

Big data climate prediction model based on linear regression

Yuxing Wang ^{*,#}, Yifan Shi [#]

School of Physics and Technology, Nantong University, Nantong, China, 226019

* Corresponding author: 19517112297@163.com

[#]These authors contributed equally.

Abstract. This paper deeply studies the trends of climate change and ocean temperature change in British Columbia, which is of great significance for the prediction of climate change and the formulation of corresponding countermeasures. First, the paper obtained the historical temperature changes and ocean temperature data from 1981 to 2010 at 13 sites in British Columbia through the Canadian government website, and conducted data preprocessing. Then through linear fitting, we build a climate prediction model. Finally, through the analysis, we found that the temperature showed different trends with the change of latitude or longitude, and in the time dimension, we found that the temperature in British Columbia showed a rising trend from 1950 to 1999, and the temperature rising trend was roughly 1.5°C. In addition, it is found that the global sea surface temperature is on the rise.

Keywords: Climate Trend Prediction Models, Big Data, Linear Regression.

1. Introduction

As the effects of global climate change become increasingly apparent, the frequency and intensity of extreme weather events are increasing. These extreme weather events, such as heavy rains, floods, hurricanes, droughts and extreme heat, bring huge losses and impacts to human societies and economies [1].

Liu Jianguo's research [2] found that global temperatures continued to rise, with 2015 to 2022 becoming the eight hottest years on record. In addition, extreme heat is occurring frequently in many regions. Liu Zhaohui [3] also pointed out that the phenomenon of high temperature and heat waves in cities is obvious, and these extreme weather events may reshape the global weather pattern.

Zhang Piaoyin et al. [4] predicted the global terrestrial extreme climate change based on the CMIP6 ensemble optimization data set. They found that in the future scenario, both global land temperature and precipitation extremes are expected to increase, which has important implications for disaster risk management in various regions. The research of Liu Anan et al. [5] has made important progress in using Marine big data and deep learning algorithms, successfully predicting Marine extreme weather phenomena such as El Nino, typhoons and short rainfall. These studies provide important reference for future climate prediction and disaster management.

In summary, the increasing frequency and intensity of extreme weather events present new challenges and opportunities for future climate prediction and disaster management. There is still a long way to go in the study of global climate change, and international cooperation needs to be further strengthened to jointly deal with the challenges brought by climate change.

2. Data Preprocessing

2.1. Data source

This paper selects British Columbia, Canada, as the research object, because British Columbia is located in the southwest of Canada, bordering the Atlantic Ocean in the west, and most of the province is forested, the province is roughly a parallelogram. Therefore, according to their geomorphology and latitude and longitude, the weather station data of 13 locations were selected as research data. The specific distribution is shown in Figure 1 (the red dots are the selected locations). This data comes from the Government of Canada

(http://climate.weather.gc.ca/historical_data/search_historic_data_e.html).
The Python program given in the sample is used for reading.

