Team Control Number

**12646**

Problem Chosen

**B**

**2022**

**HiMCM/MidMCM**

**Summary Sheet**

# $CO_2$ and global warming research based on data-driven and multivariate regression analysis


Many monitoring data and studies show that there is an obvious warming trend in global climate. This paper mainly analyzes $CO_2$ concentration and land-ocean temperature data to explore the relations between global warming and $CO_2$. We build multiple models to predict the future $CO_2$ concentration and land-ocean temperature, and further collect relevant data to build multivariate mathematical models, and put forward suggestions to slow down the trend of global warming.

Firstly, we study the growth of $CO_2$ concentration in 2004. We use three methods, calculating the **10-year average growth**, **10-year fitting line slope** and **10-year growth series T-test statistics**. The results show that we partially agree that the year 2004 demonstrates the maximum growth. For the prediction of $CO_2$ concentration, the data set is first divided into the training set (80%) and the test set (20%), and then the basic models of time: **linear, quadratic, exponential, S-type model** and time series analysis model **ARIMA(2,2,3)**, are established. The fitting R-square of the five models on the whole data set are 0.98246, 0.99941, 0.99945, 0.99287, 0.99983, respectively. We also predict the $CO_2$ concentration in 2100, which are 534.8825 ppm, 688.3268 ppm, 833.0377 ppm, 442.4825 ppm, 694.1067 ppm, respectively, with the five models. At the same time, it is predicted that the time points for $CO_2$ concentration to reach 685 ppm are 2212, 2099, 2082, *No solution*, 2098. Among them, the quadratic, exponential and ARIMA models can reach 685 ppm in 2050, while the S-type model will never reach 685 ppm. By comparing RMSE, MAE and MRE, the exponential model has the best performance in the training set and test set, and is the most accurate.

For land-ocean temperature, **the correlation coefficient** between $CO_2$ concentration and land-ocean temperature is **0.9613**, indicating an extremely strong linear correlation between them. Considering that the most accurate model for predicting $CO_2$ is problem 1, we establish an **exponential model** to predict land-ocean temperature, and the R-square is 0.9184 through fitting. It is further predicted that the time points of 1.25°C, 1.50°C and 2°C are 2028, 2036 and 2048, respectively. Based on the correlation between $CO_2$ and land-ocean temperature, we further establish a **linear regression model** $T=a*CO_2+b$. The fitting **R-square is 0.9241**, and the predicted time points of 1.25°C, 1.50°C and 2°C are 2030, 2038 and 2051, respectively. Finally, based on the theory of **one-dimensional radiative conduction cycle and radiative forcing**, we discuss several factors affecting land-ocean temperature, including **solar radiation, greenhouse gases, earth reflectance, aerosols, fuel combustion** and so on, among which greenhouse gases include $CO_2$, $CH_4$ and $N_2O$. So, we build a **multivariate linear regression model**. According to the collected data, the final prediction model is obtained, which predicts a temperature rise of **1.7810°C in 2050**. Further, we analyze the sensitivity of $CO_2$ concentration level on land-ocean temperature, indicating that the $CO_2$ concentration level is very sensitive to land-ocean temperature.

Our models above show that human activities are aggravating carbon emissions, thus increasing $CO_2$ concentrations and causing earth warming. The earth is our common home. We suggest that we should reduce carbon emissions as much as possible, and set up relevant global organizations to jointly protect the earth of human beings and benefit our children and grandchildren.



**Key words:** Global warming; $CO_2$ concentration; Land-ocean temperature; Multivariate linear regression analysis; Exponential model

# Contents

# 1. Introduction

## 1.1. Background

In the last century, global climate has been undergoing significant changes mainly characterized by global warming. In the second half of the last century, global greenhouse gas concentration continued to rise rapidly, and the global average surface temperature maintained an overall rising trend, which was accompanied by a series of climate problems, such as rising sea level, melting plateau snow cover, retreating polar sea ice, and rising sea surface temperature. By the early 21st century, it was almost no longer in dispute that greenhouse gases emitted by human activity would cause global warming.



Photo credit: pixabay.com https://image.baidu.com/

On February 2, 2007, the Intergovernmental Panel on Climate Change (IPCC) published a summary of its fourth global Climate Change Assessment report[1], which combines six years of scientific research by scientists all over the world. The report says that climate warming is already an "undisputed" fact. The report says the most likely "rise" in global average temperatures is 1.8℃ to 4℃ from now to 2100, with sea levels of 18 to 59 cm. A more than 90 percent likely rise in global average temperature over the past 50 years may be related to an increase in greenhouse gases generated by human use using fossil fuels, the first time the IPCC has used such severe wording to describe the association between human activity and warming. The IPCC Committee noted that the global average surface temperature rose by 0.74℃ over the past 100 years; the past 50 years is the warmest in the world, which was the highest in the past 500 and 1,300 years, the northern 20th century was probably the hottest in the past 1,000 years, and the 1990s was the warmest decade, 1998 and 2005 are the warmest year on record. According to the World Meteorological Organization (WMO) [2], the global average temperature in 2006 was 0.42℃ higher than the average in 1961-1990, the sixth warm year on record. The IPCC committee believes that the abnormal and rapid rise in global temperatures during this period coincides with a human period of intensive greenhouse gas emissions, and that human activity is the main cause of global warming.

In order to facilitate people's understanding of global climate changes, a simplified model of $CO_2$ concentration and earth temperature change needs to be built to make reasonable predictions of future climate change and $CO_2$, and provide explanation as to its relationship with global warming, so to show people the development trend and impact of global climate change, explain the causes of global warming, enhance people's awareness of environmental protection and the ability to warn of risks, and urge policy makers to launch corresponding policies and regulations on global climate issues. Therefore, how to establish climate models accurately and effectively is of great significance for environmental protection and the safety of people's lives and property.

## 1.2. Problem restatement

Human activities and various carbon emissions have an important impact on global warming. This study is based on the change rule of carbon dioxide and land-ocean temperature and the relationship between them to predict future carbon dioxide concentration and land-ocean temperature. We aim to accomplish the following tasks.

### Problem 1: CO2 Level Analysis and Prediction Research

**1.a** Do you agree that the March 2004 increase of $CO_2$ resulted in a larger increase than observed over any previous 10-year period? Why or why not?

**1.b** Fit various (more than one) mathematical models to the data to describe past, and predict future, concentration levels of $CO_2$ in the atmosphere.

**1.c** Use each of your models to predict the $CO_2$ concentrations in the atmosphere in the year 2100. Do any of your models agree with claims and predictions that the $CO_2$ concentration level will reach 685 ppm by 2050? If not by 2050, when do your models predict the concentration of $CO_2$ reaching 685 ppm?

**1.d** Which model do you consider most accurate? Why?

### Problem 2: The Relationship Between Temperature and $CO_2$ Research

**2.a** Build a model to predict future land-ocean temperatures changes. When does your model predict the average land-ocean temperature will change by 1.25°C, 1.50°C, and 2°C compared to the base period of 1951-1980?

**2.b** Build a model to analyze the relationship (if any) between $CO_2$ concentrations and land-ocean temperatures since 1959. Explain the relationship or justify that there is no relationship.

**2.c** Extend your model from part 2.b. into the future. How far into the future is your model reliable? What concerns, if any, do you have with your model's ability to predict future $CO_2$ concentration levels and/or land-ocean temperatures?

### 1.3. Our Work

Human activities and various carbon emissions have an important impact on global warming. This study is based on the change rule of carbon dioxide and land-ocean temperature and the relationship between them to predict future carbon dioxide concentration and land-ocean temperature.

For problem 1, first of all, we conducted a basic statistical analysis of the $CO_2$ data, including mean value, variance, etc., and verified that the increment in 2004 was larger than that in any previous decade by using methods such as 10-year incremental mean value, 10-year fitting line slope and T-test, Rank-sum test. Secondly, we establish the basic models with explicit time, including linear, quadratic, exponential and S-type models, etc., and establish the time series ARIMA model to predict the change of $CO_2$ concentration. Thirdly, all the models were used to predict the $CO_2$ concentration in 2100, and the time points for the $CO_2$ concentration to reach 658 ppm were obtained. Finally, the most accurate model is discussed through the analysis of prediction accuracy and error, and the rationality of the model.

For problem 2, firstly, the correlation between $CO_2$ and temperature was studied, on which basis an exponential model was established to predict the temperature change according to the best model of Problem 1. Furthermore, a linear regression model was built with $CO_2$ as the independent variable and temperature as the dependent variable to explore the relationship between them, and the time points when the temperature change was 1.25°C, 1.50°C and 2°C were obtained. Finally, based on relevant literature[3], we established a more realistic multivariate linear regression model by considering solar radiation, greenhouse gases, earth reflectance, aerosols, fuel combustion and other factors, and collected relevant data for model fitting and future temperature prediction.

Finally, we also did the sensitivity analysis of temperature with respect to $CO_2$ concentration and other factors; a one-page non-technical report was designed to jointly call on human beings to reduce carbon emissions and protect the earth that we depend on for survival.

## 2. Variables and Assumptions

### 2.1. List of variables

The table below defines all the variables in this paper.

**Table 4.1 Definitions of variables**

| Symbol | Definition | Unit |
|---|---|---|
| $y/CO_2$ | $CO_2$ concentration | ppm |
| $y_i^d$ | Average of 10-year $CO_2$ concentration increment to year $i$ | ppm |
| $T$ | Land-ocean temperature | ℃ |
| $t$ | Time | year |
| $RF$ | Radiation forcing | $w/m^2$ |
| $r$ | Person correlation | / |

## 2.2. Assumptions

1. Assume no sudden change in natural factors in the next 200 years;
2. Assume no sudden changes in human activity in the next 200 years;
3. Assume that the data given and collected are true and valid.

# 3. Model Preparation: Visualization and Statistical Analysis

Data set 1 gives the $CO_2$ from 1959-2021 of the average March concentration level, the molar mass fraction in dry air, measured in ppm, and the data are obtained from NOAA (https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_annmean_mlo.txt.). Data set 2 gives the global average surface temperature, which is the difference between the average temperature from 1951-1980, with data obtained from Aeronautics and Space Administration Goddard Institute for Space Studies. $CO_2$ basic statistical analysis was performed, including the mean, standard deviation, median, maximum, minimum, etc., such as Table 4.2.

**Table 4.2 Basis statistics of $CO_2$**

| Statistics | Num | Mean | Std dev | Max | Min | Median |
|---|---|---|---|---|---|---|
| Concentration of $CO_2$ | 63 | 357.34 | 29.849 | 416.45 | 315.98 | 354.45 |
| Annual increase of $CO_2$ | 62 | 1.6205 | 0.67346 | 3.4 | 0.42 | 1.685 |

The time series diagram of $CO_2$ and the annual increase of $CO_2$ are shown in Figure 4.1 and Figure 4.2.

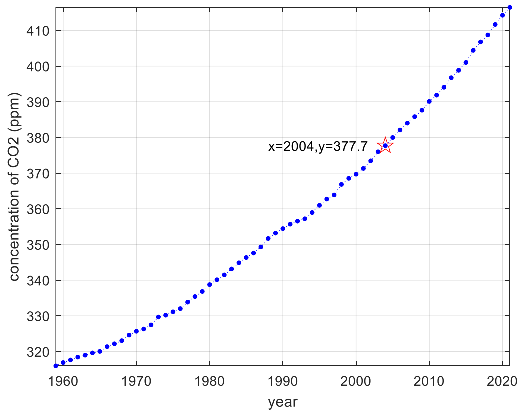



**Figure 4.1 Concentration of $CO_2$**

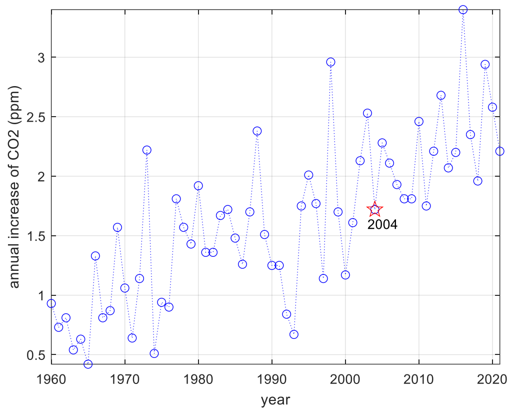



**Figure 4.2 Annual increase of $CO_2$**

According to Table 4.2 and Figure 4.1, $CO_2$ concentration shown an increasing trend, increasing from 315.98 ppm in 1959 to 416.45 ppm in 2021, with a relative increase of 31.80%. As can be seen from Figure 4.2, the overall trend of growth is still increasing.

