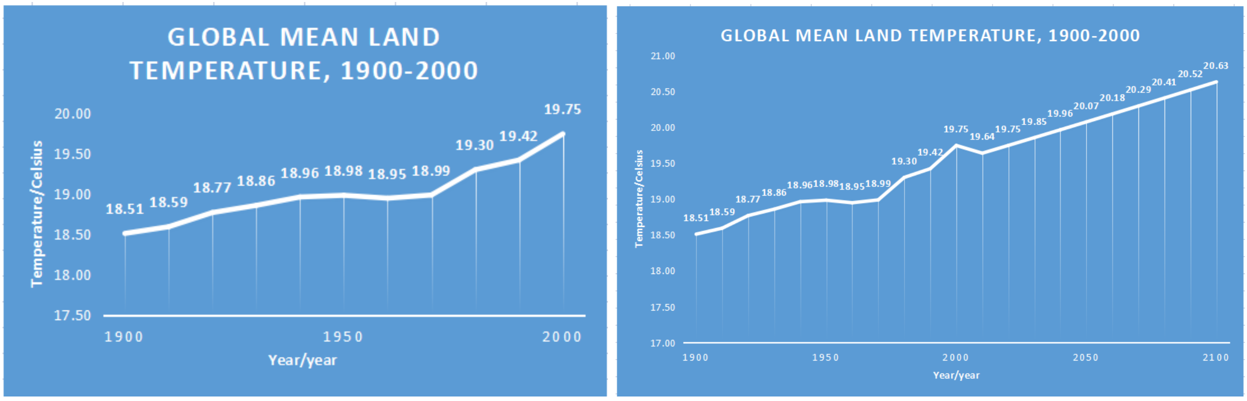
The GM(1,1) model of global temperature is a method of forecasting systems with uncertainty. Grey forecasting is the prediction of systems that contain both known and uncertain information. This is done by identifying the degree of dissimilarity between the trends of the system factors, i.e. by correlation analysis, and by generating the raw data to find the patterns of the system changes, generating a regular data series and then building the corresponding differential equation model to predict the future temperature changes.

4.1.2 Establishment of the model

A GM(1,1) grey prediction model was developed using the global land mean temperature. Using the least squares method, the prediction model is obtained as follows. The GM(1,1) grey prediction model was developed using the global land mean temperature and, after quadratic residual series analysis, as shown in equation (1).

$$X(1)_{t+1} = (18.51 - 18.4373/(-0.0039)) * (1 - \mathrm{e}\hat{}(-0.0055)) * \mathrm{e}\hat{}(0.0055*(11+t))\mathrm{t} = 1, 2, 3 \cdots\cdots 11 \tag{5}$$

After the model test and residual test, in the post-residual test: C=0.1047<0.35 when p=1 The relative error of equation (1) is controlled to within 5



(a) Global mean land temperature 1900-2000 (b) Global mean land temperature 1900-2100

Figure 2: Global land mean temperature prediction

Model projections of global land mean temperature for 2050 and 2100 from the model.

4.2 LSTM neural network model

4.2.1 Prediction of GM(1,1) grey system theory model

The problem is based on the global temperature data from 1899 to 2012 to build a model to predict the global temperature from 2050 to 2100. In this problem, since only time is used as a variable and not many other factors such as population and mega-hazards are given, it is more in line with the regulation of the use of time series. solving.

4.2.2 Data pre-processing

We collected global meteorological data from the Global Weather Station Data Network: NOAA, and used python and matlab software to filter the global average land temperature and CO2 emissions data from 1894 to 2022, and plotted the time series using

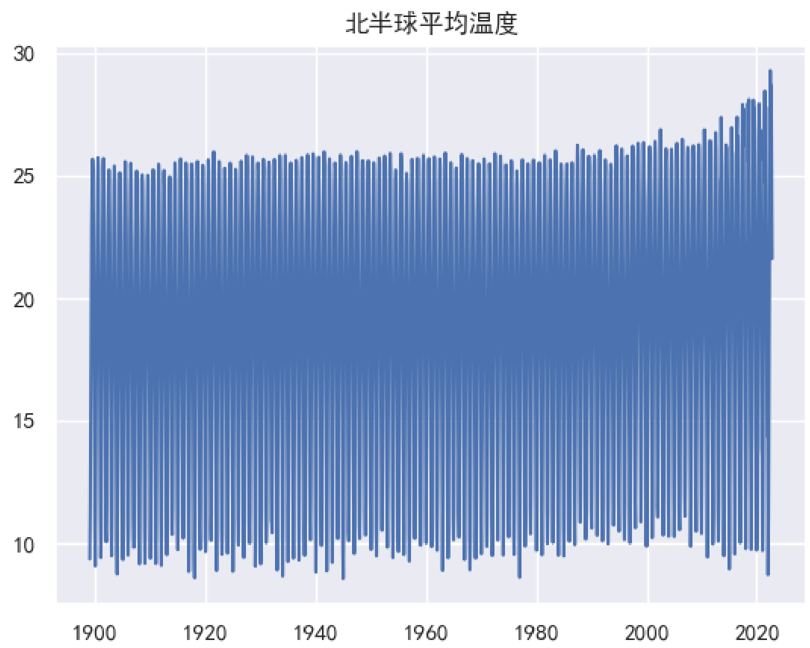


Figure 3: Average temperature change in the Northern Hemisphere

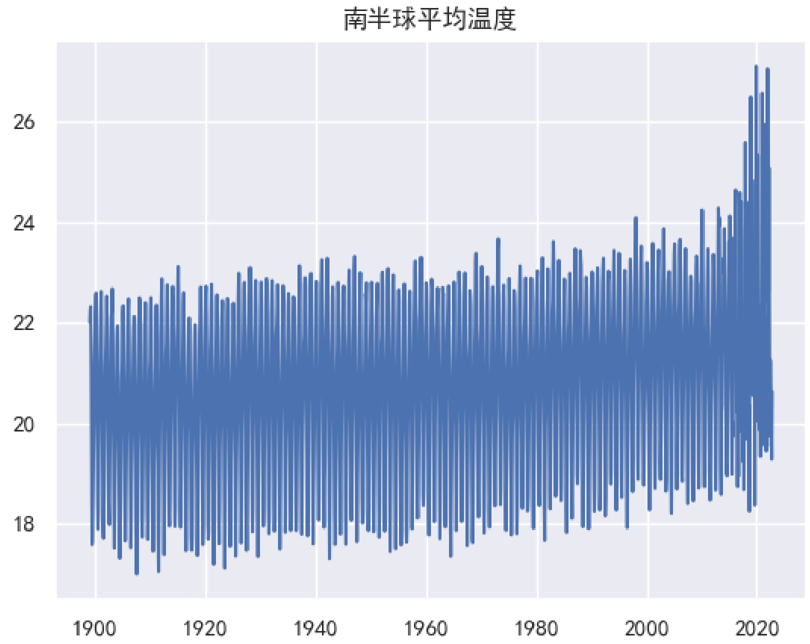
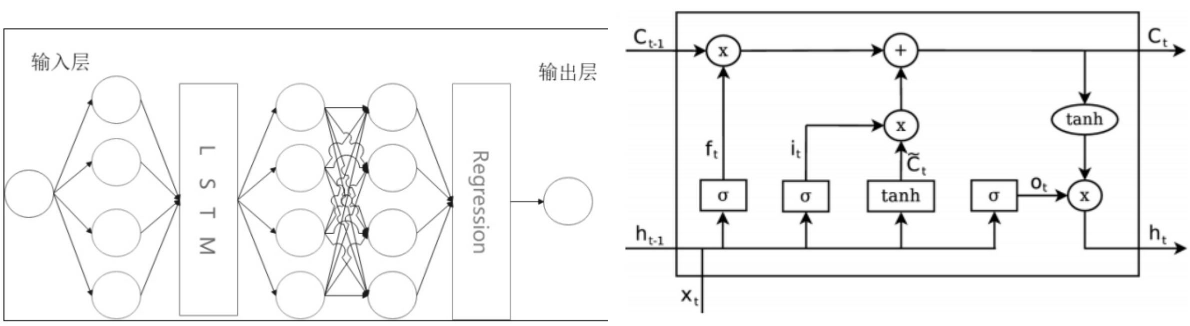