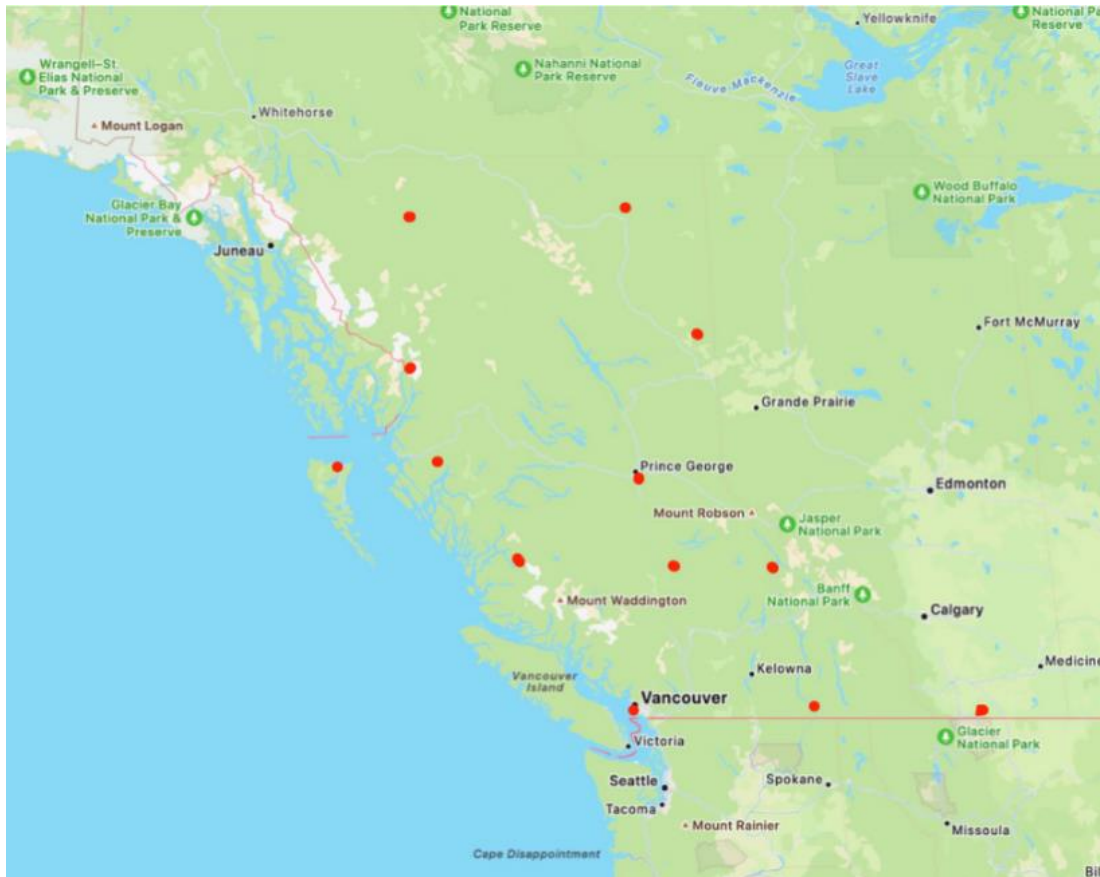


Figure 1. Distribution of data sources

2.2. Data preprocessing

Preprocessing mainly includes removing unique attributes, dealing with missing values and outliers, feature encoding, and so on. The specific processing methods and steps are as follows:

(1) Remove unique attributes

The unique attributes are usually some id attributes, which do not describe the distribution of the sample itself, so simply delete these attributes. In addition to the temperature data required for this question, there are a large number of other climate data (rainfall, snowfall, precipitation, etc.), so it is necessary to delete other useless data.

(2) Deal with missing values and outliers

Missing data is when a portion of the required data is missing from the dataset. A data anomaly is one that is collected because it deviates significantly from other data or is contrary to common sense. In general, due to equipment failure, data recording errors, data loss, and other external factors, the data we collect often has missing values and outliers.

The paper utilizes outlier boxes to handle outliers effectively while preserving the overall data structure. Outlier bin is a method of dealing with outliers that divides the observed values in a data set into different bins (or groupings) based on their numerical range or feature distribution. By separating outliers from non-outliers, outliers can be handled more efficiently while preserving the overall structure of the data.

(3) Normalization

The normalized linear transformation method is used to standardize the row data to solve the comparability between data indicators. After the original data of the supply of different influencing factors are standardized, each index is in the same order of magnitude, which is convenient for comprehensive comparison and evaluation. So using normalization can make them spread between 1

and 100, and the weighting process can be added to get the total score they assess to get their importance.

$$S = \frac{R_t - R_{\min}}{R_{\max} - R_{\min}} \times T \quad (1)$$

3. Climate data and ocean data analysis

3.1. Trending model for historical temperature data in British Columbia, Canada

3.1.1. Model processing steps

The model uses the control variable method to control a series of different dimensions (49°N, 52°N, 54°N) unchanged and longitude changes, and a series of different longitudes (120°W, 123°W) unchanged and latitude changes.

3.1.2. Research methodology

(1) Control variable method

The control variable method is used in the Monte Carlo method to reduce variance a technical approach. This method reduces the error in the estimation of unknown quantities by knowing the known quantities. In physics, for multi-factor (multi-variable) problems, the method of controlling factors (variables) is often adopted, turning multi-factor problems into multi-factor problems [6].

(2) Linear fitting method

Linear fitting, also known as first-order fitting, is a form of curve fitting. Let x be the observed quantities y , and is a function of: x , the curve fitting is through, and the observations of, to find the best $y = ax + b$ estimate of parameter b , and to find the best theoretical curve x $y = ax + b$.

$$S(a, b) = \sum_{k=1}^n [(a + bx_k) - y_k]^2 \quad (2)$$

The geometric background of the problem is to find a straight line so that the sum of squares of the longitudinal distance between the line and the plane scatter determined by the data table is minimized as shown in Figure 2 Scatterplot linear fit [7].