**Table 4.3 Basis statistics of temperature**

| Statistics | Num | Mean | Std dev | Max | Min | Median |
|---|---|---|---|---|---|---|
| Temperature | 63 | 0.34656 | 0.32475 | 1.02 | -0.2 | 0.32 |

Based on the visual analysis of land-ocean temperature changes over time, as shown in Figure 4.3, we see sustained global temperature growth in fluctuations since 1980, which has increased by nearly 1℃ over the

40 years of 2021.

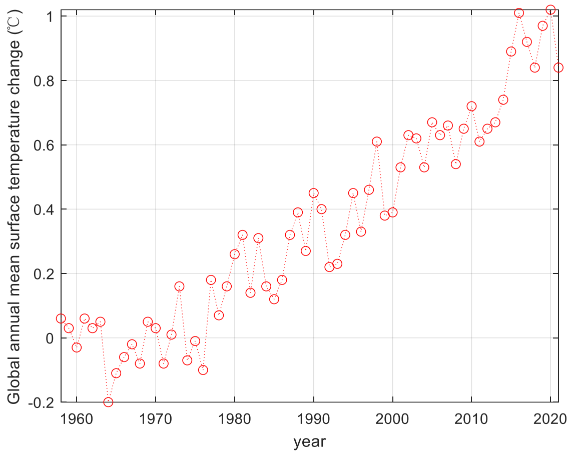



**Figure 4.3 Annual increase of land-ocean temperature**

# 4. Problem 1： $CO_2$ Level Analysis and Prediction Research

## 4.1. Problem analysis: Overall idea of problem 1

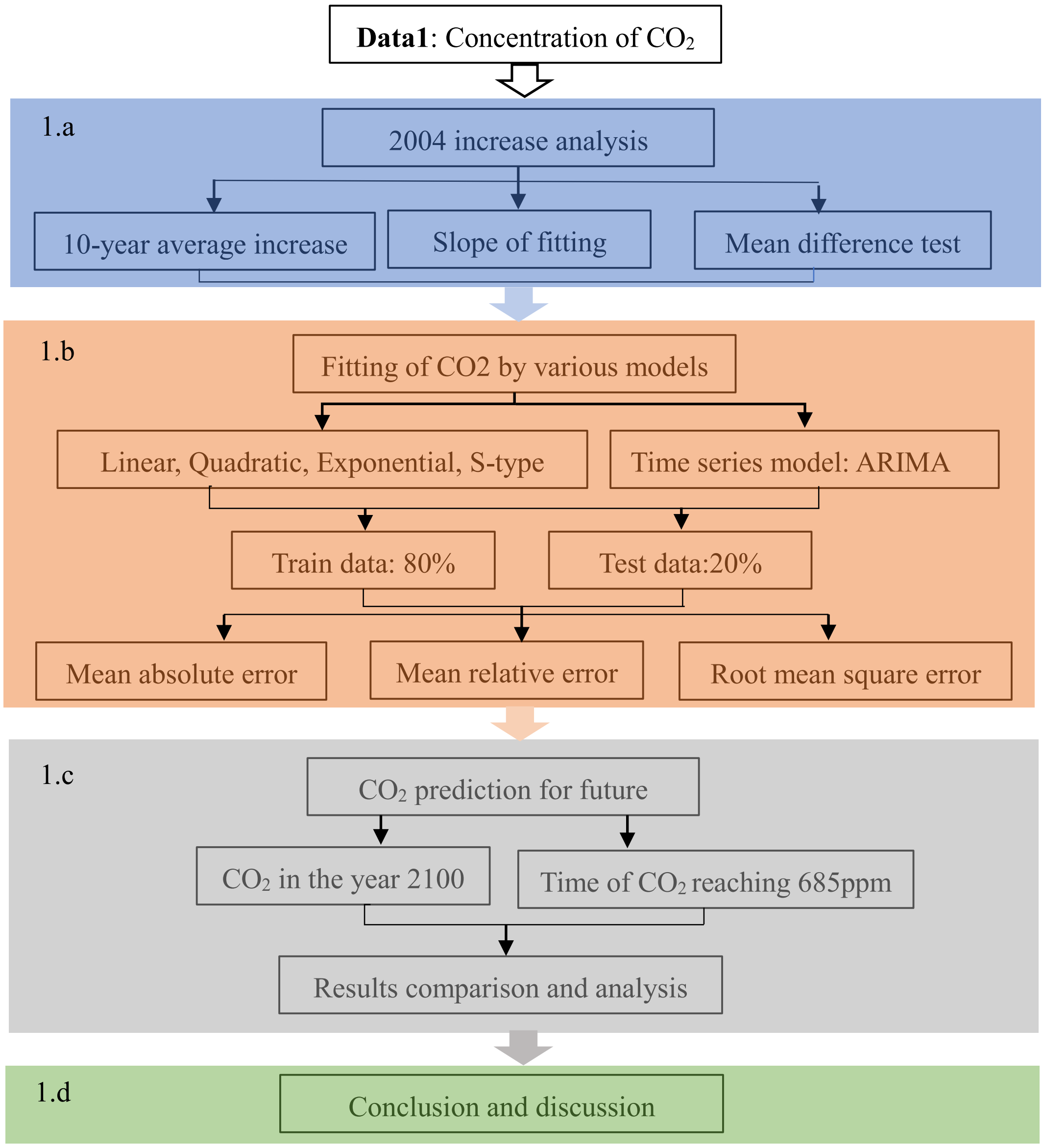



**Figure 4.4 The overall idea of problem 1**

For Problem 1, regarding the problem of whether the growth of $CO_2$ in 2004 is greater than ever, we take the mean value of increment of $CO_2$ every 10 years as the main reference quantity, by using means of direct comparison of average values, slope of fitting lines for 10 years based on least squares, and T-test, Rank-sum test. For the prediction of $CO_2$ concentration, based on Data set 1, we consider the basic model with explicit time and the $CO_2$ time series model; we compare the error accuracy of different models in the training set and the test set, and finally select the best model, which was used to predict the $CO_2$ concentration in 2100 and the time when $CO_2$ concentration reaches 685 ppm. The overall idea is shown in Figure 4.4.

## 4.2. [1.a] Analysis of 2004 increase of $CO_2$

### 4.2.1. Visualization of 10-year average growth

According to the background of the problem, we calculated the average $CO_2$ growth of each decade to describe the change of $CO_2$ growth, such as 2004, referring to 10-year average increase from 1995 to 2004. The computational formula is shown in equation (4-1).

$$y_i^d = \frac{y_i - y_{i-9}}{9}, \qquad i = 1968, 1969, \dots, 2021. \tag{4-1}$$

In equation (4-1), $y_i$ represents the $CO_2$ concentration of year $i$, $y_i^d$ represents the decadal average increment until year $i$. This formula is applied to Data set 1, and the results are shown in Figure 4.5, where the 10-year average increment value until 2004 is 1.86°C, while that of 2003 is 1.89°C. 2004 exceeded the average increase of any previous decade except for 2003.

As can be seen from Figure 4.5, the average growth rate decreased slightly in the 1990s, but maintained a momentum of continuous growth after 2004. By 2021, the 10-year average growth rate reached 2.49°C.

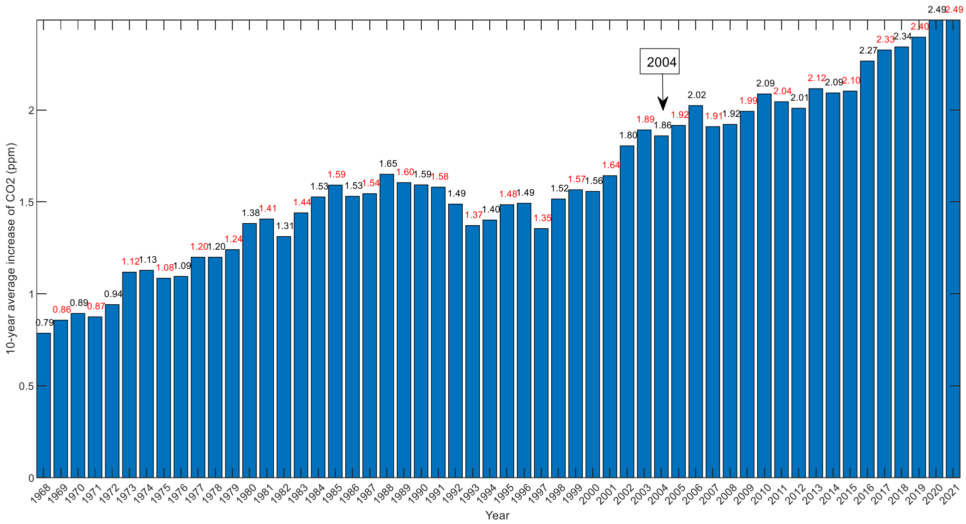



**Figure 4.5 10-year average increase of $CO_2$ (ppm)**

### 4.2.2. 10-year change slope based on least squares fitting

For the 10-year growth rate, the slope of the fitting line is used to describe the 10-year growth rate, so as to study the change of the 10-year growth rate. The basic model is:

$$y = kt + b, \tag{4-2}$$

where $y$ is the annual $CO_2$ concentration, $t$ is the actual year, and $k$ is the slope. The serial value of slope $k$ can be obtained by fitting the data of each successive 10 years, as shown in Figure 4.6. It can be seen from the figure that in 2004, the corresponding 10-year linear fitting slope was 1.853, which was larger than that of any previous 10 years. The lowest value was reached in 1997, after which the slope still showed an increasing trend.

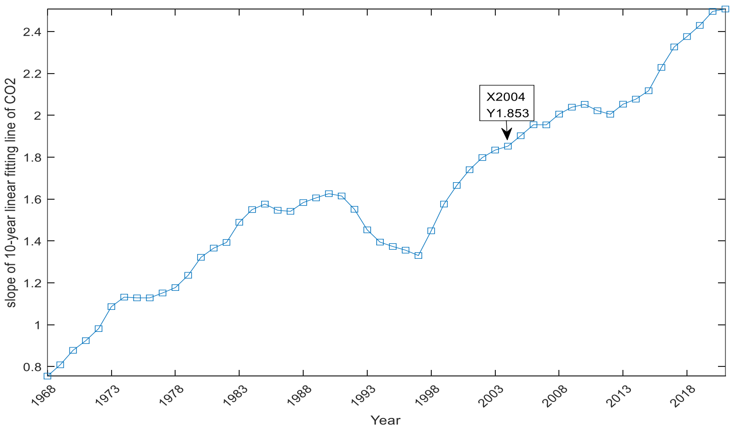



**Figure 4.6 Slopes of each 10-year linear fitting line**

### 4.2.3. The significance analysis of growth based on the T-test and Rank-sum test

Based on the 10-year growth sequence, we used the method based on statistical tests to verify whether there were significant differences in growth changes. Considering the unknown distribution of the sample sequence, T-test and Rank-sum test were used. The sequence of 2004 was used as the reference sequence, and the mean difference test was conducted between the sequence of 2004 and the sequence of any previous 10 years.

**Two-sample independent T-test method:** The T-test method is suitable for testing the normal distribution of the data, so we first conducted the normal fitting and test for the annual increment sequence, as shown in Figure4.7. We can see that the data conforms to normal distribution.