Figure 4: Average temperature change in the Southern Hemisphere

As can be seen from the graph, there is a clear upward trend in CO2 emissions and land temperature over time, which is consistent with the question we are studying.

4.2.3 Model building

The Long Short Term Memory Network, or LSTM for short, is a temporal recurrent neural network that solves the long term dependency problem of a general RNN (recurrent neural network). The structure of the model is shown in Fig.The internal structure of a single LSTM is shown in Figure .



(a) Structure of the LSTM neural network model (b) The internal structure of a single LSTM

Figure 5: Long-term and short-term memory networks

The core part is the memory block, which consists of input and forgetting gates and output gates, where a memory block is computed at time t as follow

$$\tilde{c}_t = \tanh\left(w_{xc}x_t + w_{ch}h_{t-1} + b_c\right) \tag{6}$$

$$\tilde{c}_t = \tanh\left(w_{xc}x_t + w_{ch}h_{t-1} + b_c\right) \tag{7}$$

The internal structure of a single LSTM is given by

$$f_t = \sigma\left(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f\right) \tag{8}$$

$$o_t = \sigma\left(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_t + b_o\right) \tag{9}$$

$$c_t = c_{t-1} \otimes f_t + i_t \otimes c_t \tag{10}$$

$$h_t = o_t \otimes \tanh c_t \tag{11}$$

where: denotes the update state of the memory cell at the moment. it, fi, ot it, fi, ot and ht It, fi, ot, ct, and ht represent the input, forget, output, memory cell, and hidden layer outputs at time t respectively. Ht- 1, ct- 1 denote the output of the hidden layer and the memory cell at time t- 1, respectively. wxc, wch denote the weight matrices of the memory cells with input xi and the hidden layer, respectively. wxt, whi, and wci are the weight matrices of the input gate and xt, hidden layer, and memory cell respectively. wxf, whi, and wci are the weight matrices of the input gate and xt, hidden layer, and memory cell respectively. wxf, whf, wcf are the weight matrices of the forgetting gate and xi, the output layer, and the memory cell, respectively. Wx0, who, wco are the weight matrices of memory cells with xi, output layer, and memory cells, respectively. Wx0, who, wco are the weight matrices of the memory cell and xi, output layer, and memory cell respectively. is the dot product, is the activation function, and bc, bi, bf, and bo are the biased chairs[1].

4.2.4 Solution of the model and presentation of results

Neural network construction using python software,The solution steps of the LSTM model are shown in Figure .

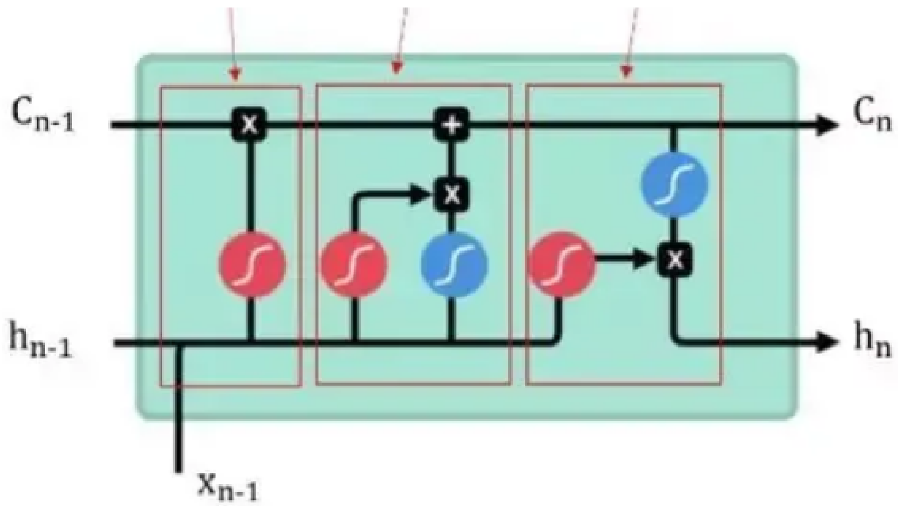


Figure 6: The solution steps of the LSTM model

5 Task1.c:Model prediction results

5.1 Prediction of the GM(1,1) grey system theory model

We believe that the global temperature will reach 20 degrees Celsius in 2050.Based on the GM grey prediction model we built in the previous question, we drew an image depicting the previous global temperature has been the possible future trend of global temperature change[1], based on this prediction, we can analyse that the global temperature will exceed 20 degrees in 2050 under this prediction model, the image is as follows.

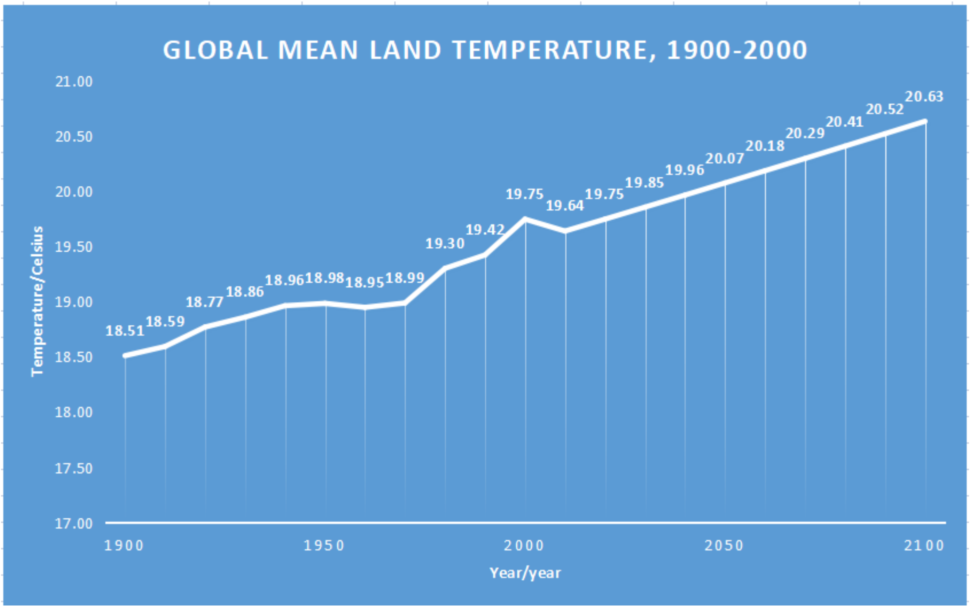


Figure 7: Global land mean temperature prediction

Based on the resulting image, the results we obtain from the GM grey projection model are that the global temperature will reach 20.07 degrees Celsius in 2050 and 20.63 degrees Celsius in 2100.