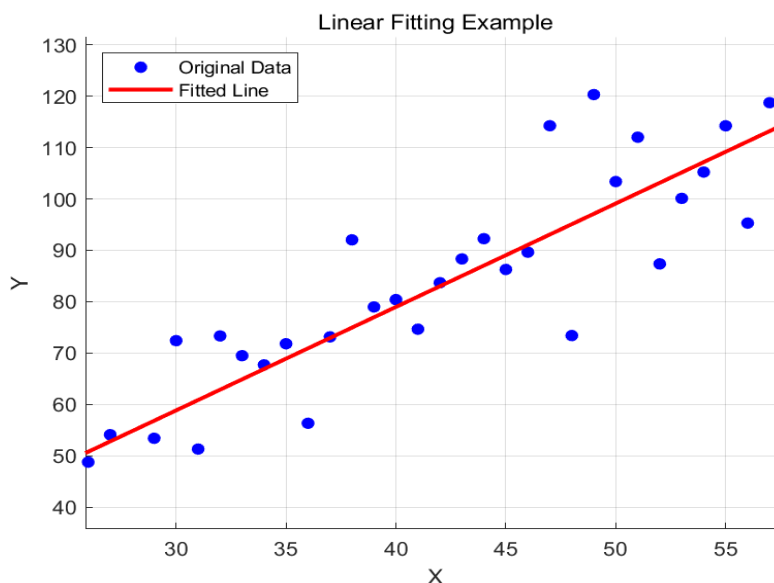


Figure 2. Scatter plot linear fit plot

The function derives the two variables:

$$\frac{\partial S}{\partial a} = 2 \sum_{k=1}^n [(a + bx_k) - y_k] \quad (3)$$

$$\frac{\partial S}{\partial b} = 2 \sum_{k=1}^n [(a + bx_k) - y_k] x_k \quad (4)$$

For equations (4.2) and (4.3), the simultaneous solution obtains:

$$\begin{cases} ma + \sum_{k=1}^n x_k b = \sum_{k=1}^n y_k \\ \sum_{k=1}^n x_k a + \sum_{k=1}^n x_k^2 b = \sum_{k=1}^n x_k y_k \end{cases} \quad (5)$$

Solving the system of equations (4) gives the values of two constants a , b and thus gives the linear fitting function $y = ax + b$.

3.1.3. Result analysis

According to the needs of the topic, the model processes and analyzes the data from the time dimension and the spatial dimension respectively.