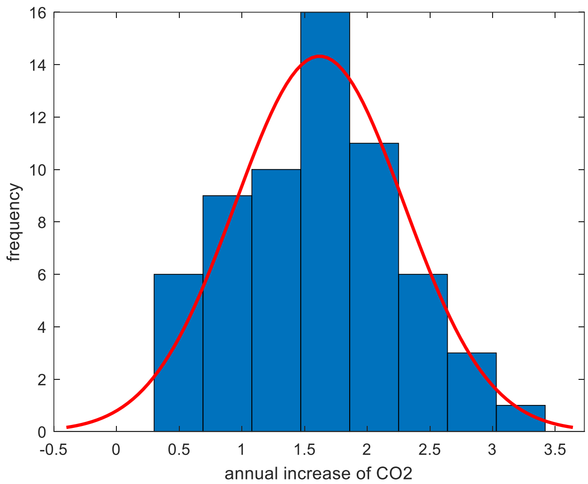



**Figure 4.7 Normal fitting annual increase of $CO_2$**

**Rank-sum test:** It is a non-parametric test, which is independent of the specific form of the overall distribution and can be applied regardless of the distribution of the studied object and whether the distribution is known.

The corresponding 10-year average increment in 2004 was taken as the fixed sequence, and 36 10-year average increment sequences from 2003 to 1968 were taken as the comparison sequence according to the sliding window mode. The results of two kinds of tests were obtained as follows.

**T-test:** H=[0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1];

**Rank-sum test:** H=[0 0 0 0 0 0 1 0 0 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 11].

H=0 indicates that the medians of 2004 10-years increase and comparing samples are equal; H=1 not equal. As can be seen from the above calculation results, the 10-year average increase of 2004 was significantly higher than that of the previous 10 years in most cases, but there were also a few parts that were not significant.

### 4.2.4. Conclusion

We compared and examined it in three methods, and in the end we partially agreed that the 10-year increase for 2004 was larger than any previous decade.

## 4.3. [1.b] Multi-model predictions of $CO_2$ concentration

Based on Data set 1, we decide to establish the basic model with explicit time, including **linear, quadratic, exponential, S-type**, etc. According to the characteristics of the data themselves, we also choose to further establish the time series prediction model **ARIMA**.

### 4.3.1. The division of the data set

In order to make an objective evaluation of the prediction ability of the model, Data set 1 of $CO_2$ concentration was first divided into 63 samples, among which 51 samples of the first 80% were used as the training set, and the last 20% of the sample, the size of which is 12, as the test set; the error is compared between the training set and the test set.

### 4.3.2. The basic model of time: linear, polynomial, exponential, S-type

**Linear model:** After fitting, $a$=0.0015, $b$=-2535.2, that is, $y$=0.0015$t$-2535.2, goodness of fitting $R^2$=0.9877 of the training set, fitting and prediction effects and error analysis are shown in Figure 4.8.

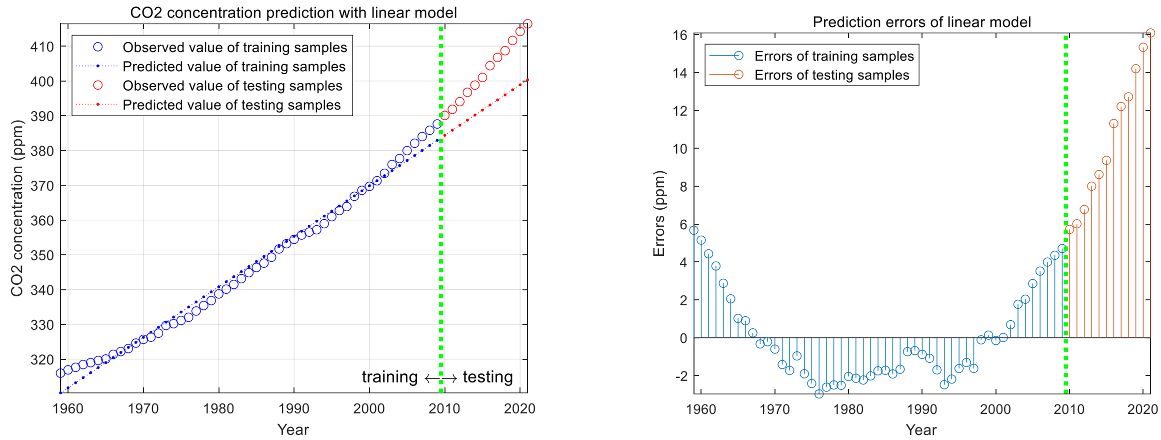



**Figure 4.8 $CO_2$ prediction and prediction error with linear model**

**Quadratic model, $y=at^2+bt+c$:** Through the fitting of training set, $a$=0.0118, $b$=-45.3892, $c$=43929; That is, $y = 0.0118t^2 - 45.3892t + 43929$, and goodness of fitting $R^2$=0.999. The fitting and prediction effects and error analysis are shown in Figure 4.9:

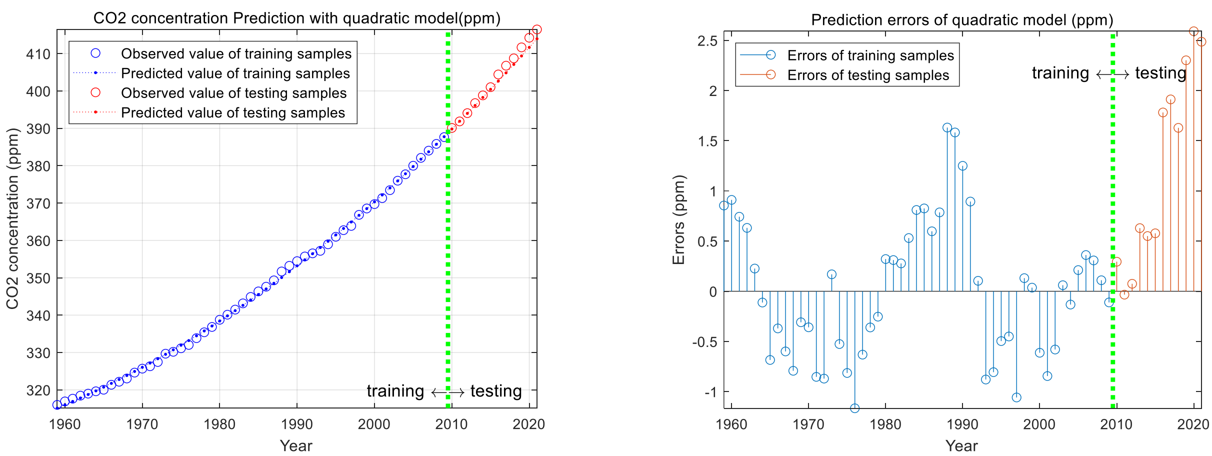



**Figure 4.9 $CO_2$ prediction and prediction error with quadratic model**

**Exponential model:** Considering that the order of magnitude of the independent variable $t$ is large, for the accuracy of the model, time $t$ is first normalized, and the formula is $T = (t - \min)/(\max - \min)$; through data training, $a$=59.8834, $b$=0.9917, $c$=254.7966, that is, $y = 59.8834e^{0.9917t} + 254.7966$, the fitting and prediction results and error analysis are shown in Figure 4.10.

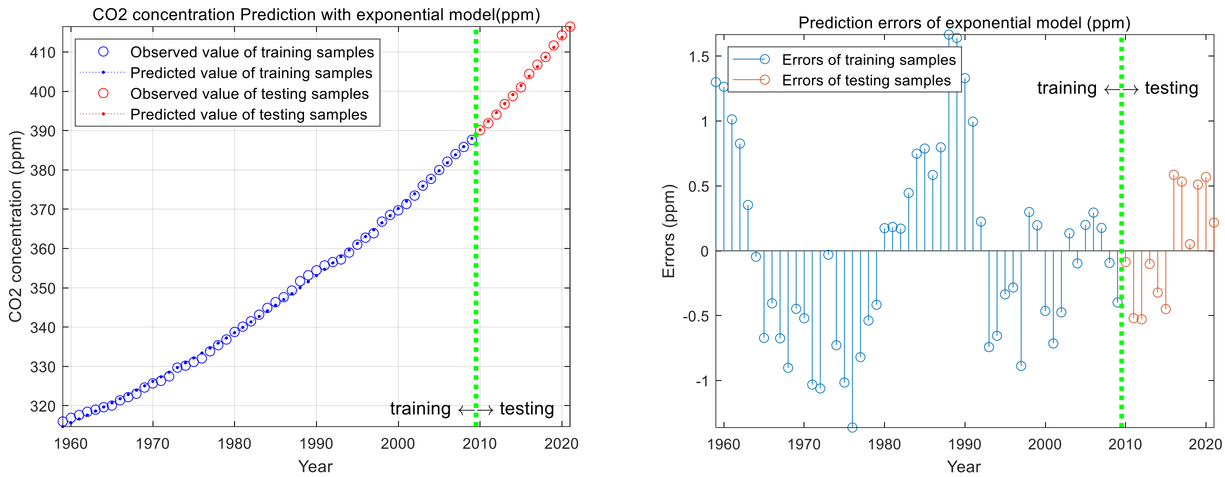



**Figure 4.10 $CO_2$ prediction and prediction error with quadratic model**

**S-type model:** It is also necessary to normalize the time *t* first and obtain the final parameters as *a*=0.8406, *b*=20.0158, *c*=5.5396, that is, $y = 0.8406/(1 + 20.0158e^{-5.5396t})$. The fitting and prediction results and error analysis are shown in Figure 4.11.

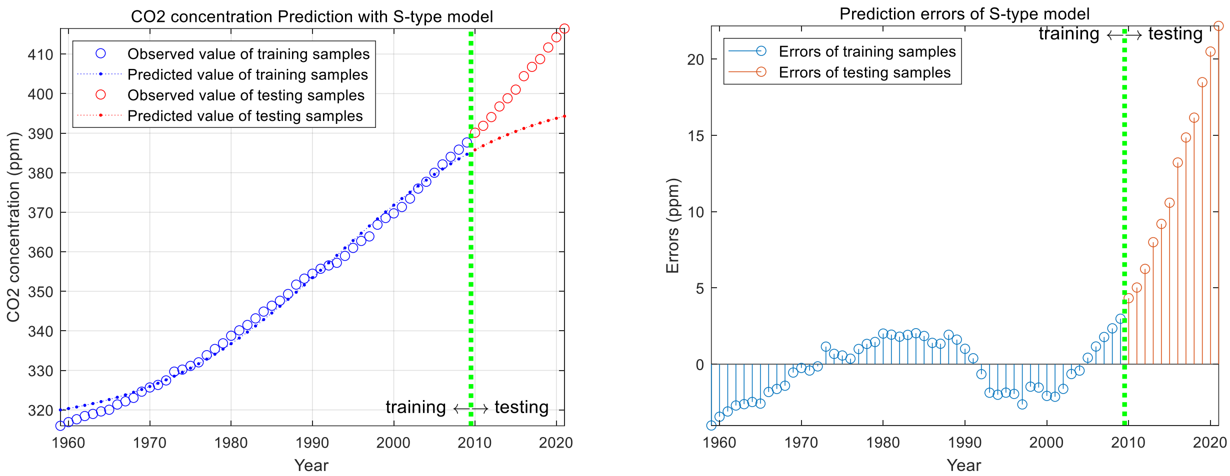



**Figure 4.11 $CO_2$ prediction and prediction error with S-type model**

### 4.3.3. ARIMA model

ARIMA model[5], also known as auto-regressive moving average model, is a kind of model that captures a set of different standard time structures in time series data. Its advantage is that it only needs endogenous variables instead of other exogenous variables. The ARIMA model can be thought of as a "filter" that separates the noise from the signal and then extrapolates the signal into the future to make predictions, suitable for fitting data showing non-stationarity.

The ARIMA model is disassembled, "AR" is auto-regression, used to analyze the previous values in the data, and make assumptions about them; "I" stands for synthesis, which replaces the data value with the difference between the data value and the original value; "MA" is the moving average, which is the sum of the error terms in the auto-regressive model. ARIMA model has three parameters: *p*, *d* and *q*, where *p* is used to represent the lag number of time series data itself, *d* represents the time series data which need to undergo several orders of difference differentiation to reach the most stable level, and *q* represents the lag number of prediction errors adopted in the prediction model. In the case of known parameters, the mathematical representation of ARIMA is:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}. \tag{4-3}$$

Of this, $\phi_i, (i = 1, \cdots, p)$, represents the coefficient of AR, and $\theta_j, (j = 1, \cdots, q)$, represents the coefficient of MA. When building an ARIMA model, the following steps are generally required.