5.2 prediction of LSTM neural network model

Based on the LSTM neural network model we built in the previous question, we have drawn a picture depicting the previous global temperature and the likely future global temperature trend. Based on this prediction, we can analyse that the global temperature will exceed 20 degrees Celsius in 2050 under this prediction model with the following image.

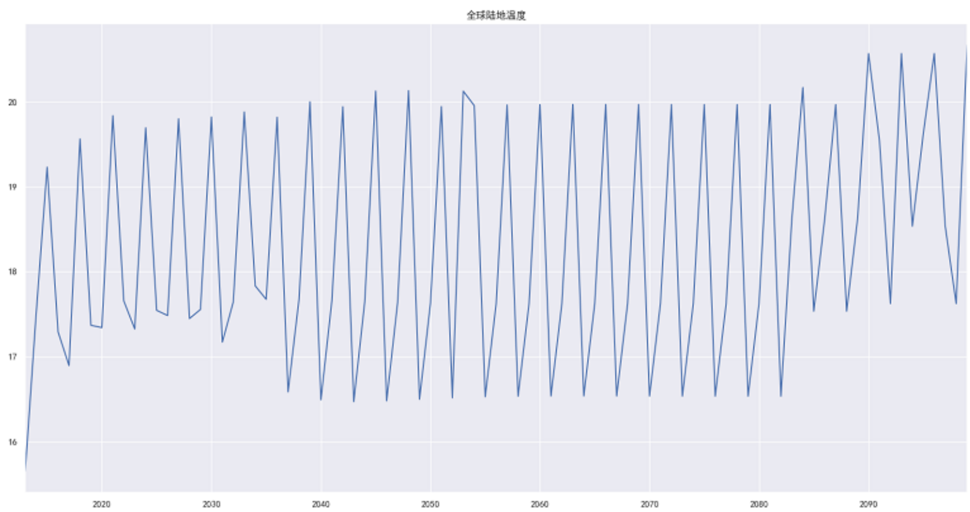


Figure 8: Global land mean temperature prediction

6 Task1.d :Comparison of the models

The GM grey prediction model[1] is considered first. The advantage of this model is that it can solve the problems of low historical data, the integrity of the series and low reliability by dealing with less eigenvalue data and not requiring a large enough sample space for the data, and can generate the irregular raw data to get a more regular generated series. The disadvantage is that it is only suitable for short to medium term forecasting and is more suitable for near exponential growth forecasting. Since the temperature we predict is not a huge change from the current temperature, it can be used with a high degree of accuracy.

The LSTM neural network model, on the other hand, has the advantage of having a long term memory for sequential modelling problems. It is simple to implement. At the same time, this model solves the problem of gradient disappearance and gradient explosion during the training of long sequences. The downside is that it has some disadvantages in parallel processing, and it is not as effective as some of the latest networks.

Taking the two together, we believe that although the grey prediction model is stable and can reflect the global temperature trend, it is not able to make long-term predictions and is only suitable for making more accurate predictions for the near future, and is not highly generalisable. In contrast, the LSTM neural network model can solve the problem of missing data better, and can make more accurate long-term forecasts, which is more adaptable and accurate, so we believe that the LSTM neural network model is more accurate.

7 Task2.a: Features and Leaders for change

7.1 Choice of model for Pearson correlation coefficient analysis

The Pearson correlation coefficient [1]is a measure of vector similarity. The output ranges from -1 to +1, where 0 represents no correlation, negative values represent negative correlation and positive values represent positive correlation. The Pearson correlation coefficient is optimised on the Euclidean distance by centring the values of the vectors, i.e. subtracting the mean of the elements for all dimensions in both vectors, the mean of all dimensions after centring is essentially zero; the result of centring is then found as the cosine distance, but the calculation of the cosine distance requires that all values in each vector must be non-empty. This method is the most appropriate and efficient way of correlation analysis when dealing with continuous data, normal distributions and linear relationships. It better fits the conditions of the correlation analysis we need for this question, so we use this model for the analysis of this subtopic.

7.2 Development of the Pearson correlation coefficient analysis model

The Pearson correlation coefficient ranges between [-1,1]. The following gives an understanding of the application of the Pearson correlation coefficient: Suppose there are two variables X, Y. Then there are: (1) When the correlation coefficient is 0, the X variable and the Y variable are not correlated. (2) When the value of X and the value of Y are increasing or decreasing, then the two variables are positively correlated, with a correlation coefficient between 0 and 1. (3) When the value of X increases and the value of Y decreases, or when the value of X decreases and the value of Y increases, the two variables are negatively correlated, with a correlation coefficient between -1 and 0. Note: The larger the absolute value of s of the correlation coefficient, the stronger the correlation, the closer the correlation coefficient is to 1 or -1, the stronger the correlation, and the closer the correlation coefficient is to 0, the weaker the correlation. The strength of correlation of variables is usually judged by the following range of values.[1]

0.8-1.0 Very strong correlation

0.6-0.8 Strongly correlated

0.4-0.6 Moderately correlated

0.2-0.4 Weakly correlated

0.0-0.2 Very weak or no correlation

We ran a Pearson correlation analysis of global temperature with each variable separately, looking at which factor had a correlation coefficient closer to 1 in absolute value to the global temperature, and which data had the greatest impact on global temperature.

$$\rho_{X,Y}=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{E\left(\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right)}{\sigma_X\sigma_Y}=\frac{E(XY)-E(X)E(Y)}{\sqrt{E\left(X^2\right)-E^2(X)}\sqrt{E\left(Y^2\right)-E^2(Y)}} \tag{12}$$

$$\rho_{X,Y}=\frac{N\sum XY-\sum X\sum Y}{\sqrt{N\sum X^2-\left(\sum X\right)^2}\sqrt{N\sum Y^2-\left(\sum Y\right)^2}} \tag{13}$$

$$\rho_{X,Y}=\frac{\sum XY-\frac{\sum X\sum Y}{N}}{\sqrt{\left(\sum X^2-\frac{\left(\sum X\right)^2}{N}\right)\left(\sum Y^2-\frac{\left(\sum Y\right)^2}{N}\right)}} \tag{14}$$

Using Matlab software, we analysed the Pearson correlation coefficients for year and temperature, CO2 concentration, northern hemisphere, southern hemisphere, northern temperate zone, southern temperate zone and tropics respectively and made tests with the following results.