(1) Analysis of spatial dimension results

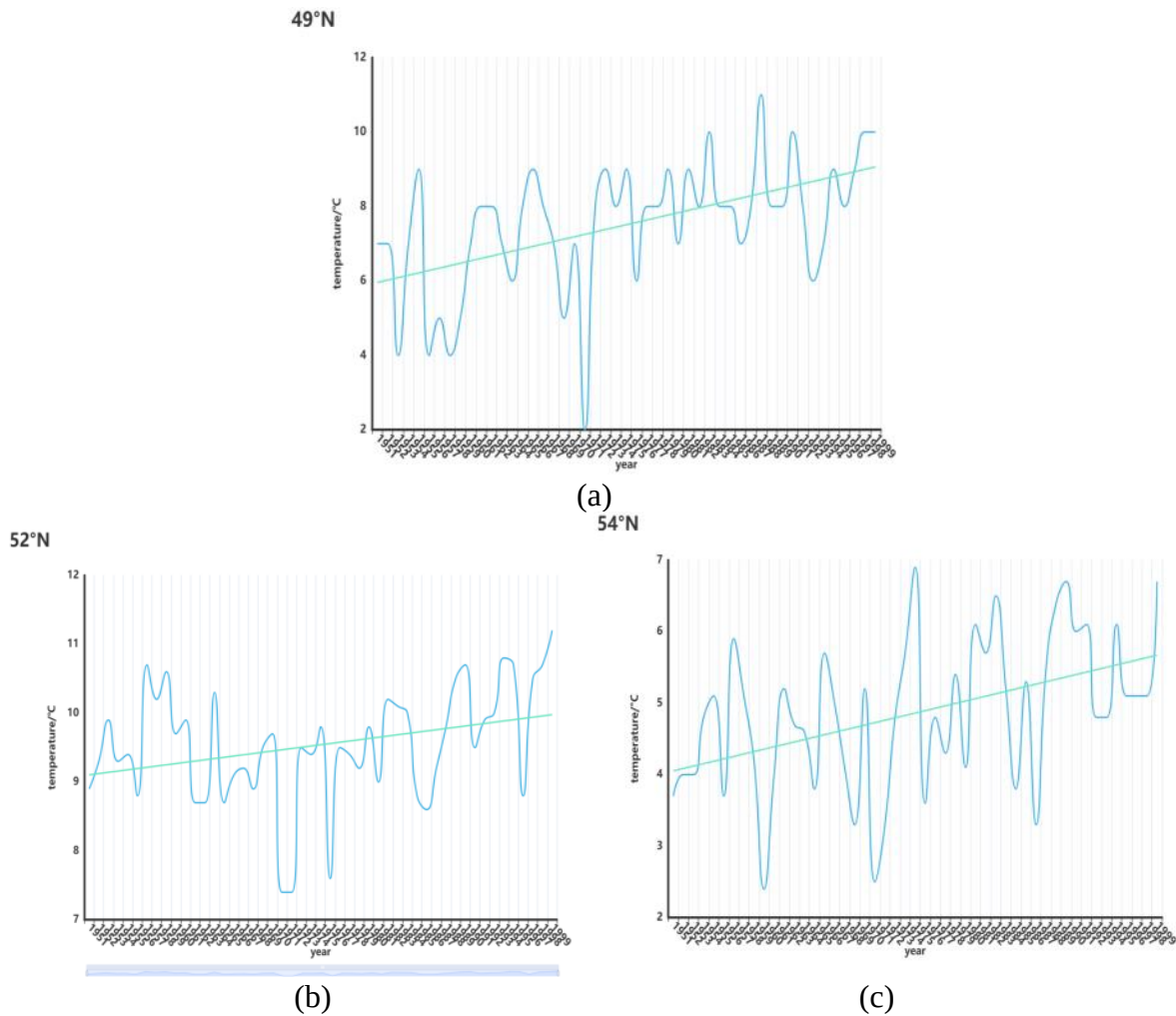


Figure 3. Historical average temperature for British Columbia, Canada as a function (a) 49°N, (b) 52°N, (c) 54°N

The specific data changes are shown in Table 1. Figure 3 Historical average temperature changes with longitude in British Columbia, Canada.

Table 1. Historical mean temperature as a function of longitude in British Columbia, Canada

Latitude	Toponym	Longitude/°	Average	Median	Standard deviation	Heating fit/°C
49°N	Vancouver	-123.18	9.9743	9.6000	5.2751	0.99
	Penticton	-119.6	9.0380	8.7000	7.9504	0.994
	Fernie	-115.07	4.0732	2.4000	8.0212	1.032
52°N	Bella Coola	-126.69	7.7115	8.0000	6.6515	1.716
	Williams Lake	-122.05	4.1786	4.6500	8.7211	1.34
	Blue River	-119.28	4.2953	4.4000	9.0300	1.259
54°N	Masset	-132.3	6.9702	6.2000	5.0676	0.417
	Kitimat	-128.7	6.6393	6.5000	7.0086	1.278
	Prince George	-122.68	3.6228	4.9000	9.1948	2.121

From the average and median columns of Table 1 Historical mean temperature as a function of latitude in British Columbia, Canada, it can be analyzed that the temperature data shows a downward trend with the decrease of longitude (westward shift) when the latitude is unchanged, as shown in Figure 4 The average and median of historical mean temperature as a function of longitude. The standard deviation column shows an overall upward trend with decreasing longitude (westward shifting), as shown in Figure 5 Historical mean temperature standard deviation with longitude. The temperature rise fit from 1950 to 1999 was basically around 1.5°C [8].