(1) Obtain the time series data of the observed system, observe and test whether it is a stationary time series;

(2) A non-stationary sequence is transformed into a stationary sequence by the D-order difference operation;

(3) Calculate ACF (autocorrelation coefficient) and PACF (partial autocorrelation coefficient) for stationary time series, and identify ARIMA and other models;

(4) Determine the model parameters, test the obtained model, and make prediction and error analysis. The modeling process for stationary sequences can be represented by the Figure 4.12.

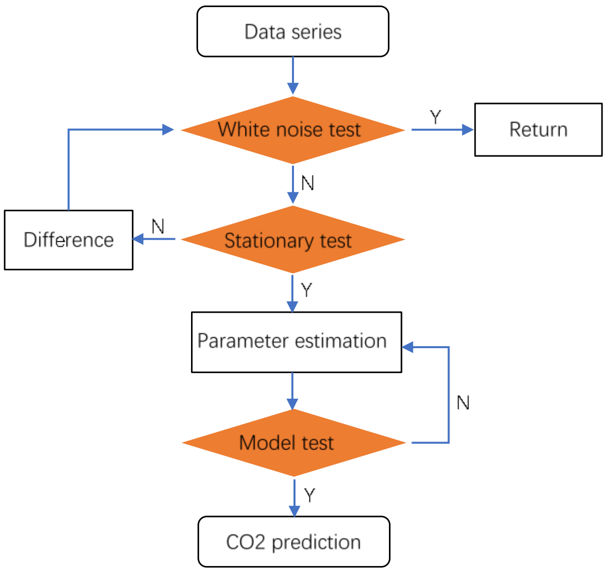



**Figure 4.12 ARIMA flow chart**

**Step 1 Sequence white noise test**

First of all, the sequence is tested for white noise. Ljung-Box test shows that the sequence is not white noise, and the next step can be carried out.

**Step 2 The stationary sequence is obtained by difference operation**

After the difference operation of the original sequence, the unit root test is used to show that the sequence has been stable after the second difference. The differential process is shown in Figure 4.13.

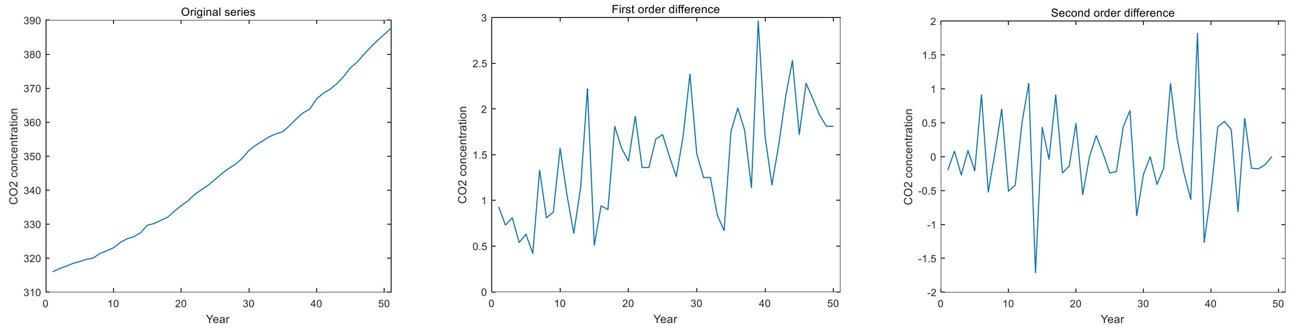



**Figure 4.13 $CO_2$ series differential processing**

**Step 3 Order identification based on AIC and BIC criterions**

According to the actual situation of the problem, the highest order of AR and MA models was set as 4, and AIC criterion and BIC criterion were used to seek optimization. The heat maps of AIC and BIC were shown in Figure 4.14. By combining the two figures, AIC was the main focus, and finally the model parameters were set as $p$=2 and $q$=3. According to the previous second-order difference, the final model is ARIMA (2,2,3).

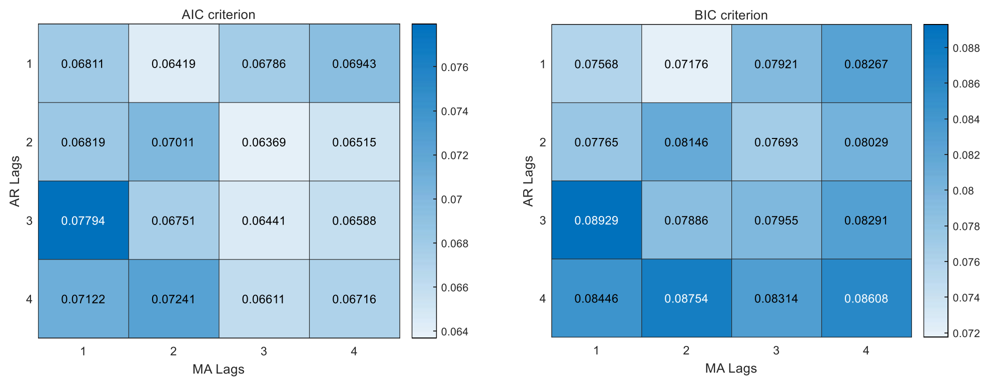



**Figure 4.14 Heat map of AIC and BIC criterion**

**Step 4 CO2 concentration prediction based on ARIMA(2,2,3)**

As before, we used the same segmentation method for the data set. 80% was used for training model and

20% was used for testing. The final prediction results and errors were obtained as shown in Figure 4.15.

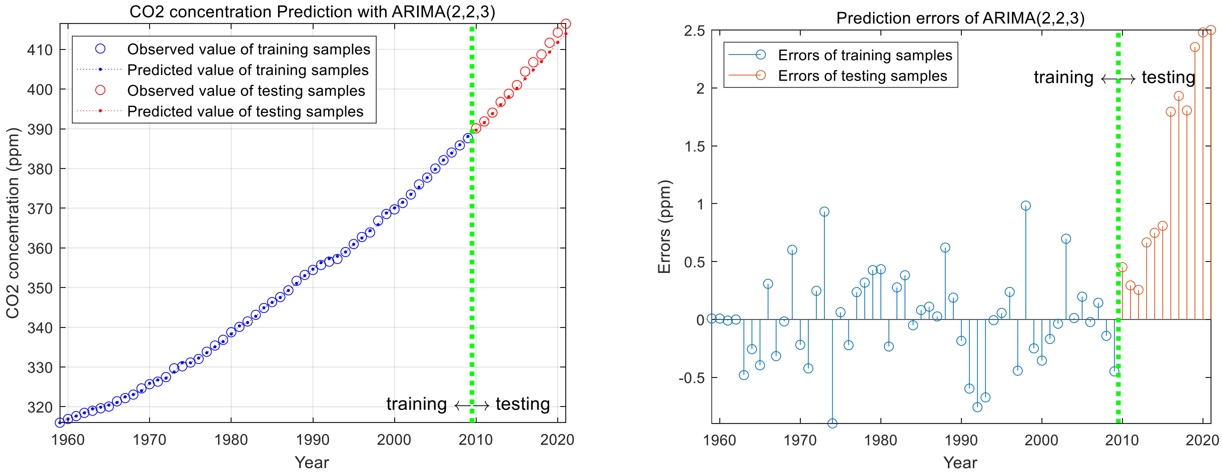



**Figure 4.15 $CO_2$ prediction and prediction error with ARIMA(2,2,3)**

### 4.3.4. Model testing and comparing

For the five models above, we construct three indicators for error comparison, namely root mean square error (RMSE), mean absolute error (MAE), mean relative error (MRE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4-4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{4-5}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{4-6}$$

Through calculation, the results are shown in Table 4.4.

**Table 4.4 The training error and testing error of each model**

| Model | Training data | | | Testing data | | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | MRE | RMSE | MAE | MRE |
| **Linear** | 2.3868 | 1.9686 | 0.0057 | 11.0870 | 10.5288 | 0.0260 |
| **Quadratic** | 0.6852 | 0.5752 | 0.0017 | 1.5488 | 1.2381 | 0.0030 |
| **Exponential** | 0.7456 | 0.6201 | 0.0018 | 0.4217 | 0.3729 | 0.0009 |
| **S-type** | 1.8342 | 1.6173 | 0.0047 | 13.6988 | 12.3921 | 0.0305 |
| **ARIMA(2,2,3)** | 0.3937 | 0.2981 | 0.0009 | 1.5854 | 1.3408 | 0.0033 |

### 4.3.5. Future prediction and conclusion

All the data from 1959 to 2021 were taken as the training set, and the model was fitted again according to the above method, and the model results were obtained as shown in Table 4.5.

**Table 4.5 Forecasting model of $CO_2$ concentration**

| Type | Model | R-square |
|---|---|---|
| **Linear** | $y = 1.6t - 2854.6$ | 0.98246 |
| **Quadratic** | $y = 0.013t^2 - 50.2758t + 48771$ | 0.99941 |
| **Exponential** | $y = 58.7202e^{1.0048t} + 256.024$ | 0.99945 |
| **S-type** | $y = \dfrac{1.2609}{1 + 2.5722e^{-4.2343t}}$ | 0.99287 |
| **ARIMA** | ARIMA(2,2,3) | 0.99983 |

The models in Table 4.4 are used to forecast the future until 2100. The forecast results are shown in Figure 4.16.

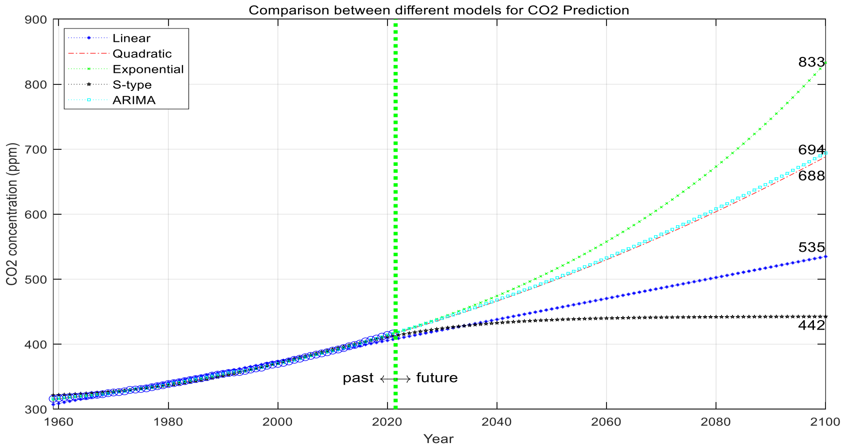



**Figure 4.16 $CO_2$ prediction of each model**

## 4.4. [1.c] Prediction results in 2100 and time points prediction of 685 ppm

### 4.4.1. CO2 prediction results in 2100

**Prediction results in 2100:** The corresponding $CO_2$ concentration predicting results of linear, quadratic, exponential, S-type and ARIMA(2,2,3) are 534.8825 ppm, 688.3268 ppm, 833.0377 ppm, 442.4825 ppm, 694.1067 ppm, respectively.

**Prediction results in 2050:** After model prediction, it is found that none of the five models can reach 685 ppm in 2050, the corresponding $CO_2$ concentration predicting results of linear, quadratic, exponential, S-type and ARIMA(2,2,3) are 454.1807 ppm, 496.8051 ppm, 512.6329 ppm, 437.4233 ppm, 498.8932 ppm, respectively. Figure 4.17 shows that the prediction result of the exponential model is the highest, while the linear model is the lowest. The results are shown in Table 4.6.

**Table 4.6 The $CO_2$ concentration prediction in the year 2100 by different models (ppm)**

| Model | Linear | Quadratic | Exponential | S-type | ARIMA(2,2,3) |
|---|---|---|---|---|---|
| $CO_2$ in 2100 (ppm) | 534.8825 | 688.3268 | 833.0377 | 442.4825 | 694.1067 |
| $CO_2$ in 2050 (ppm) | 454.1807 | 496.8051 | 512.6329 | 437.4233 | 498.8932 |
| Year of $CO_2$ 685 ppm | 2212.3 | 2099.3 | 2081.7 | *No Solution* | 2098 |

### 4.4.2. Predicted time points comparison of $CO_2$ concentration reaching 685 ppm

**Time points of $CO_2$ reaching 685 ppm:** After further calculation, we obtained the time points of $CO_2$ of each model reaching 685 ppm are 2212 (linear), 2099 (quadratic), 2082 (exponential), *No solution* (S –type) 2098(ARIMA). The predicted results are shown in Figure 4.17.