Relationship	r	t	p-value
Year and CO2 concentration	0.9912	58.4827	0
Year and temperature	0.8441	16.6605	0
Year and Northern Hemisphere temperature	0.8286	15.5083	0
Year and Southern Hemisphere temperature	0.8594	17.7879	0
Year and Northern Hemisphere Temperate Temperatures	0.7897	13.6226	0
Year and Southern Hemisphere Temperate Temperatures	0.8279	15.7301	0
Years and tropical temperatures	0.8358	16.1106	0

Table 2: Notations Table

From the table we can see that the Pearson correlation coefficients for each relationship are greater than 0.7, so we can conclude that CO2 concentration, location relationships, and time of year all have some effect on global temperature. Elevation also has a certain influence on temperature. In order not to ignore this factor, we made statistics on the elevation data of Guo-jia area and carried out Pearson correlation analysis. Meanwhile, the heat map was drawn as follows:

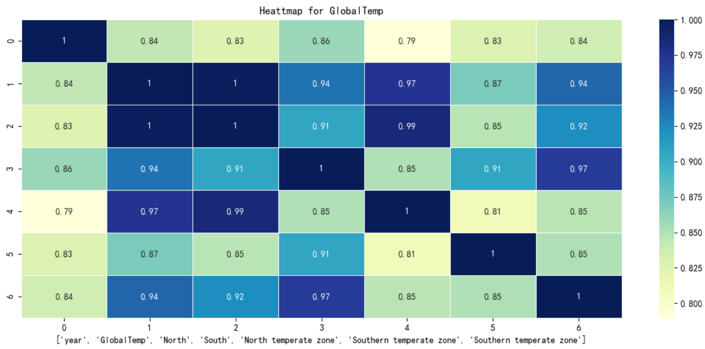
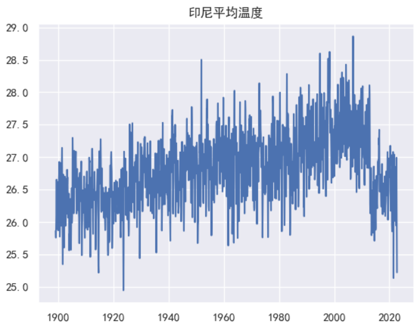
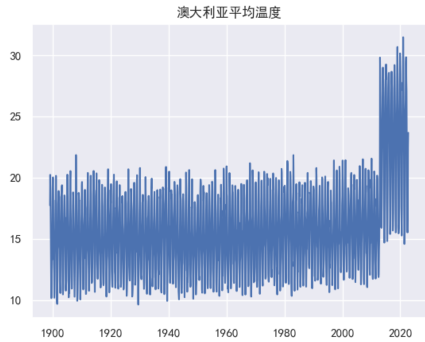


Figure 9: Heattmap for Global Temp

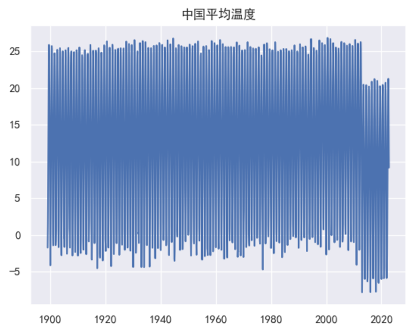
It turns out that elevation also affects global temperature.



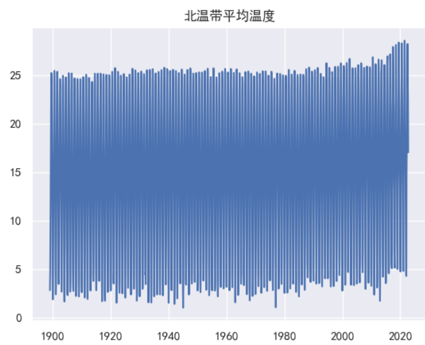
(a) Mean temperature in Indonesia



(b) Mean temperature of Australia



(c) Average temperature in China



(d) Average temperature in the North Temperate Zone

We analyzed the average temperature of different regions respectively, and drew the image with Matlab software, from which we could see the difference of global temperature in different locations more intuitively, which further proved that global temperature is related to location.

8 Task2.b:Application of the Pearson correlation coefficient analysis model

8.1 Problem analysis

When we analyse the effect of forest fires on global temperatures, we classify this natural disaster as a statistical problem for CO2. Because forest fires essentially affect CO2 emissions, and because co2 raises temperatures, we only need to know the effect on temperature per tonne of CO2 (which can be analysed using the correlation between co2 concentration and temperature)[1] and calculate the cumulative CO2 produced by fires in a given year or years to calculate the effect of forest fires on temperature.

8.2 The effect of forest fires on the concentration of carbon dioxid

According to the Pearson correlation coefficient model, we conducted a correlation analysis regarding the concentration of carbon dioxide generated by forest fires and global temperature, and calculated the correlation coefficient r=0.7980, the test value t=3.7452, and the probability value p=0.0057 that there is no correlation between the two variables. Therefore, it is concluded that Forest fires have an impact on global temperatures. Also we counted the temperature change in Brazil from 2012 to 2022 as shown in the figure.

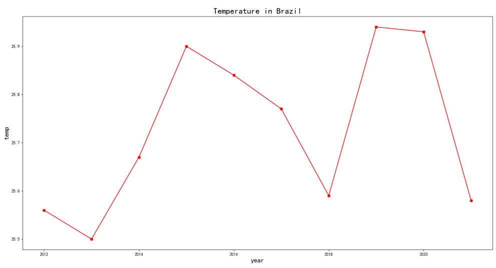


Figure 10: Average annual temperature vs CO2PPM in Brazil

This is complemented by the finding that there is an elevation in temperature in Brazil during the years of forest fires.

9 Task2.c: Other factors

We believe that the most critical influence affecting global temperature change is the concentration of carbon dioxide. The analysis based on the Pearson correlation coefficient model yielded a larger correlation coefficient between CO2 concentration and temperature than all other effects and a stable test value with a small p-value. It fits well with the test of the model. Conclusions can be drawn from the graphs as in Task 2.a.