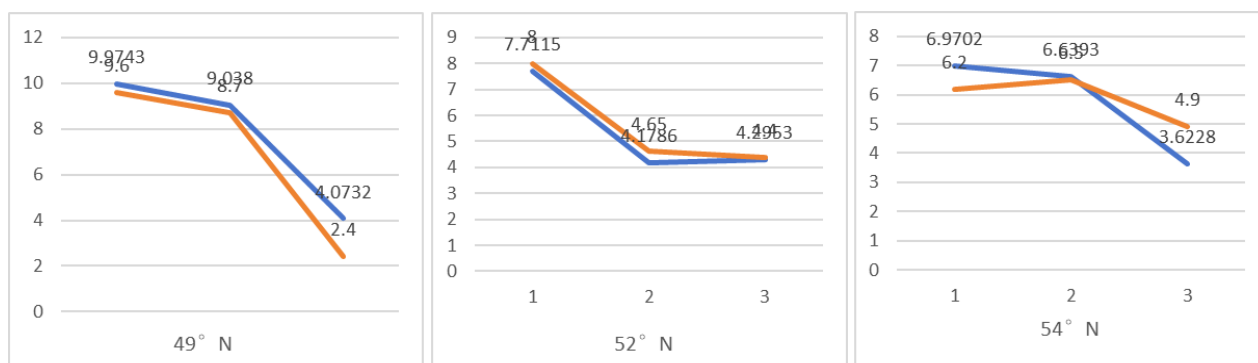


Figure 4. Historical mean and median temperatures as a function of longitude

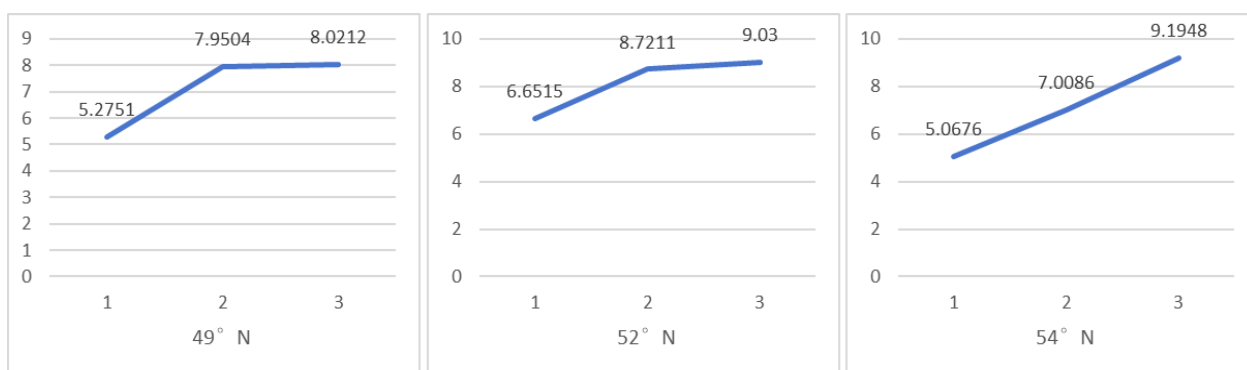


Figure 5. Historical mean temperature standard deviation as a function of longitude

Longitude does not change, latitude changes

To ensure that the historical temperature data of British Columbia, Canada have no change in longitude and latitude, and the average temperature data of a series of fixed latitude values (see Figure 1 Data Source Distribution Map for spatial distribution) are respectively taken to plot the results of temperature with latitude, as shown in Figure 6 The trend of historical average temperature data in British Columbia, Canada with latitude [9].

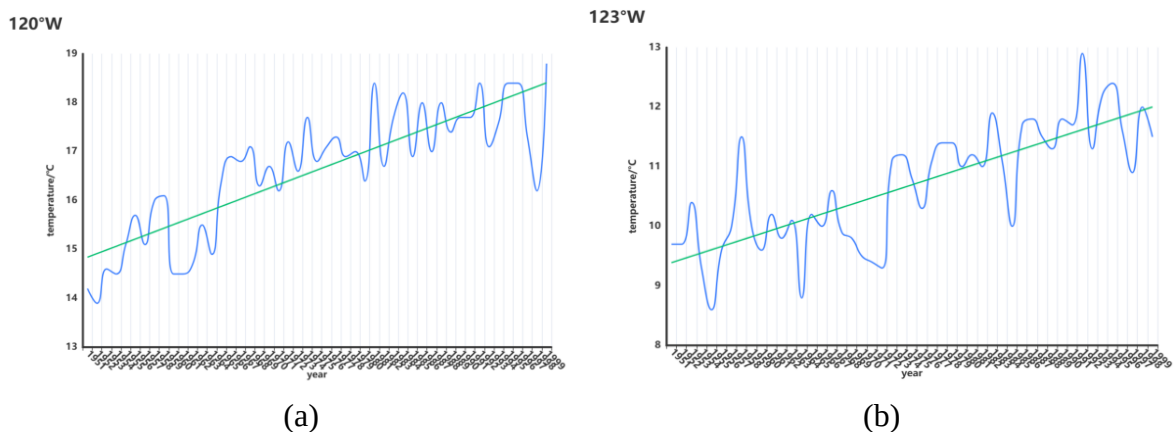


Figure 6. Historical average temperature in British Columbia, Canada as a function of latitude (a) 120°W, (b) 123°W

The specific data changes are shown in Table 2 The change of historical mean temperature with latitude in British Columbia, Canada.

Table 2. Historical mean temperature as a function of latitude in British Columbia, Canada

Longitude	Toponym	Latitude/°	Average	Median	Standard deviation	Heating fit/°C
120°W	Penticton	49.46	9.0380	8.7000	7.9504	0.994
	Blue River	52.15	4.2953	4.4000	9.0300	1.259
	Fort st John	56.24	1.5860	3.8000	11.5475	2.1436
123°W	Vancouver	49.2	9.9743	9.6000	5.2751	0.99
	Prince George	53.89	3.6228	4.9000	9.1948	2.121
	Fort Nelson	58.84	-1.0883	1.3000	14.2282	1.8798

From the average and median columns of the table of historical mean temperature as a function of longitude in Table 2, it can be analyzed that the temperature data shows a downward trend as the latitude rises (moves upward) when the longitude is constant, as shown in Figure 7 The historical mean and median temperature as a function of latitude. The standard deviation column shows an overall upward trend in temperature data as latitude rises (moving westward), as shown in Figure 8 Historical mean temperature standard deviation with latitude. The temperature rise fit from 1950 to 1999 was basically around 1.5°C.