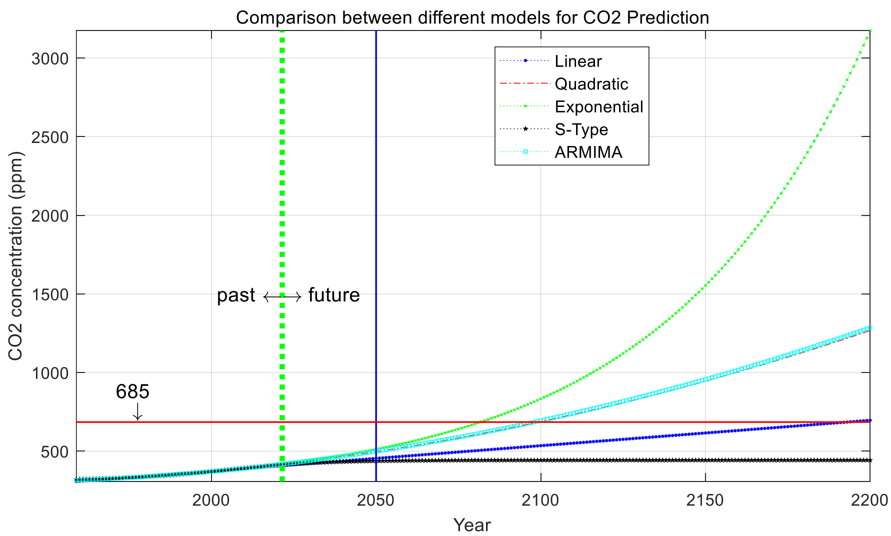



**Figure 4.17 $CO_2$ prediction of each model**

### 4.4.3. Conclusion

According to the prediction results of 2100, the highest value of the exponential model is 833.0377 ppm, while the lowest value of the S-type model is 442.4825 ppm.

In the prediction result of 2050, none of the models will reach 685 ppm.

The time points of 685 ppm are 2212 for linear model, 2099 for quadratic model, 2082 for exponential model, 2098 for ARIMA model, and the S-type prediction could not reach 685 ppm.

## 4.5. [1.d] Precision analysis of the models

With the analysis above, the exponential model is the most accurate. In the given 63 data from 1959 to 2021, both the training set and the test set have higher prediction accuracy than others, and the prediction error is the smallest in the test set. Since the 1960s, carbon emissions have been out of control for a long time, and this leads to the conformity of the concentration of $CO_2$ to the exponential distribution, which is consistent with the reality. We further used the exponential model to make the interval prediction at the 95% confidence level, with results shown in Figure 4.18.

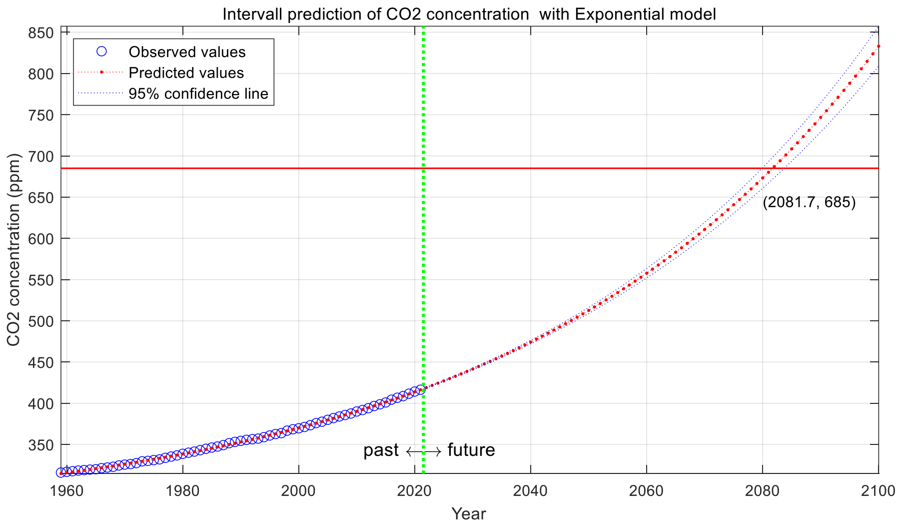



**Figure 4.18 Interval prediction of $CO_2$ prediction by exponential model**

# 5. Problem 2: Relationship between land-ocean temperature and $CO_2$

## 5.1. Model preparation: Pearson correlation analysis

According to Data set 1 and Data set 2, we create a scatter diagram of land-ocean temperature and $CO_2$ concentration, as shown in Figure 5.1.

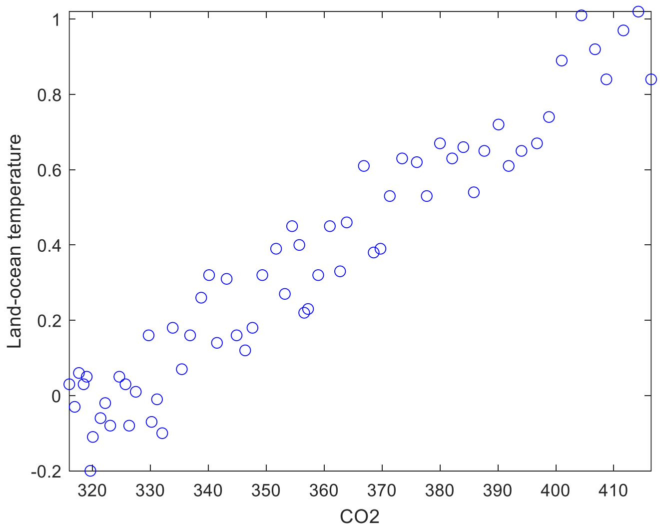



**Figure 5.1 Scatter figure of land-ocean temperature and CO2 concentration**

Pearson correlation coefficient is used to measure the correlation between two variables *X* and *Y*, and its range is between - 1 and 1. It is generally used to analyze the linear relationship between two variables. The formula is defined as follows.

$$r=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-Y)^2}} \tag{5-1}$$

If the value of *r* is between 0.8 and 1.0, the two variables have extremely strong correlation; 0.6~0.8 means strong correlation; 0.4~0.6 means medium correlation; 0.2~0.4 means weak correlation; 0~0.2 means little or no correlation.

Through calculation, ***r*=0.9613**, which shows that there is an extremely strong correlation between them.

## 5.2. Problem analysis: Overall idea of problem 2

For Problem 2, the temperature series are extracted from Data Set 2; and due to the extremely strong correlation between $CO_2$ and temperature, we refer to the methods of problem 1 and select the exponential model for prediction, use the least square method to fit the model parameters, and predict the time points when the temperature rise to 1.25℃, 1.5℃ and 2℃.

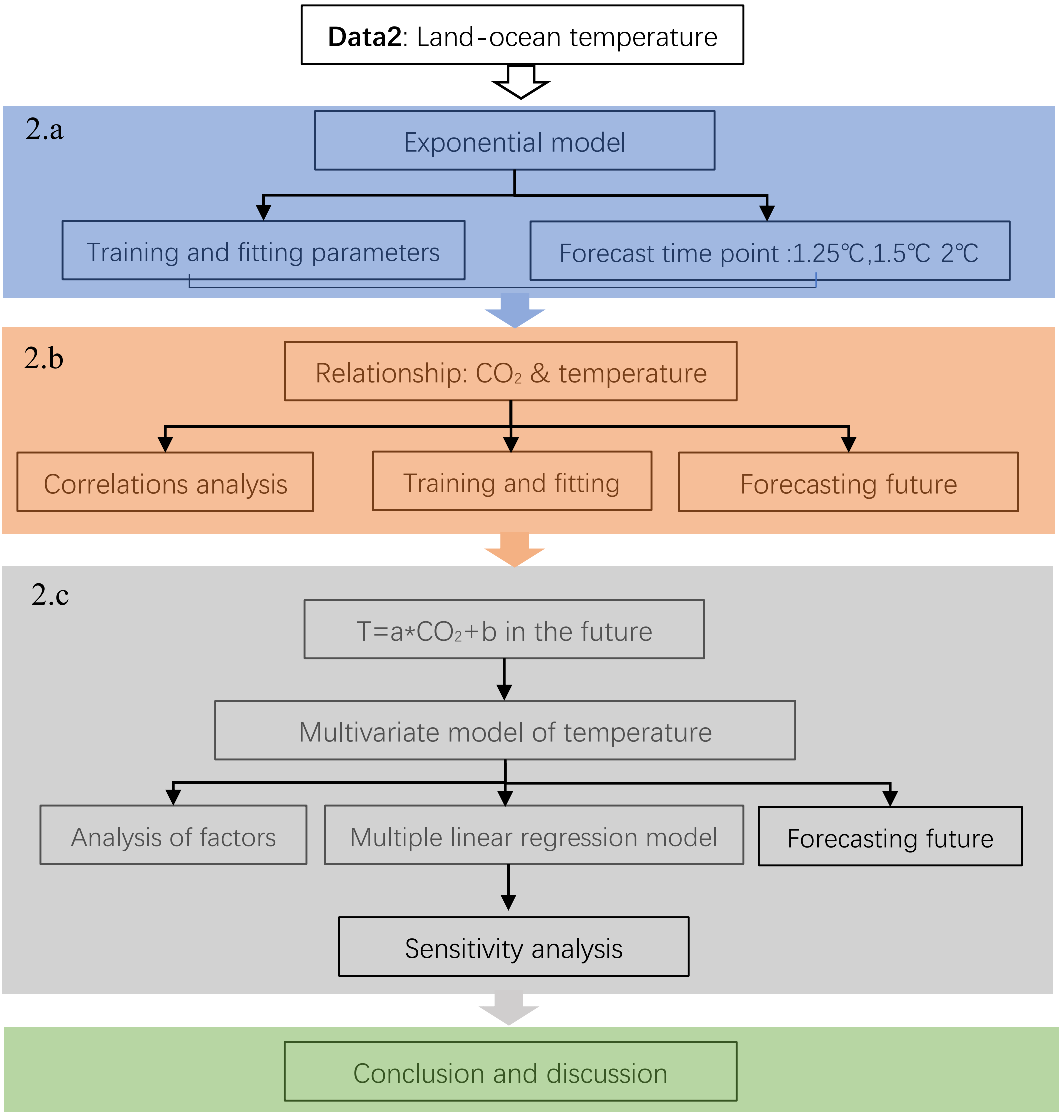



**Figure 5.2 The overall idea of problem 2**

Considering the influence of $CO_2$ on temperature, the relationship between them is explained through correlation analysis and other methods, and a linear regression model of temperature and $CO_2$ is further established to fit the model parameters and predict the future temperature.

The linear model of temperature with respect to $CO_2$, according to the characteristics of the model, the prediction for the future will lead to a continuous increase in temperature, which is not consistent with the actual situation. Therefore, in order to improve the model, literature [3] shows that, from the mechanism's point of view, radiation forcing is the direct factor affecting land-ocean temperature. Radiative forcing is a measure of the factors that affect the balance of incident and outgoing energy of the earth's atmospheric system. It is the most important index to characterize the potential factors of the climate change mechanism, with the unit of W/m$^2$ (*IPCC, 2021, Climate Change, https://www.ipcc.ch/report/ar6/wg1/* ). The radiative forcing is used by IPCC to objectively represent the changes in the radial energy budget of the earth's climate system. These changes can be caused by long-term changes in radioactive active substances, such as greenhouse gases (mainly including $CO_2$, $CH_4$, $N_2O$, etc.), aerosols, fuel combustion, solar radiation, etc., as well as other factors that can affect the surface absorption of radiation, such as the earth's reflectivity. These changes will lead to the imbalance of the energy budget of the earth's climate system and cause changes in the land-ocean temperature and other climate parameters, and finally form a new equilibrium state of the climate system. Therefore, we establish a multiple regression model based on the above factors to predict the future trend of land-ocean temperature.