Relationship	r	t	p-value
Year and CO2 concentration	0.9912	58.4827	0
Year and temperature	0.8441	16.6605	0
Year and Northern Hemisphere temperature	0.8286	15.5083	0
Year and Southern Hemisphere temperature	0.8594	17.7879	0
Year and Northern Hemisphere Temperate Temperatures	0.7897	13.6226	0
Year and Southern Hemisphere Temperate Temperatures	0.8279	15.7301	0
Years and tropical temperatures	0.8358	16.1106	0

Table 3: Notations Table

10 Task2.d:Measures to curb global warming

In order to stop global warming, we should do something to reduce CO2 emissions as shown in the following graph

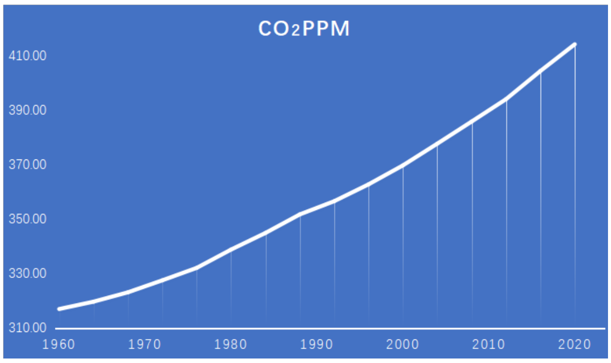


Figure 11: CO2PPM

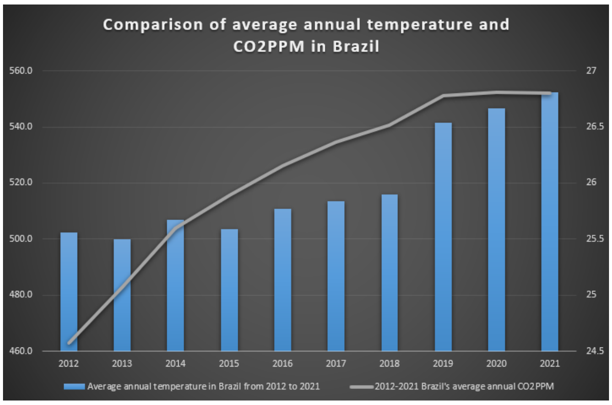


Figure 12: Average annual temperature vs CO2PPM in Brazil

CO2 is still a significant change to the increase in temperature. In this regard, our group's suggestion is to try to replace CO2 with new energy sources, such as reducing the use of coal and petrol, replacing fuel cars with new energy cars as much as possible, saving energy and planting trees, including choosing some green ways of travelling when travelling, such as walking or cycling, etc.

11 Strengths and Weaknesses

- Strengths:

1.For continuous data, normal distribution and linear relationship, pearson correlation coefficient is the most appropriate and efficient

2.The few missing data have little impact on the LSTM model

- Weaknessess:

1.Because of the large amount of data, our model cannot analyze each genre thoroughly.

2.The indicators of the popularization model cannot be quantified, resulting in a low persuasibility of the model.

3.The gap between experimental data should not be too widePearson correlation coefficient is greatly affected by outliers

A document to the APMCM Society

To whom it may concern,

After four days of discussion and communication, our group has made some new discoveries about global temperature warming. First of all, we preprocessed the global land mean temperature data. The given temperature data were all subjected to strict quality control and uniformity test, and the linear regression method of adjacent stations was used to interpolate the individual missing data to ensure the continuity of the meteorological data after processing and correction. According to the given data, we conducted a data visualization. Finally, it shows that the global temperature increase in March 2022 is larger than that observed in any previous decade.

The lstm neural network forecast model and GMP grey forecast model are then used to predict global temperatures in 2050 and 2100. The lstm neural network forecast model predicts that the temperature will reach 20 degrees Celsius in 2040, while the GMP grey forecast model predicts that the temperature will reach 20 degrees Celsius in 2045. We also found that the global warming trend slowed down in the second half of the forecast model, and we concluded that the global temperature rise is likely to reach a peak in a given year. Secondly, we use Pearson correlation coefficient to analyze the relationship between global temperature and different regions and time, and find that they are closely related, and also find that the occurrence of fire has a certain impact on the rise of global temperature.

Global warming is an important environmental problem. It is caused by human beings emitting too much carbon dioxide into the atmosphere, which produces the greenhouse effect. It can cause floods, hurricanes and other bad weather, at the same time make many animals and plants die, make the environment change, bring a lot of trouble to people. Fundamentally speaking, we should try our best to solve this problem by all means and protect the environment. One of the most critical links is to strictly control carbon emissions, we also need to start from small.

Yours Sincerely,

Team #apmcm2200789

References

[1] Ma Li, Yu Ruilin, Li Jie. Analysis and prediction of local ambient temperature changes in the Three Gorges Reservoir area based on time series J. Chinese Journal of Agrometeorology, 2018.39(1):9-17.

[2] Zhang Yingchun, Xiao Dongrong, Zhao Yuandong, et al. Research on Meteorological Prediction based on Time Series Neural network J1. Journal of Wuhan University of Technology 2003.27(2):238-240.

[3] Guo Qingchun, Li Li, Zhang Ran, et al. Neural network prediction model of temperature change in China J1. Journal of Tropical Meteorology, 2009.25(4):484-487.

[4] Luo Xianglong, Li Danyang, Zhang Shengrui, et al. Short-term traffic flow prediction based on KNN-LSTM J. Journal of Beijing University of Technology, 2018.44(12):79-92.

[5] Song Fangcan. Extreme Cold swept Europe and Asia killing more than 200 people [N]. Guangzhou Daily, 2012(A3).[8] Sun Mian. Research on Improvement and Application of Two-class Clustering Algorithm [D]. Xi ′an: Shaanxi University of Science and Technology.2020.

[6] PENG Zhuliang. Multi-label text classification based on encoding and decoding neural network structure [D]. Guangzhou: Guangdong University of Technology, 2020.