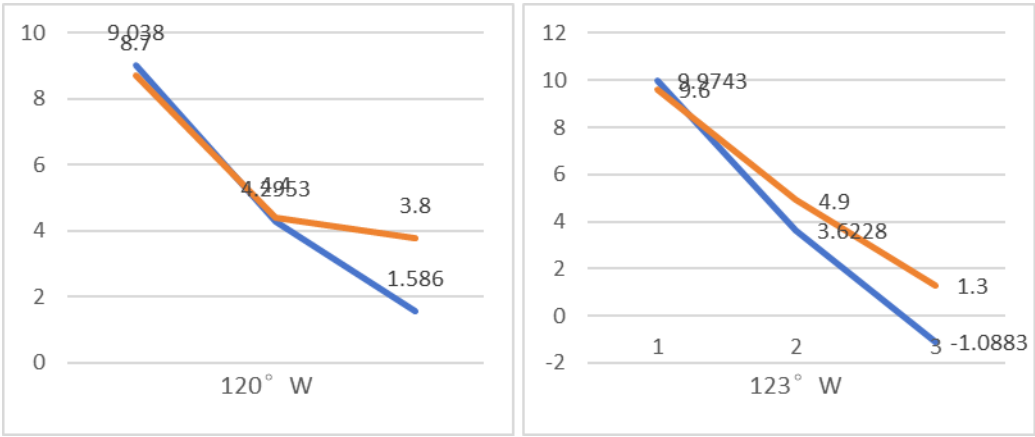


Figure 7. Historical mean and median temperatures as a function of latitude

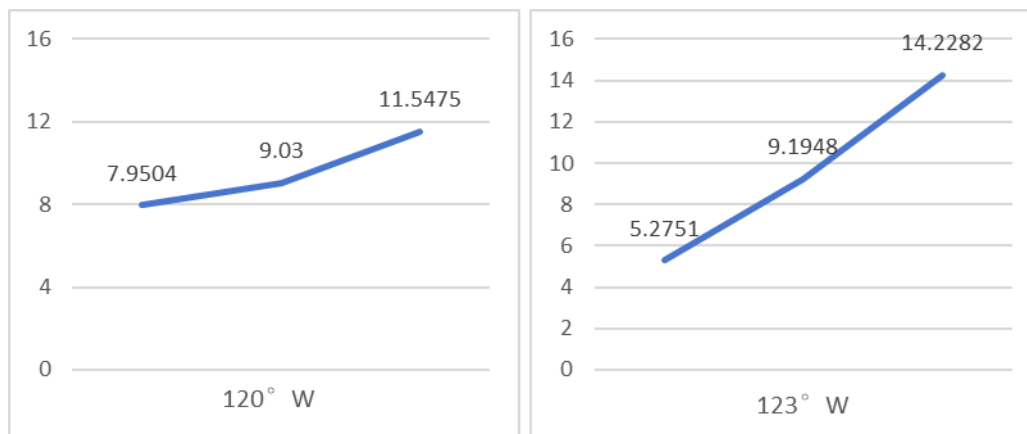


Figure 8. Historical mean temperature standard deviation as a function of latitude

(2) Analysis of time dimension results

The analysis of the time dimension in British Columbia, Canada, is as follows:

From Figure 2, Figure 5, Table 1 and Table 2, it can be seen that the historical average temperature has also been increasing over time, that is, the British Columbia Province in Canada was warming between 1950 and 1999. The average temperature in the region has risen by about 1.5°C over the past 50 years.

Historical climate data for a more detailed study of four representative locations in the region are illustrated in Figure 9. The trend of historical temperature data over time in British Columbia, Canada, shows that the historical data of all types of temperatures have increased over time.