The overall idea is shown in Figure 5.2.

## 5.3. [2.a] Temperature prediction based on exponential model

### 5.3.1. Exponential model of land-ocean temperature

With time $t$ as the independent variable and land-ocean temperature $T$ as the dependent variable, an exponential model was established, and the data from 1959 to 2021 were used for training. Considering the problem of data scale, the time $t$ was normalized, and the model was finally obtained as follows.

$$T = 0.5180e^{1.1163t} - 0.6034. \tag{5-2}$$

$R^2$=0.9184, RMSE=0.0922, MAE=0.0811, MRE=0.7484, the accuracy of model fitting is very good. The fitting and prediction results are shown in Figure 5.3.

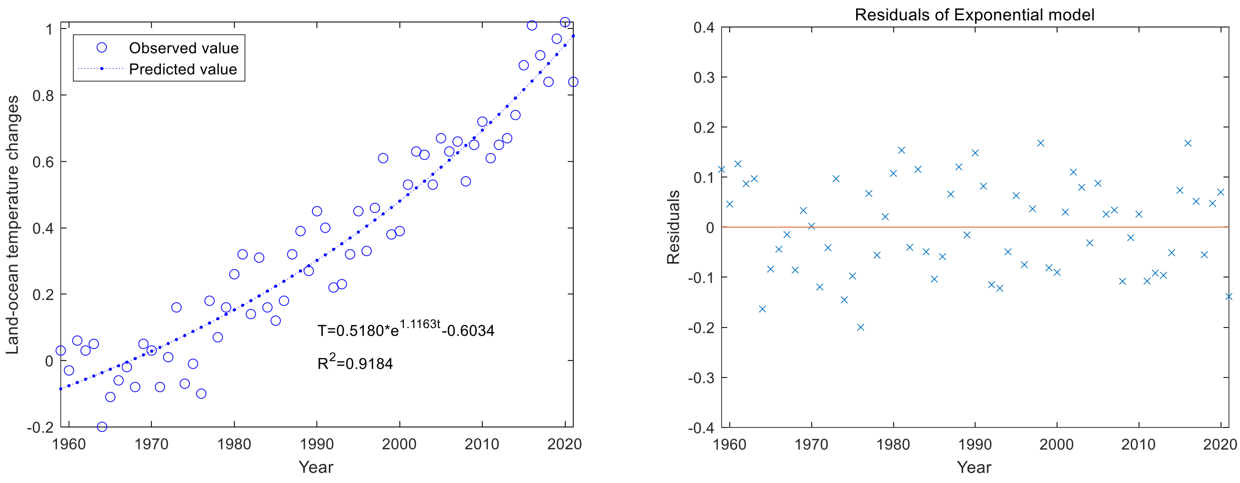



**Figure 5.3 Temperature prediction and error based on exponential model**

### 5.3.2. Forecasting time points of different land-ocean temperatures

The predicted time points of the exponential model for 1.25°C, 1.5°C and 2°C are 2028, 2036 and 2048, respectively. The prediction results are shown in Figure 5.4. The results of the temperature prediction are consistent with the conclusions shown in:" **This is well over the concentration level of 450 ppm required to have at least a 50% chance of stabilizing the climate at 2 degrees (2°C) global average temperature increase** "(*Organization for Economic Co-Operations and Development. (2012). The OECD environmental outlook to 2050,[Internet].https://www.oecd.org/env/cc/Outlook%20to%202050_Climate%20Change%20 ChapterHIGLIGHTS-FINA-8pager-UPDATED%20NOV2012.pdf.*).

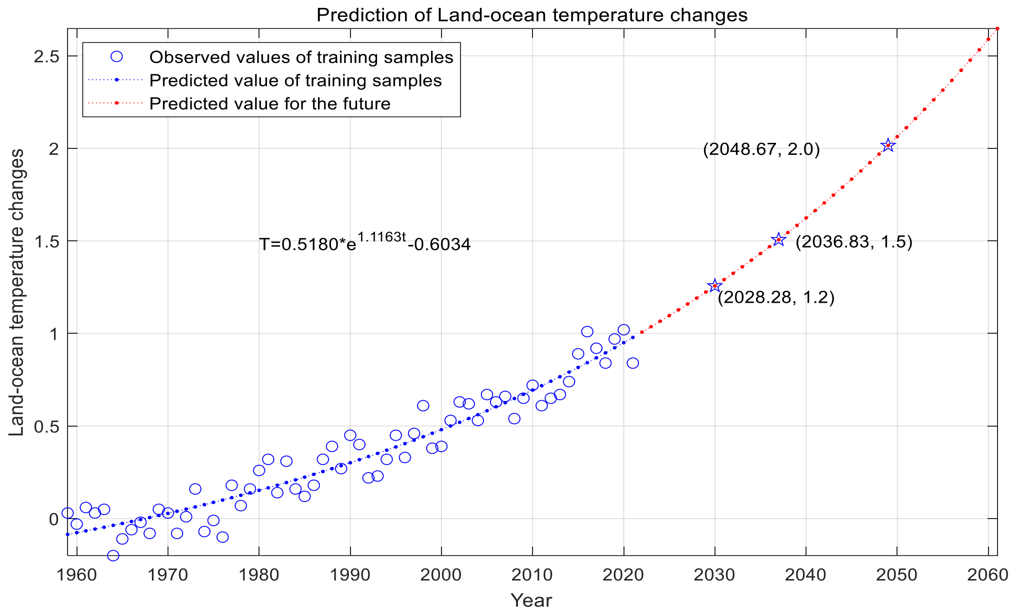



**Figure 5.4 The key time points prediction by the exponential model**

## 5.4. [2.b] Correlation analysis and temperature prediction based on linear regression

Through calculation, $r$=0.9613, indicating an extremely strong correlation between $CO_2$ concentration and the land-ocean temperature. Thus, a linear model is built. With $CO_2$ concentration as the independent variable and land-ocean temperature as the dependent variable, a linear regression model is established, and the data from 1959 to 2021 are used for training.

$$T = 0.105CO_2 - 3.3926. \tag{5-3}$$

$R^2$=0.9241, $F$=753.0363, $P$=0.000. These statistics indicate that the model performs well. The prediction results and fitting errors are shown in Figure 5.5.

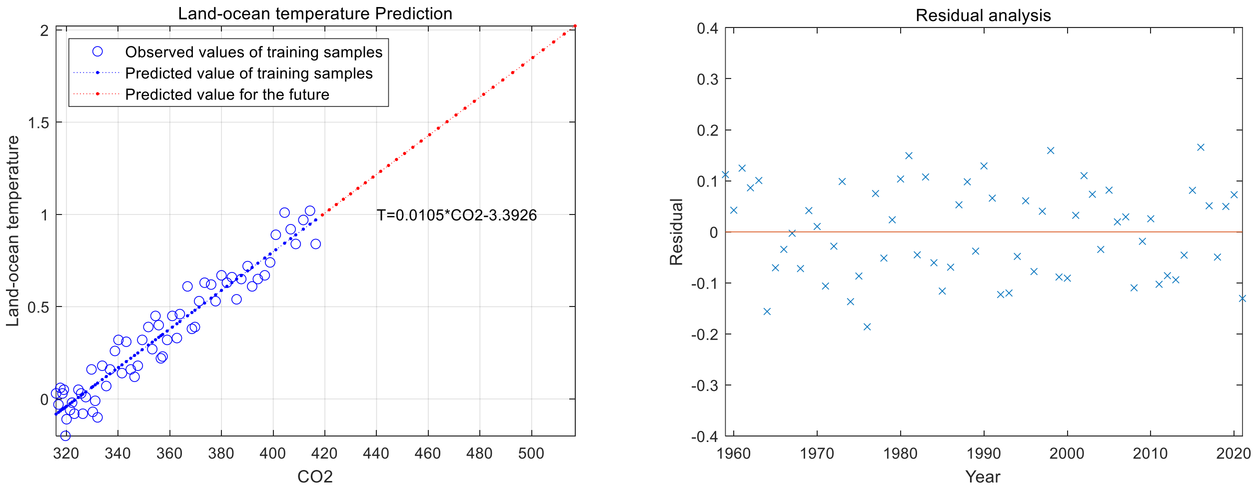



**Figure 5.5 Temperature prediction and error analysis based on linear model**

Through the model (5-3), the land-ocean temperature is predicted with $CO_2$ as the independent variable. The results show that the time points when the temperature rise to 1.25℃, 1.5℃ and 2℃ are 2030, 2038 and 2051, respectively, which are very close to the results of the exponential model, indicating that this linear prediction model is reasonable. The results are shown in Figure 5.6.

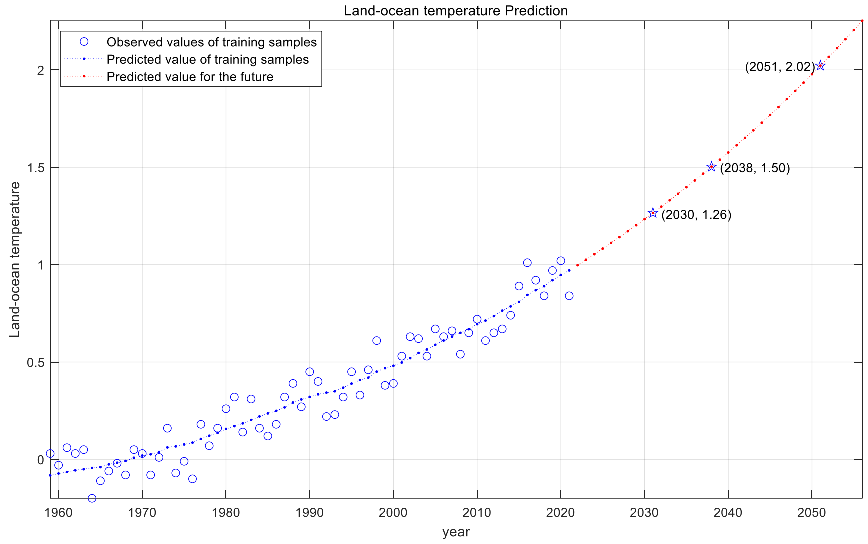



**Figure 5.6 The key time points prediction by linear model**

## 5.5. [2.c] Multivariate model for land-ocean temperature prediction

### 5.5.1. Irrationality of long-term prediction by linear models

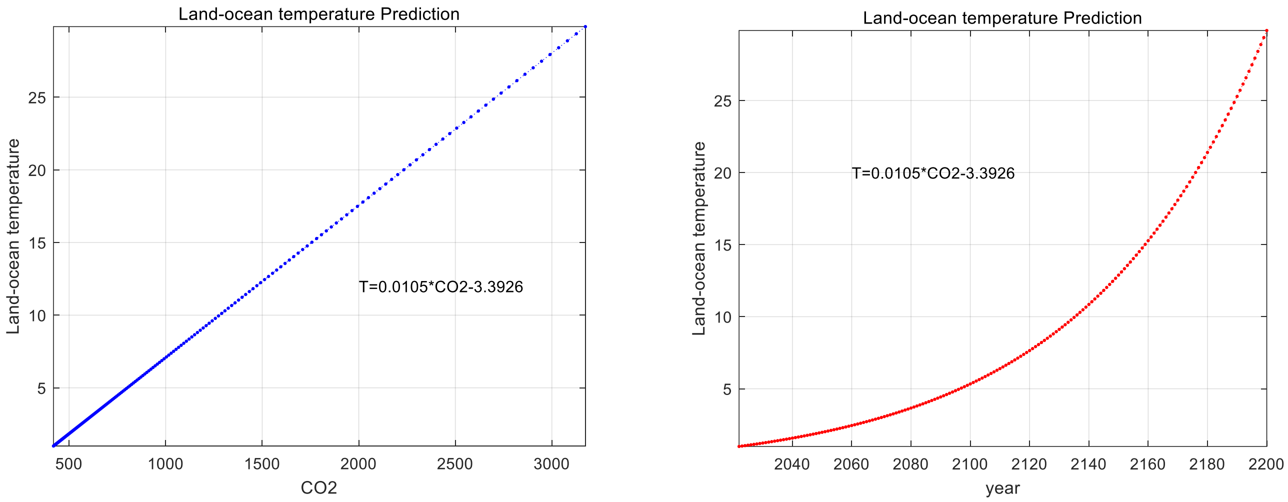



**Figure 5.7 Temperature forecasting based on linear model**

Figure 5-7 shows the results predicted by model (5-3) to 2200. As shown in Figure 5-7, the temperature has risen to 30 degrees by 2200, indicating that linear models are unreasonable for long-term forecasting. The linear model of temperature with respect to $CO_2$, according to the characteristics of the model, the prediction for the future will lead to a continuous increase in temperature.