13 Appendix

Listing 1: The matlab Source code of Algorithm

matlab_code

```
'dataTrans'
#%%
clc
clear
%
%123AverageTemperature
%4AverageTemperatureUncertainty
%56
%7-10
Dataset = readtable("2022_APMCM_C_Data.csv",'Format','%s%f%f%s%s%s%s');
%
Datacell = table2cell(Dataset);
%
ContryLabel = Datacell(:,5);
ContryLabel_catgorical = categorical(ContryLabel);
ContryLabel_double = double(ContryLabel_catgorical);
ContryNum = size(unique(ContryLabel_double),1);
%
CityLabel = Datacell(:,4);
CityLabel_catgorical = categorical(CityLabel);
CityLabel_double = double(CityLabel_catgorical);
CityNum = size(unique(CityLabel_double),1);
[N,M] = size(Dataset);
%CountryData = cell(ContryNum,2);
```

```
%
Date_int = zeros(N,10);%
for i = 1:N  %
%
timeTemp = regexp(Datacell{i,1},'[/-]','split');
%
Date_int(i,1) = str2double(timeTemp{1,1});
%
Date_int(i,2) = str2double(timeTemp{1,2});
%d
Date_int(i,3) = double(Datacell{i,2});
%
Date_int(i,4) = double(Datacell{i,3});
%
Date_int(i,5) = CityLabel_double(i);
%
Date_int(i,6) = ContryLabel_double(i);
%
La = regexp(Datacell{i,6},'[NESW]','split');
%
Date_int(i,7) = str2double(La{1});
%1
Date_int(i,8) = Datacell{i,6}(length(Datacell{i,6})) == 'N';
%
Lo = regexp(Datacell{i,7},'[NESW]','split');
%
Date_int(i,9) = str2double(Lo{1});
%1
Date_int(i,10) = Datacell{i,7}(length(Datacell{i,7})) == 'E';
end
%%
YearStart = min(Date_int(:,1));%
YearEnd = max(Date_int(:,1));%
% n5
GlobalAverage = zeros((YearEnd-YearStart+1)*12,5);

for i = 1:N
%
if ~isnan(Date_int(i,3))
%12
newindex = (Date_int(i,1)-YearStart)*12+Date_int(i,2);
%
GlobalAverage(newindex,1) = Date_int(i,1);
%
GlobalAverage(newindex,2) = Date_int(i,2);
GlobalAverage(newindex,3) = GlobalAverage(newindex,3) + 1;%
GlobalAverage(newindex,4) = GlobalAverage(newindex,4) + Date_int(i,3);%
GlobalAverage(newindex,5) = GlobalAverage(newindex,5) + Date_int(i,4);%
end
end
for i = 1:size(GlobalAverage,1)
if GlobalAverage(i,3)>0
GlobalAverage(i,4) = GlobalAverage(i,4)/GlobalAverage(i,3);
GlobalAverage(i,5) = GlobalAverage(i,5)/GlobalAverage(i,3);
end
end
%
YearFlag = zeros(size(GlobalAverage,1)/12,1);
for i = 1:size(YearFlag,1)
flag = 1;
```

```
for j = 1:12
idx = (i-1) * 12 + j;
if GlobalAverage(idx,3) < CityNum
flag = 0;
end
end
YearFlag(i) = flag;
end
%YearFlag1899-2012
GlobalAverage_full = GlobalAverage(1873:3240,:);
Totalmonth = size(GlobalAverage_full,1);
GlobalAverage_year = zeros(Totalmonth/12,size(GlobalAverage_full,2));
for i = 1:12:Totalmonth
GlobalAverage_year = datenum(GlobalAverage_full(i,1),GlobalAverage_full(i,2),1);
end
%python
writematrix(GlobalAverage_full,'APMCM.csv')
%%
%
AreaAverage_month = zeros(Totalmonth,11);
AreaAverage_month(:,1:2) = GlobalAverage_full(:,1:2);
TempCounts = zeros(1,9);
for i = 1:N
if Date_int(i,1) >= 1899 && Date_int(i,1) <= 2012
idx0 = 12 * (Date_int(i,1)-1899) + Date_int(i,2);
% /
flag = Date_int(i,8) == 0;
if flag
AreaAverage_month(idx0,3) = AreaAverage_month(idx0,3) + Date_int(i,3);
TempCounts(1) = TempCounts(1) + 1;
else
AreaAverage_month(idx0,4) = AreaAverage_month(idx0,4) + Date_int(i,3);
TempCounts(2) = TempCounts(2) + 1;
end
la = Date_int(i,7);
%
if la <= 23.5
AreaAverage_month(idx0,5) = AreaAverage_month(idx0,5) + Date_int(i,3);
TempCounts(3) = TempCounts(3) + 1;
else
if la <= 66.5
%
if flag
AreaAverage_month(idx0,6) = AreaAverage_month(idx0,6) + Date_int(i,3);
TempCounts(4) = TempCounts(4) + 1;
%
else
AreaAverage_month(idx0,7) = AreaAverage_month(idx0,7) + Date_int(i,3);
TempCounts(5) = TempCounts(5) + 1;
end
else
%
%                 %
%                 if flag
%                     AreaAverage_month(idx0,8) = AreaAverage_month(idx0,8) + Data_i
%                     TempCounts(6) = TempCounts(6) + 1;
%                 %
%                 else
%                     AreaAverage_month(idx0,9) = AreaAverage_month(idx0,9) + Data_i
%                     TempCounts(7) = TempCounts(7) + 1;
```

```
%                  end
end
end
%3
if Date_int(i,6) == 3
AreaAverage_month(idx0,8) = AreaAverage_month(idx0,8) + Date_int(i,3);
TempCounts(6) = TempCounts(6) + 1;
%46
elseif Date_int(i,6) == 46
AreaAverage_month(idx0,9) = AreaAverage_month(idx0,9) + Date_int(i,3);
TempCounts(7) = TempCounts(7) + 1;

%9
elseif Date_int(i,6) == 9
AreaAverage_month(idx0,10) = AreaAverage_month(idx0,10) + Date_int(i,3);
TempCounts(8) = TempCounts(8) + 1;

%18
elseif Date_int(i,6) == 18
AreaAverage_month(idx0,11) = AreaAverage_month(idx0,11) + Date_int(i,3);
TempCounts(9) = TempCounts(9) + 1;
end

end
end
for i=1:size(TempCounts,2)
AreaAverage_month(:,i+2) = AreaAverage_month(:,i+2).* Totalmonth./ TempCounts(i);
end

%
AddData = importdata('adddata.mat');
AreaAverage_full = [AreaAverage_month ;AddData];
rowName = {'','','','','','','',...
    '','','',''};
OutPutTable = array2table(AreaAverage_full,'VariableNames',rowName);
% writematrix(AreaAverage_month,'.csv');
writetable(OutPutTable,'matlab.csv','WriteVariableNames',true);
%%
X_p = zeros(size(GlobalAverage_full,1),1);
for i = 1:size(GlobalAverage_full,1)
X_p(i) = datenum(GlobalAverage_full(i,1),GlobalAverage_full(i,2),1);
end
% X_p = datenum(re
%X_p = GlobalAverage_full(:,1) + GlobalAverage_full(:,2)./100;
Y_p = GlobalAverage_full(:,4);
Z_p = GlobalAverage_full(:,5);
%hold on
figure(1)
plot(X_p,Y_p,'b-','linewidth',1)
axis([min(X_p),max(X_p),10,26]);