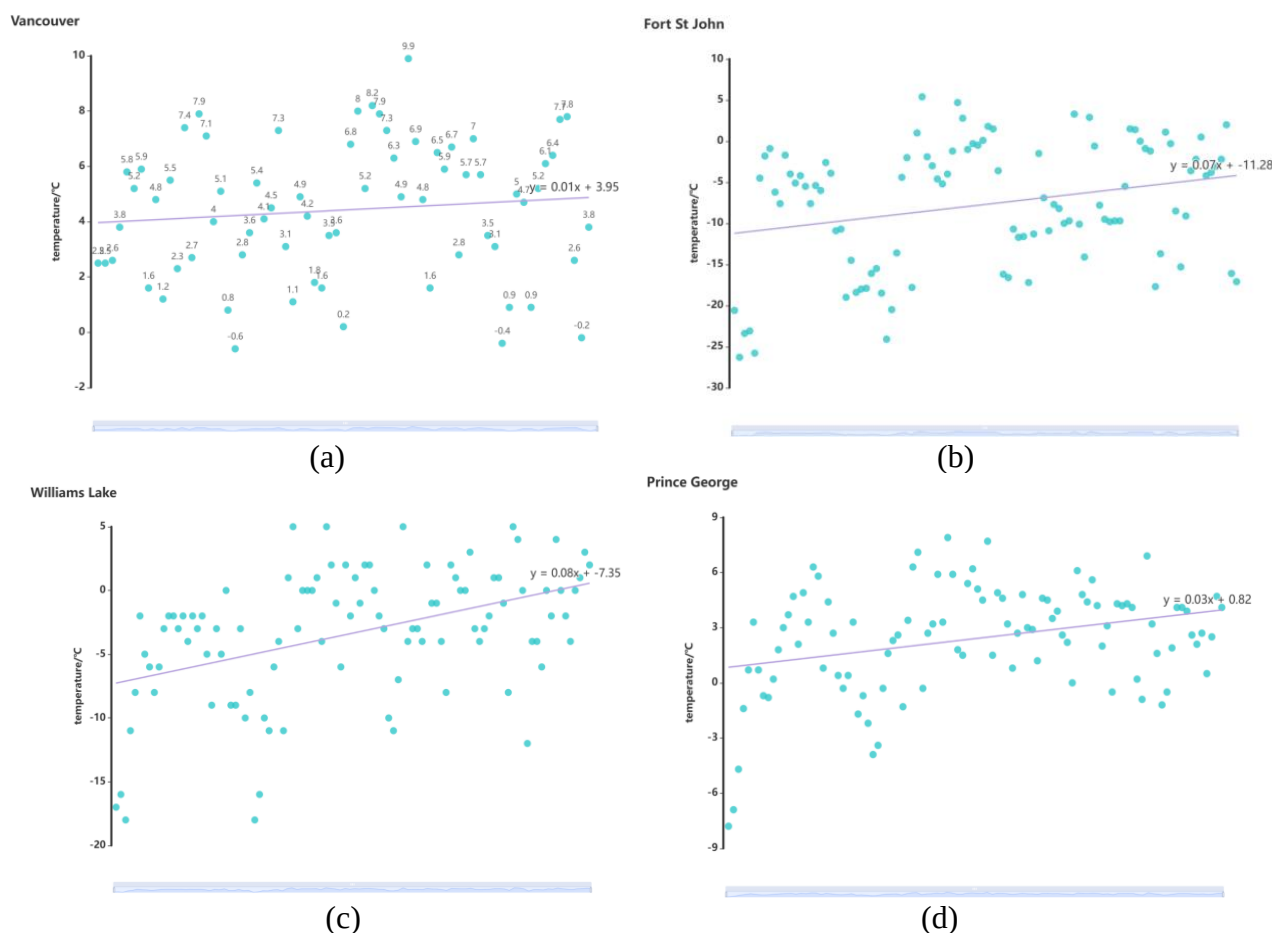


Figure 9. Historical temperature trends over time across British Columbia, Canada (a)Vancouver; (b)Fort st John; (c)Willams Lake; (d)Prince George

3.2. Trending models for historical ocean surface temperature data

The model uses Python to read the sample of sea surface temperature data given in the question, and then analyzes the read data to find the changing trend of sea surface temperature from 1981 to 2010. This sample data is weekly sea surface temperature data after January 1, 1981, and can be analyzed for trends using the same solution as in the first part. Finally, it can be seen that the overall trend of sea surface temperature is gradually rising. These trends are shown in Figure 10 sea surface temperature trends [10].