### 5.5.2. Multivariate analysis for land-ocean temperature and data collection

According to the one-dimensional radiative conduction cycle model [6], there is a correlation between radiative forcing and feedback (climate response, such as temperature change) (*Friedrich T, Timmermann A, Tigchelaar M, et al. Nonlinear climate sensitivity and its implications for future greenhouse warming[J]. Science Advances, 2016, 2(11): e1501923*). Therefore, by sorting out the data of various major factors closely related to climate change during 1989-2021, especially greenhouse gases, the radiative forcing intensity can be converted into land-ocean temperature by relevant calculation formulas to analyze the linear relationship between global radiative forcing change and temperature change.

According to the earth's energy budget model [7], the relationship between land-ocean temperature and radiative forcing (and hence related factors) is shown in Figure5.8. There are six main factors related to climate change.

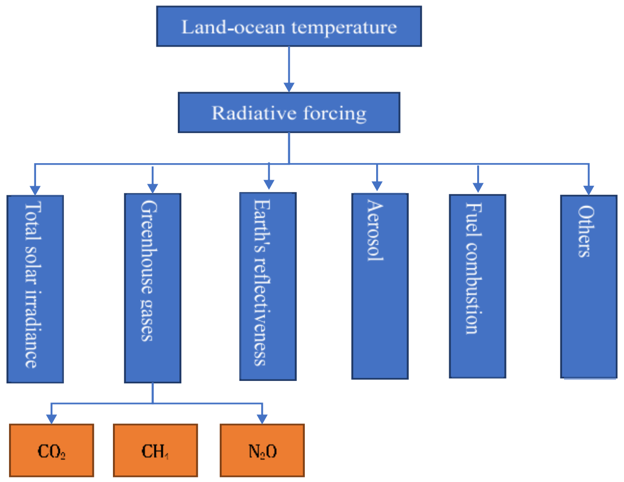



**Figure 5.8 Factors structure analysis of land-ocean temperature**

**(1) Total solar irradiance ($RF_{solar}$)**

The sun is the source of all energy in the climate system, so changes in solar radiation can cause changes in radiative forcing in the climate system. *Data period: 1989-2021. Data source: NASA. Measuring Solar Radiation Incident on Earth. Space Research, 2021.*

**(2) Greenhouse gases ($RF_{CO_2} + RF_{CH_4} + RF_{N_2O}$)**

Greenhouse gases refer to the natural and man-made gases in the atmosphere that can absorb and re-emit infrared radiation. They can cause greenhouse effect through the long-wave radiation from the surface to the outer space, thus increasing the temperature of the earth surface. Some concise formulas for calculating radiative forcing of greenhouse gas are based on the PICC Full Report (2001, p. 358).

$$CO_2: RF_{CO_2} = \alpha_1\big(g(C) - g(C_0)\big), \tag{5-4}$$

$$CH_4: RF_{CH_4} = \alpha_2 \ln\big(\sqrt{M} - \sqrt{M_0}\big) - \big(f(M, N_0) - f(M_0, N_0)\big), \tag{5-5}$$

$$N_2O: RF_{N_2O} = \alpha_3 \ln\big(\sqrt{N} - \sqrt{N_0}\big) - \big(f(M_0, N) - f(M_0, N_0)\big), \tag{5-6}$$

$$f(M, N) = 0.47 \ln(1 + 2.01 \times 10^{-5}(MN)^{0.75} + 5.31 \times 10^{-15} M(MN)^{1.52}), \tag{5-7}$$

$$g(C) = \ln(1 + 12C + 0.005C^2 + 1.4 \times 10^{-6}C^3), \tag{5-8}$$

where $C$ is $CO_2$ in ppmv, $M$ is $CH_4$ in ppbv, $N$ is $N_2O$ in ppbv, and $\alpha_1 = 3.35, \alpha_2 = 0.36, \alpha_3 = 0.012, C_0 = 278, M_0 = 700, N_0 = 270$. *Data period: 1989-2021. Source: https://gml.noaa.gov/ccgg/trends/ and reference [3].*

**(3) Earth's reflectiveness**

Different landforms on the earth have different reflectance of solar radiation, which is reflected by the surface and transmitted to the atmosphere. *Data source: https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-chapter2-1.pdf are constant in recent years, return not to consider.*

**(4) Aerosol ($RF_{aerosol}$)**

Aerosols are liquid or solid particles suspended in the air that reflect solar radiation directly and absorb both long and short-wave radiation [8]. But unlike the long-term heating effects of greenhouse gases, the contribution of aerosols to radiative forcing is generally negative because of their short atmospheric lifetime. *Data period: 1989-2021. Data source: Jia H, Ma X, Yu F, et al. Significant underestimation of radiative forcing by aerosol -- cloud interactions derived from satellite-based methods [J]. Nature communications, 2021, 12(1): 1-11.*

**(5) Fuel Combustion ($RF_{fuel}$)**

The burning of fossil fuels releases a large amount of heat and greenhouse gases such as $CO_2$. Here we only consider the impact of heat released by fuel consumption on global temperature change. *Data period: 1989-2021. Data source:https://www.bp.com.cn/content/dam/bp/country-sites/zh_cn/china/home/reports/statistical-review-of-world-energy/2021/BP_S tats_2021.pdf.*

**(6) Others**

Some other factors such as volcanic eruption, sudden El Nino phenomenon and other astral radiation, are excluded in the study of global climate change, because the sixth factor does not follow a certain rule, or there is only a small-time scale; in addition, comparatively, the main factor (solar radiation) shows a big difference (for example, the moon radiation is only one millionth of the solar radiation).

### 5.5.3. A multivariate temperature prediction model based on radiative forcing

**(1) Model construction and parameters fitting based on historical data**

Based on the multi-factor analysis in Section 5.5.2, we ignore the effects of earth reflectance and other factors, and consider solar radiation, fuel combustion and greenhouse gases as independent variables. Among them, greenhouse gases are decomposed into $CO_2$, $CH_4$ and $N_2O$, and land-ocean temperatures are dependent variables. The multiple linear regression model is built as follows:

$$T = \alpha_0 + \alpha_1 RF_{solar} + \alpha_2 RF_{fuel} + \alpha_3 RF_{aerosol} + \alpha_4 RF_{CO_2} + \alpha_5 RF_{CH_4} + \alpha_6 RF_{N_2O}, \quad (5\text{-}9)$$

where $T$ is land-ocean temperature, $\alpha_i = 0,1,2,\cdots,6$ is the regression coefficient. $RF_{solar}$, $RF_{fuel}$, $RF_{aerosol}$, $RF_{CO_2}$, $RF_{CH_4}$, $RF_{N_2O}$ respectively represents radiative forcing of the total solar radiation, fuel, aerosol, carbon dioxide, methane, nitrous oxide. Using data collected in the previous section (1989-2021), the least squares regression was carried out, and the model was obtained as follows:

$$T = -11.2063 - 0.4078 RF_{solar} - 142.7224 RF_{fuel} - 24.6518 RF_{aerosol} + 11.0078 RF_{CO_2} + 5.3996 RF_{CH_4} + 10.3846 RF_{N_2O}. \quad (5\text{-}10)$$

The goodness of fitting $R^2$ of this model is 0.8781.

**(2) Eliminate multi-collinearity of factors based on and principal components analysis**

In model (5-10), the coefficients of $RF_{solar}$, $RF_{fuel}$, $RF_{aerosol}$ are negative, which is not in conformity to reality. We think there are some multi-collinearity in the 6 factors. We compute the Person correlation coefficients of factors, and create a heat map, as is shown in Figure 5.9.

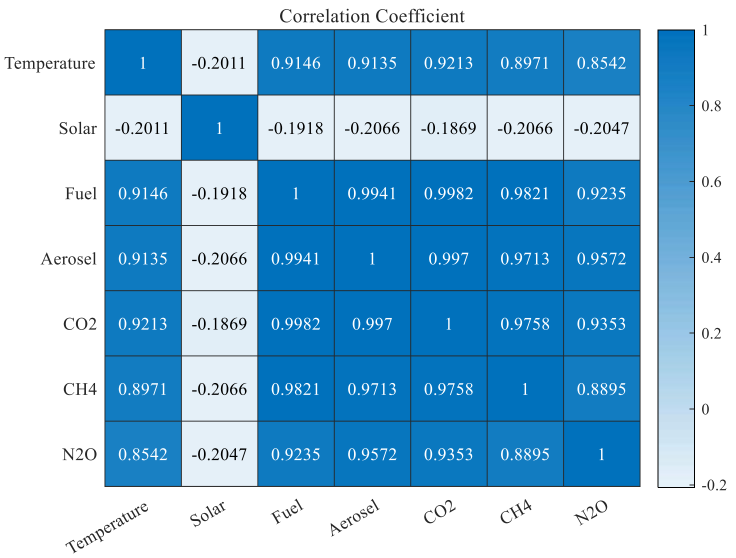



**Figure 5.9 Correlation analysis heat map of all factors**

Figure 5.9 shows that there are extremely strong or strong correlation between factors. Principal component analysis[9] is a commonly used technology to extract important variables (in form of components) from a large set of variables available in a data set, eliminating the multi-collinearity of factors. For six factors, 2 components are extracted, and cumulative variance contribution rate is greater than 99%. We build a principal component regression (PCR) model, as shown in equation 5-11.

$$T = -0.5361 - 1144 RF_{solar} + 0.0385 RF_{fuel} + 0.4375 RF_{aerosol} + 1.4396 RF_{CO_2} + 0.2128 RF_{CH_4} + 0.1526 RF_{N_2O}. \quad (5\text{-}11)$$

Equation (5-11) is the final prediction model, $R^2$ is 0.8484, the coefficients of $RF_{fuel}$, $RF_{aerosol}$, $RF_{CO_2}$, $RF_{CH_4}$, $RF_{N_2O}$ are positive, which shows that these factors have positive effect on temperature, but, the $RF_{solar}$ is negative, we find it consist with what is found in [4], Therefore, we believe that the model (5-11) is completely reasonable.

**(3) Prediction of future temperature by multiple linear regression model**

In order to predict the future based on the model (5-11), the total solar radiation $RF_{solar}$, fuels $RF_{fuel}$, aerosols $RF_{aerosol}$, carbon dioxide $RF_{CO_2}$, methane $RF_{CH_4}$ and nitrous oxide $RF_{N_2O}$ need to be predicted first, and the predicted data period is 2022-2100. Except for the predicted value of $CO_2$, which was directly taken from problem 1.c, the prediction of other data was predicted through the data collected in section 5.5.2. The specific prediction methods of these six factors are as follows.

**Solar radiation ($RF_{solar}$):** According to the variation law of sunspots, the cycle of sunspots is about 11 years, and the predicted value of $RF_{solar}$ is obtained by averaging the values of the same cycle moment in each cycle contained in the historical data;

**Fuel combustion ($RF_{fuel}$):** According to the sample data, the prediction model $RF_{fuel} = 0.003t -$

0.6672 was obtained, with goodness of fitting $R^2$=0.9889;

**Aerosol ($RF_{aerosol}$):** According to the sample data, the prediction model $RF_{aerosol} = 0.0040t - 8.0854$ was obtained, with goodness of fitting $R^2$=0.9995;

**$CO_2$($RF_{CO_2}$):** The data are obtained according to the predicted results of the exponential model in question 1.c;

**CH4($RF_{CH_4}$):** According to the sample data, the prediction model $RF_{CH_4} = 0.0019t - 3.2976$ was obtained, and goodness of fitting $R^2$=0.9428;

**Nitrous oxide ($RF_{N_2O}$):** Prediction model $RF_{N_2O} = \frac{0.1463}{1+0.5294e^{-30605t}}$ is obtained according to sample data; and goodness of fitting $R^2$=0.9467.