%legend('AverageTemperature','AverageTemperatureUncertainty');
datetick('x','yyyy')
xlabel('year','FontSize',15);
ylabel('temperature','Fontsize',15);
title('Global AverageTemperature')

figure(2)
plot(X_p,Z_p,'m-','linewidth',1.5);
axis([min(X_p),max(X_p),0,1.5]);
datetick('x','yyyy')
```

```
xlabel('year','FontSize',15);
ylabel('temperature','Fontsize',15);
title('Global AverageTemperatureUncertainty')
'MKcheck'
%%Manner-Kendall M-K
A=xlsread(".csv");
%
x = zeros(size(A,1),1);
for i = 1:size(A,1)
x(i) = datenum(A(i,1),A(i,2),1);
end
titleName = {'tropical','Southern temperate zone','North temperate zone'};
for ii=1:3
y=A(:,2 + ii);%
time_series=y;
n=size(A,1);
UF=zeros(size(time_series));
E = n*(n-1)/4;
Var = n*(n-1)*(2*n+5)/72;
r1 = zeros(1,n);
for i= 1:n
r1(i) = sum(time_series(i)>time_series(1:i));
end
s = zeros(size(time_series));
for k = 2:n
s(k) = sum(r1(1:k));

E = k*(k-1)/4;                    % s(k)
Var = k*(k-1)*(2*k+5)/72;         % s(k)
UF(k) = (s(k)-E)/sqrt(Var);
end

time_series2 = zeros(1,n);
for i=1:n
time_series2(i)=time_series(n-i+1);
end
% UB=n=0
UB = zeros(size(time_series2));
r2 = zeros(1,n);
for i= 1:n
r2(i) = sum(time_series2(i)>time_series2(1:i));
end
s2 = zeros(size(time_series2));
for k = 2:n
s2(k) = sum(r2(1:k));

E = k*(k-1)/4;                    % s2(k)
Var = k*(k-1)*(2*k+5)/72;         % s2(k)
UB(k) = -(s2(k)-E)/sqrt(Var);
end
UB2 = zeros(1,n);
for i=1:n
UB2(i)=UB(n-i+1);
end
% x = 1:n;
figure(ii)
plot(x,UF,'g-','linewidth',1.5);
hold on
plot(x,UB2,'m-','linewidth',1.5);
plot(x,UB,'m-','linewidth',1.5);
```

```
plot(x,1.96*ones(n,1),'-.r','linewidth',1);
plot(x,0*ones(n,1),'-','color',[0.2,0.2,0.2],'linewidth',1);
plot(x,-1.96*ones(n,1),'-.r','linewidth',1);
grid(gca,'minor')%
datetick('x','yyyy')
axis([min(x),max(x),-5,5]);
legend('UF','UB','0.05 Significant level');
set(gca,'Fontsize',12)
set(gca,'ytick',-5:1:5)
xlabel('{\itt} (year)','FontSize',15);
ylabel('Time series data','Fontsize',15);
title(titleName(ii))
end
'GMP'
%function SGrey
clc,clear;
X0 =[18.510 18.593 18.769 18.858 18.962 18.983 18.947 18.987 19.300 19.421 19.745 ];
%input('');   %
n = length(X0);  %n

%
X1 = zeros(1,n);
for i = 1:n
if i == 1
X1(1,i) = X0(1,i);
else
X1(1,i) = X0(1,i) + X1(1,i-1);
end
end
X1

%BY
B = zeros(n-1,2);
Y = zeros(n-1,1);
for i = 1:n-1
B(i,1) = -0.5*(X1(1,i) + X1(1,i+1));
B(i,2) = 1;
Y(i,1) = X0(1,i+1);
end
B,Y

%GM(1,1)au
A = zeros(2,1);
A = inv(B'*B)*B'*Y;
a = A(1,1);
u = A(2,1);
a,u

%
XX0(1,1) = X0(1,1);
for i = 2:n
XX0(1,i) = (X0(1,1) - u/a)*(1-exp(a))*exp(-a*(i-1));
end
XX0
%
e = 0;          %
for i =1:n
e = e + (X0(1,i) - XX0(1,i));
end
e = e/n;
```

```
e
aver = 0;     %
for i = 1:n
aver = aver + X0(1,i);
end
aver = aver / n;
aver
s12 = 0;     %
for i = 1:n
s12 = s12 + (X0(1,i)-aver)^2;
end
s12 = s12 / n;
s12
s22 = 0;       %
for i = 1:n
s22 = s22 + ((X0(1,i) - XX0(1,i)) - e)^2;
end
s22 = s22 / n;
s22
C = s22 / s12;    %
C
cout = 0;    %
for i = 1:n
if abs((X0(1,i) - XX0(1,i)) - e) < 0.6754*sqrt(s12)
cout = cout+1;
else
cout = cout;
end
end
P = cout / n;
P
if (C < 0.35 & P > 0.95)
disp('');
m = input(': m = ');   %
disp('m');
f = zeros(1,m);
for i = 1:m
f(1,i) = (X0(1,1) - u/a)*(1-exp(a))*exp(-a*(i+n-1));
end
f
else
disp('');
end
''
MIN = min(mat); %
MAX = max(mat); %
MEAN = mean(mat); %
MEDIAN = median(mat); %
SKEWNESS = skewness(mat); %
KURTOSIS = kurtosis(mat); %
STD = std(mat); %
RESULT = [MIN;MAX;MEAN;MEDIAN;SKEWNESS;KURTOSIS;STD]
R2=corrcoef(mat)
R1=corrcoef(mat1)
R3=corrcoef(huozai)
t3=0.7980*((10-2)/(1-0.7980^2))^(0.5)
disp('p')
disp((1-tcdf(t3,8))*2)
R4=corrcoef(q)
t4=(-0.0999)*((8-2)/(1-(-0.0999)^2))^(0.5)
```

```
disp('p')
disp((1-tcdf(t4,6))*2)
' PPM'
MIN = min(PPM1); %
MAX = max(PPM1); %
MEAN = mean(PPM1); %
MEDIAN = median(PPM1); %
SKEWNESS = skewness(PPM1); %
KURTOSIS = kurtosis(PPM1); %
STD = std(PPM1); %
RESULT = [MIN;MAX;MEAN;MEDIAN;SKEWNESS;KURTOSIS;STD]
a1=(Year1-1959)/(2021-1959);
a2=(Year-1899)/(2012-1899);
R=corrcoef(a1,PPM1)
R2=corrcoef(a2,wendu)
R3=corrcoef(a2,beibanqiu1)
R4=corrcoef(a2,nanbanqiu1)
R5=corrcoef(a2,beiwendai1)
R6=corrcoef(a2,nanwendai1)
R7=corrcoef(a2,redai1)
t1=0.9912*((63-2)/(1-0.9912^2))^(0.5)
t2=0.8441*((114-2)/(1-0.8441^2))^(0.5)
t3=0.8260*((114-2)/(1-0.8260^2))^(0.5)
t4=0.8594*((114-2)/(1-0.8594^2))^(0.5)
t5=0.7897*((114-2)/(1-0.7897^2))^(0.5)
t6=0.8297*((114-2)/(1-0.8297^2))^(0.5)
t7=0.8358*((114-2)/(1-0.8358^2))^(0.5)
disp('p')
disp((1-tcdf(t1,61))*2)
disp('p')
disp((1-tcdf(t2,112))*2)
disp('p')
disp((1-tcdf(t3,112))*2)
disp('p')
disp((1-tcdf(t4,112))*2)
disp('p')
disp((1-tcdf(t5,112))*2)
disp('p')
disp((1-tcdf(t6,112))*2)
disp('p')
disp((1-tcdf(t7,112))*2)

'shujutianchong'
%
[cleanedData,missingIndices] = fillmissing(w2010,"linear");

%
clf
plot(cleanedData,"Color",[0 114 189]/255,"LineWidth",1.5,"DisplayName","")
hold on

%
plot(find(missingIndices),cleanedData(missingIndices),".","MarkerSize",12,...
"Color",[217 83 25]/255,"DisplayName","")
title(": " + nnz(missingIndices))

hold off
legend
clear missingIndices
python_code
```

```
import os
import torch
import torch.nn as nn
import pandas as pd
import numpy as np
from matplotlib.backends.backend_agg import FigureCanvasAgg
import matplotlib.pyplot as plt
import seaborn as sns
'plot.py'
sns.set_theme(style="darkgrid")
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.sans-serif'] = 'SimHei'