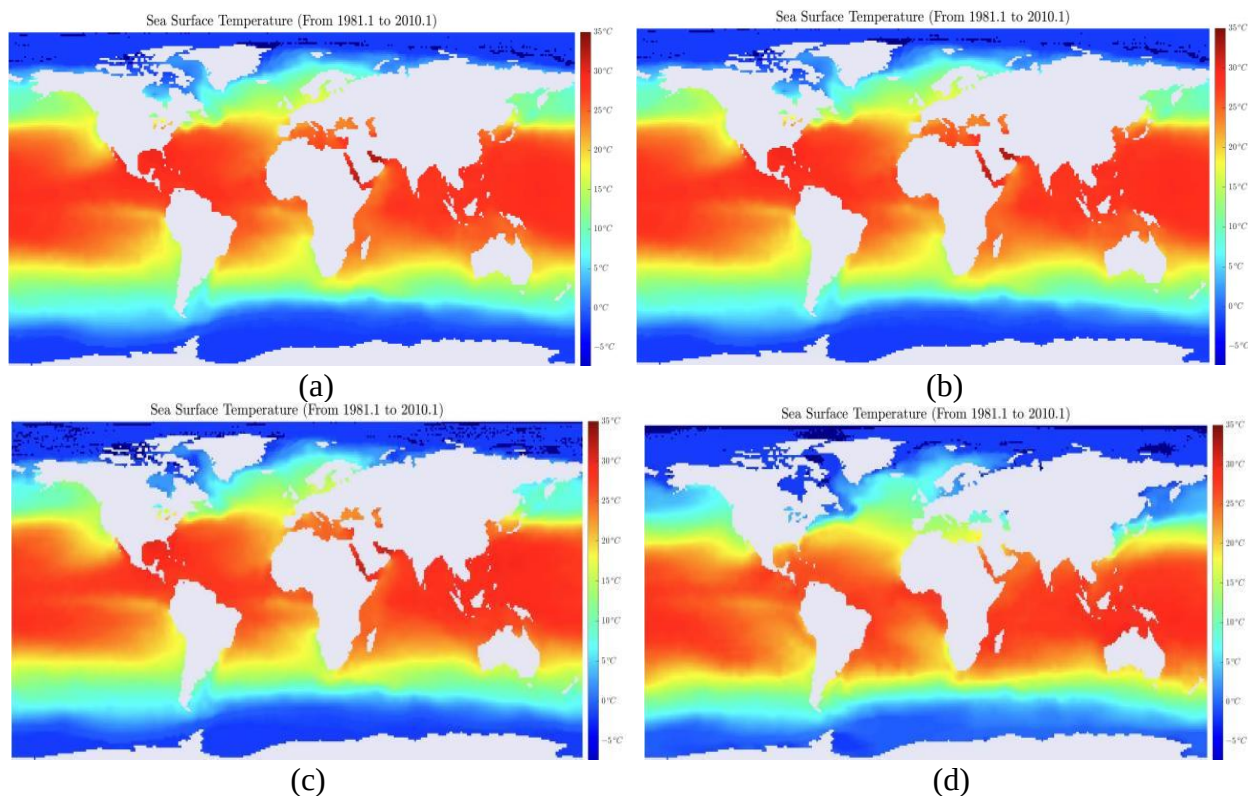


Figure 10. Trends in ocean surface temperature

(a) Map of ocean surface temperature for January 1981, (b) Map of ocean surface temperature for January 1991, (c) Map of ocean surface temperature for January 2001, and (d) Map of ocean surface temperature for December 2010

4. Conclusion

In this paper, the trend of temperature change and SST in British Columbia are analyzed comprehensively, and the corresponding prediction and decision model is established. The trend of temperature rise from 1950 to 1999 and the gradual trend of global SST rise are obtained. Therefore, the prediction decision model based on temperature data can predict climate change effectively and accurately.

In the actual climate prediction, we must fully take into account the influence of multiple factors on climate change. In addition to temperature changes, factors such as rainfall, wind direction, and humidity also have important effects on climate patterns. Therefore, we need to continuously optimize climate models to more accurately simulate and predict future trends in climate change. This will provide policy makers with more reliable strategies to address climate change, help them more effectively address the challenges posed by climate change, and promote the realization of global sustainable development.

References

- [1] Hu Jie, Zhang Jianzhi. The United Nations response to global disaster risk management and its implications [J]. Cities and Disaster Reduction, 2022, (06): 1 - 4.
- [2] Liu Jianguo. The United Nations says that the earth has entered the "boiling era" [J]. Ecological Economy, 2019, 39 (10): 1 - 4.
- [3] Liu Zhaohui. Is extreme weather sweeping the world, or reshaping global weather? [J]. Xinmin Weekly, 2023, (31): 48 - 53.
- [4] Zhang Piaoyin, Lu Jianzhong, Chen Xiaoling. Prediction of global terrestrial extreme climate change based on CMIP6 ensemble optimization dataset [J]. Journal of central China normal university (natural science edition), 2023, 57 (01): 77 - 88.
- [5] Liu Anan, Li Tianbao, Song Dan et al. Research progress of prediction methods for Marine extreme weather phenomena [J]. Data Acquisition and Processing, 2023, 38 (02): 231 - 244. (in Chinese).
- [6] Guo Jia, Zhai Guofang, Ge Yifu. Construction and Empirical Study of the Comprehensive Resilience Index for Rural Human Settlements: A Case Study of Jiangsu Province [J]. Shanghai Urban Planning, 2023, (02): 15 - 22.
- [7] Wu Chao. Exploration and Innovation of China's Comprehensive Strengthening of Social Construction in the Decade of the New Era [J]. Contemporary Chinese History Research, 2023, 30 (02): 9 - 26+156.
- [8] Tan Qiuxia, Wu Zilin. Research on the Protection of the Rights and Interests of Delivery Riders from the Perspective of Interest Balance [J]. Journal of Fujian Jiangxia University, 2023, 13 (01): 47 - 58.
- [9] Hou Xuhua, Tang Yuhui. Research on the Path to Achieve the "Dual Carbon" Goal with Forest Insurance: A Case Study of Hunan Province [J]. Hunan Social Sciences, 2022, (06): 64 - 74.
- [10] Zhang Xi. Research on Planning Elements of Climate Adaptable Cities [D]. Beijing Jianshu University, 2020.