The predicted values of the above six factors were substituted into the model (5-11) and finally into the multivariate prediction model of land-ocean temperature, as shown in Figure 5.10. The results of the model show that the predicted temperature in 2050 is 1.781℃ and that in 2100 is 3.3889 ℃, which is close to the current research results.

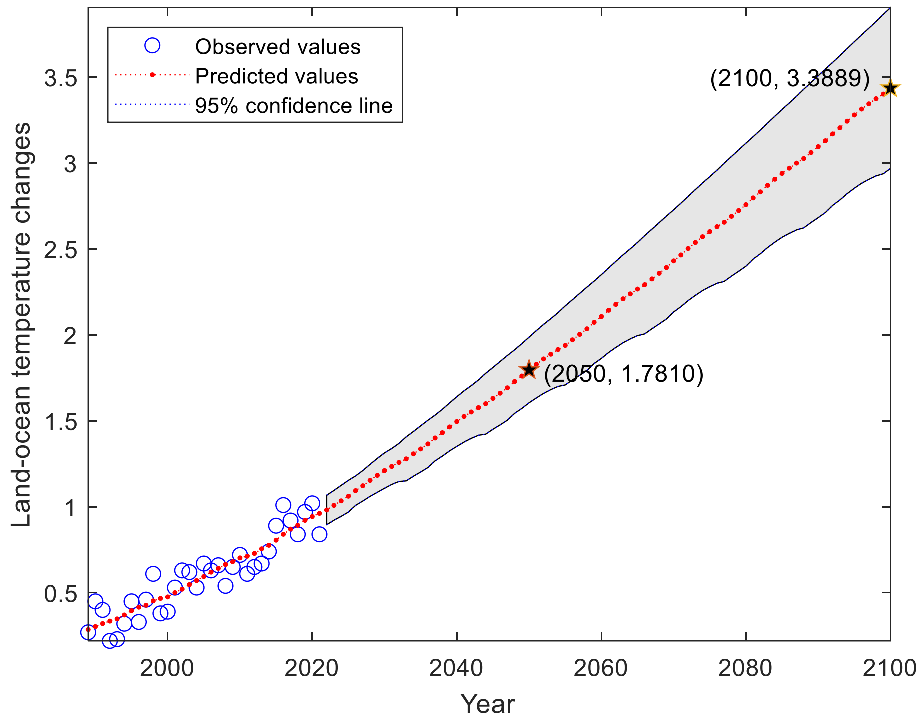



**Figure 5.10 Prediction results of temperature based on multivariate linear regression model**

### 5.5.4. Sensitivity analysis of $CO_2$ level to land-ocean temperature

**(1) Land-ocean temperature changes under different $CO_2$ levels**

Human activities are constantly changing, posing impact on carbon emissions, which further affects the changes of land-ocean temperature. According to the model (5-11), we increase or decrease the $CO_2$ level by 5% and 10% in the future, and the results are shown in Figure 5.11. The results show that if the $CO_2$ level increases by 5%, for example, the land-ocean temperature change in 2100 will increase from 3.3889℃ to 3.572℃, and if the $CO_2$ level increases by 10%, it will increase to 3.7560℃. Similarly, if the $CO_2$ level reduced by 5%, the changes of land-ocean temperature in 2100 will decrease to 3.2050℃, and if the $CO_2$ level reduced by 10%, it will decrease to 3.0220℃.

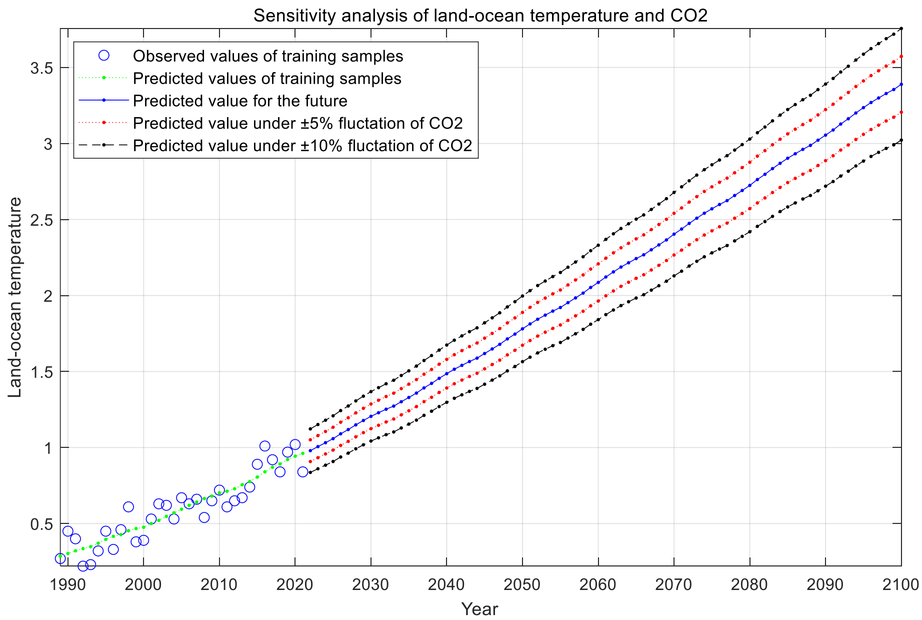



**Figure 5.11 Land-ocean temperature changes under different $CO_2$ levels**

**(2) Importance analysis of the six factors**

The standardized coefficients of factors in model (5-11) are -0.1144, 0.0385, 0.4375, 1.4396, 0.2128, 0.1526, respectively, which implies that the importance of factors is: $CO_2$>aerosol> $CH_4$>$N_2O$>solar>fuel. $CO_2$ is the most important effect on land-ocean temperature, so humankind should reduce carbon emissions through technological progress, industrial restructuring and national cooperation.

### 5.6. Conclusion and discussion

For the prediction of land-ocean temperature, the exponential model of time $t$ and the linear regression of $CO_2$ are used. The two model results have little difference. Since $CO_2$ has a strong correlation with temperature, the regression prediction of $CO_2$ is more reasonable. The land-ocean temperature change is a very complicated system problem. This paper tries to establish a multiple linear regression model based on multiple factors, and the model results are reasonable.

## 6. Strengths and Weaknesses

### 6.1. Strengths

1. The model adopted in this paper is simple and easy to understand, and can be easily solved and predicted.
2. The prediction of land-ocean temperature in section 5.5 of problem 2 of this paper takes into account the influence of more factors on land-ocean temperature, and the model is more in line with the reality.

### 6.2. Weaknesses

1. In problem 1 and 2, models used to predict $CO_2$ and temperature are created without considering more influencing factors.
2. The collected data in problem 2 may has some errors, which may affect the accuracy of the prediction model.
3. The establishment of multiple regression model in section 5.5.3 ignores the influence of earth's reflectance and other factors on radiative forcing, which may lead to inaccurate prediction.

## 7. One Page Non-technical Article

# Is carbon dioxide the main culprit of global warming?

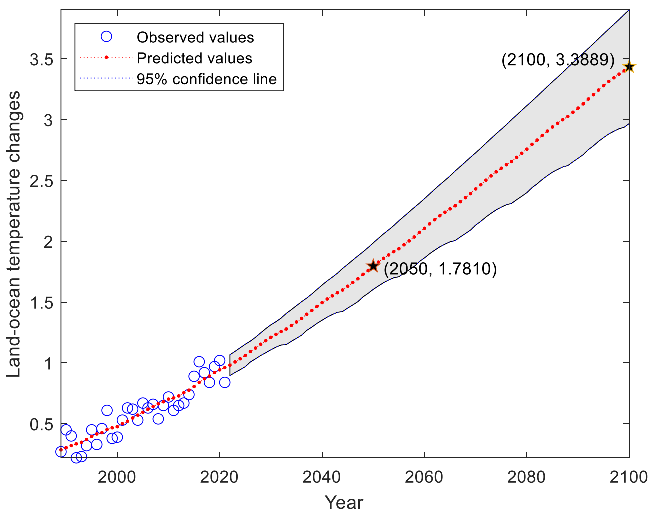
***Article | Nov 15, 2022 | Environment***


**Global warming** will lead to more extreme climate consequence, such as the melting of glaciers, the destruction of ecological chains, frequent El Nino events, the greater spread of infectious diseases, and so on. The IPCC's fourth Global Climate Change Assessment report states that it is no longer a controversial fact that human activities emit greenhouse gases, especially $CO_2$, which will lead to global warming.

**Some experts** noted that the $CO_2$ concentration increase in 2004 has led to the biggest 10-year average increase at that time, while others noted that the $CO_2$ concentration would reach 685 ppm in 2050. Are these arguments correct?

First, **linear regression models** and computer techniques are used to demonstrate that the $CO_2$ concentration did reach the largest 10-year average increase at 1995-2004, and which verify the effect of this increase is significant. **Five models** are built to predict the $CO_2$ concentrations in the future, and all models demonstrated lower results than expert's predictions for $CO_2$ concentrations in 2050. **The exponential model, which is the best-performing model**, shows that the $CO_2$ concentrations will not reach 685 ppm until 2082. Finally, by using the exponential model again, we predict the future time points of 1.25°C increase from the baseline temperature (1951-1980), where are **1.25°C, 1. 5°C and 2°C in 2028, 2036 and 2048**, respectively. If the correlation of $CO_2$ concentration and land-ocean temperature is considered, the time points of the future land-ocean temperature increase by 1.25°C, 1.5°C and 2°C will be delayed by 2 years, 2 years and 3 years, respectively.

We further discuss the main factors closely related to climate change--**solar radiation, greenhouse gas, earth reflectivity, aerosol, fuel combustion,** and develop a **multiple linear regression model** according to the radiation forcing theory. The model results show the global land-ocean temperature will rise to 1.7810°C by 2050. Regarding the sensitivity analysis of $CO_2$ concentration, it is also confirmed that changes in $CO_2$ concentration will significantly affect the global land-ocean temperature.




Our study shows that if no action is taken, the exponential increase in $CO_2$ concentrations will result 3.3889°C increase for global land-ocean temperature by 2100, when humans will be devastated. Therefore, we need to plant trees, reduce the use of fossil fuels, actively develop clean energy, reduce carbon emissions, in order to reduce $CO_2$ concentrations and slow down the trend of global warming change.

**Nature does not need human beings; it is the otherwise!**

## References


[1] Pachauri R K, Reisinger A. IPCC fourth assessment report. ***IPCC***, Geneva, 2007.

[2] World Meteorological Organization (WMO), Climatological Normals (CLINO) for the period 1961–1990, ***WMO/OMM 847***, 768 pp., Geneva, Switzerland, 1996.

[3] Huabing Li. Quantitative analysis of global warming and prediction of future global average temperature changes. ***Master Thesis, Zhejiang University***, 2007.

[4] Catherine Brahic. Sun's activity rules out link to global warming. ***NewScientist,*** 11 July 2007.https://www.newscientist.com/article/dn12234-suns-activity-rules-out-link-to-global-warming/

[5] Pena D, Tiao G C, Tsay R S. A course in time series analysis[M]. John Wiley & Sons, 2011.

[6] Friedrich T, Timmermann A, Tigchelaar M, et al. Nonlinear climate sensitivity and its implications for future greenhouse warming. ***Science Advances,*** 2016, 2(11): e1501923.

[7] Kiehl J T, Trenberth K E. Earth's annual global mean energy budget. ***Bulletin of the American meteorological society,*** 1997, 78(2): 197-208.

[8] Bellouin N, Quaas J, Gryspeerdt E, et al. Bounding global aerosol radiative forcing of climate change. ***Reviews of Geophysics,*** 2020, 58(1): e2019RG000660.

[9] Suryanarayana T M V, Mistry P B. Principal component regression for crop yield estimation[M]. ***Singapore: Springer,*** 2016.