df = pd.read_csv('matlab.csv')
df['time'] = df[''].astype(str) + '-' + df[''].astype(str)
df['time'] = pd.to_datetime(df['time'])
print(df)
for i in range(2, 11):
plt.plot(df['time'].values, df.iloc[:, i].values)
plt.title(df.columns[i])
# plt.savefig('{}.png'.format(i))
plt.show()
'plot1.py'
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.sans-serif'] = 'SimHei'
#
data =( [[1.0000, 0.8441, 0.8260, 0.8594, 0.7897, 0.8279, 0.8358],
[0.8441, 1.0000, 0.9970, 0.9367, 0.9745, 0.8705, 0.9440],
[0.8260, 0.9970, 1.0000, 0.9066, 0.9854, 0.8476, 0.9221],
[0.8594, 0.9367, 0.9066, 1.0000, 0.8508, 0.9066, 0.9690],
[0.7897, 0.9745, 0.9854, 0.8508, 1.0000, 0.8100, 0.8469],
[0.8279, 0.8705, 0.8476, 0.9066, 0.8100, 1.0000, 0.8517],
[0.8358, 0.9440, 0.9221, 0.9690, 0.8469, 0.8517, 1.0000]])
data = np.array(data)
#print(data)
#
'''
values = np.random.rand(5, 5)
print(values)
ax = sns.heatmap(values, cmap="YlGnBu", annot=True, linewidths=.5)  #
ax.set_title('Heatmap for test')  #
ax.set_xlabel('x label')  # x
ax.set_ylabel('y label')  # y
plt.show()
'''
plt.figure(figsize=(20, 10),dpi=100)#
'''
x=['year','GlobalTemp','North','South','North temperate zone','Southern temperate zo
ax = sns.heatmap(data, cmap="YlGnBu", annot=True, linewidths=.5 )  #
ax.set_title('Heattmap for GlobalTemp')
ax.set_xlabel(x)
'''
#
baxi_data=[25.56,25.50,25.67,25.90,25.84,25.77,25.59,25.94,25.93,25.58]
year=[2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021]
plt.plot(year,baxi_data,c='red')
plt.scatter(year,baxi_data,c='red')
plt.xlabel("year", fontdict={'size': 16})
plt.ylabel("temp", fontdict={'size': 16})
plt.title("Temperature in Brazil", fontdict={'size': 20})
```

```
plt.show()
'plot2.py'
sns.set_theme(style="darkgrid")
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.sans-serif'] = 'SimHei'
df=pd.read_excel('.xlsx')
print(df)
plt.figure(figsize=(20, 10),dpi=100)#
plt.title('Global carbon dioxide')
plt.ylabel('PPM')
plt.xlabel('Year')
plt.grid(True)
plt.autoscale(axis='x', tight=True)
plt.plot(df['Year'],df['PPM'])
plt.show()
'LSTM'
#
plt.rcParams['axes.unicode_minus'] = False
plt.rcParams['font.sans-serif'] = 'SimHei'
df=pd.read_csv('APMCM.csv',names=['','','','',''])
print(df.info)
#
'''
df['time'] = df[''].astype(str) + '-' + df[''].astype(str)
df['time'] = pd.to_datetime(df['time'])
plt.plot(df['time'].values, df[''].values)
plt.title('')
plt.ylabel('')
plt.xlabel('')
plt.grid(True)
plt.autoscale(axis='x',tight=True)'''
#plt.show()
#
all_data=df[''].groupby(df['']).mean().values#
print(all_data)
test_data_size = 10 #
train_data = all_data[:-test_data_size]  #
test_data = all_data[-test_data_size:]  #
#print(len(train_data))
#print(len(test_data))
from sklearn.preprocessing import MinMaxScaler#
scaler = MinMaxScaler(feature_range=(0, 1))#
train_data_normalized = scaler.fit_transform(train_data .reshape(-1, 1))#
train_data_normalized = torch.FloatTensor(train_data_normalized).view(-1)#
#print(train_data_normalized)
#
train_window = 6  #
def create_inout_sequences(input_data, tw):
inout_seq = []
L = len(input_data)
for i in range(L-tw):
train_seq = input_data[i:i+tw]
train_label = input_data[i+tw:i+tw+1]
inout_seq.append((train_seq ,train_label))
return inout_seq
train_inout_seq = create_inout_sequences(train_data_normalized, train_window)
#Lstm
class LSTM(nn.Module):
def __init__(self, input_size=1, hidden_layer_size=64, output_size=1):
super().__init__()
```

```
self.hidden_layer_size = hidden_layer_size
#
self.lstm = nn.LSTM(input_size, hidden_layer_size)

self.linear = nn.Linear(hidden_layer_size, output_size)

self.hidden_cell = (torch.zeros(1,1,self.hidden_layer_size),
torch.zeros(1,1,self.hidden_layer_size))

def forward(self, input_seq):
lstm_out, self.hidden_cell = self.lstm(input_seq.view(len(input_seq) ,1, -1), self.h
predictions = self.linear(lstm_out.view(len(input_seq), -1))
return predictions[-1]

model = LSTM()
loss_function = nn.MSELoss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.001)
print(model)
#
epochs = 150

# inout_seq
for i in range(epochs):
for seq, labels in train_inout_seq:
optimizer.zero_grad()
model.hidden_cell = (torch.zeros(1, 1, model.hidden_layer_size),
torch.zeros(1, 1, model.hidden_layer_size))

y_pred = model(seq)

single_loss = loss_function(y_pred, labels)  #
single_loss.backward() #
optimizer.step()

if i%25 == 0:
print(f'epoch: {i:3} loss: {single_loss.item():10.8f}')
print(f'epoch:{i:3} loss:{single_loss.item():10.10f}')
#
fut_pred = 87
test_inputs = train_data_normalized[-train_window:].tolist()  # 12
print(test_inputs)
model.eval()
# 2100
for i in range(fut_pred):
seq = torch.FloatTensor(test_inputs[-train_window:])
with torch.no_grad():
model.hidden = (torch.zeros(1, 1, model.hidden_layer_size),
torch.zeros(1, 1, model.hidden_layer_size))
test_inputs.append(model(seq).item())
actual_predictions = scaler.inverse_transform(np.array(test_inputs[train_window:] ).
a=[]
[a.extend(i) for i in actual_predictions]
print(a)
x1 = np.arange(1899, 2012, 1)
x2 = np.arange(2013, 2100, 1)
y = range(1,7,1)
# print(x)
plt.figure(figsize=(20, 10),dpi=100)#
plt.title('Global land temperature prediction based on LSTM')
plt.ylabel('Temp')
```

```
plt.grid(True)
plt.autoscale(axis='x', tight=True)
plt.plot(x2, actual_predictions)
plt.show()
```