# Comparative Model Study of Global Temperature Forecasting Based on the ARIMA-OLS Model

Kerui Liu, Boyu Zheng, Xinning Ding *

Department of Beijing Foreign Studies University, Beijing, China, 100089

* Corresponding Author Email: dxn320982@yeah.net

**Abstract.** Global warming is the most important climate issue for mankind today. Predicting the direction of future global temperature change is particularly important to develop relevant response strategies. To reasonably predict the future pace of global warming, this paper firstly collects global monthly average temperature data from Berkeley Earth, covering major land areas around the world and monthly sea surface temperatures from 2012 to 2022, to ensure wide spatial coverage and high data integrity. Based on the data's cyclical and stable growth characteristics, a time series model and a simple regression model are established. Then, temperature forecasts for 2050 and 2100 are generated using the ARIMA model and simple regression model separately. Results reveal that both models predict months when the average temperature exceeds 20 degrees Celsius in 2050 and 2100. However, only the simple regression model predicts that the average annual temperature will surpass 20 degrees Celsius in 2100. Considering the trend of global warming, this paper supports the superior predictive capability of the simple regression model. The paper contributes to climate analysis, providing valuable insights for understanding future temperature trends and guiding further research in this field.

**Keywords:** ARIMA-OLS Model; Simple Linear Regression; Correlation Analysis; Time Series Analysis; Global Warming.

## 1. Introduction

Global warming has emerged as the most significant climate issue of global society. [1-2] It not only has severe negative impacts on the Earth's ecosystems, such as melting ice caps, increased disasters, and species extinction, [3] but also profoundly affects human socio-economic systems, including reduced crop yields, decreased labor productivity, human health, and domestic conflicts. [4]As such, accurately predicting the direction and magnitude of future global temperature changes is crucial for developing effective response strategies and mitigating the impacts of climate change.

As the impact of global warming on the environment becomes increasingly severe, an growing number of scholars both domestically and internationally have started to pay attention to the issue of global warming. Many scholars have delved into profound thinking and research on temperature changes and predictions, and they proposed methods for temperature prediction from different perspectives. Wei and Cao[5] conducted a study on the occurrence of abrupt changes in the historical temperature series of China, the Northern Hemisphere, and the global scale using the hypothesis testing of mean differences. The analysis revealed that China experienced a transition from warm to cold in the late 1940s to early 1950s. Similar abrupt changes were observed in the Northern Hemisphere and globally in the late 19th century and the 1920s. Their research provided a theoretical foundation for subsequent studies on temperature prediction in China. Building upon their work, subsequent scholars have focused on investigating temperature changes within the geographical scope of China, employing various methods such as mean testing,[6] linear trend analysis, moving average technique, and Mann-Kendall test.[7-9] Besides, other scholars have also predicted global temperatures, for example, Sévellec et al.[10] devised a novel method for predicting global surface temperatures using transfer operators, resulting in accurate and reliable probabilistic forecasts.

Some scholars have paid attention to the paleoclimates, which provide an important scientific basis for understanding global warming. As the saying goes, "The study of the past is to predict the future", more and more scientists are trying to find the future global temperature change trend from the law of paleoclimatic change. Tierney et al.[11] systematically summarized and evaluated the important role of paleoclimate research in the scientific response to future climate change, emphasizing its importance for modern climate simulation research and future climate change prediction. The importance of paleoclimate research can be reflected in many aspects. For example, the regional and seasonal climate change information provided by paleoclimate research records the long-term and continuous history of climate change, which greatly extends the instrumental measurement record of modern climate[12]. Additionally, paleoclimatic records help scholars to understand the relationship between the ice sheet change and ice cover size, shape and range[13], and help to reduce the uncertainty of future sea level rise predictions.

All those scholars mentioned have made remarkable contributions to temperature prediction and have proposed several constructive methodologies and viewpoints. However, due to the utilization of diverse data processing techniques and research methods, there exist notable disparities in the outcomes of temperature forecasts. Additionally, while scholars have employed various methods such as mean testing, linear trend analysis, moving average technique, and the Mann-Kendall test for temperature prediction, there may be limitations in the applicability and accuracy of these methods. Therefore, further exploration and development of innovative and reliable prediction methods are necessary. This paper aims to address these gaps by presenting a rigorous methodology and providing insightful conclusions that will enhance the reliability of future global warming predictions. The research contributes to a comprehensive understanding of climate change dynamics and supports evidence-based decision-making in effectively addressing this urgent global issue.

## 2. Model Preparation

To make the model more realistic, the following assumptions were made in this study.

- Global temperatures can be simply predicted based on data trends and periodic properties, unaffected by special circumstances such as volcanoes, pandemics, El Nino, etc.;
- Assume that the data collected in this paper are true and reliable, and can accurately reflect the basic laws of global climate change;
- Assume that the earth's ecosystem will remain stable by 2050, without any new factors affecting the earth's climate;
- Assume that there is no big breakthrough in human technology, the energy structure will remain the same as it is now the energy structure is in line with today.

To facilitate model construction and make the later parameters easier to comprehend, All the symbols used in this paper and their descriptions are shown as Table 1.

**Table 1.** Symbol Description

| Symbol | Symbol Meaning | Property |
|---|---|---|
| $Y_i$ | Observed time series of Temperature | Variable |
| $Y_t$ | Observed Temperature at time t | Variable |
| k | A constant related to cycle | Constant |
| β | Coefficient of variables Yi | Coefficient |
| p | The number of autoregressive items | - |
| q | The number of moving average items | - |
| Y | Year | |
| $k_i$ | Coefficient of the average monthly temperature changes with year | Coefficient |
| $b_i$ | Intercept of monthly average temperature regression | Intercept |

## 3. Data Collection and Preprocessing

To reasonably predict the pace of global warming in the future, this paper collects global monthly average temperature data from Berkeley Earth. The dataset covers major land areas around the world, ensuring a wide spatial coverage. In terms of the time dimension, the dataset includes monthly sea surface temperatures from 2012 to 2022, ensuring high data integrity.

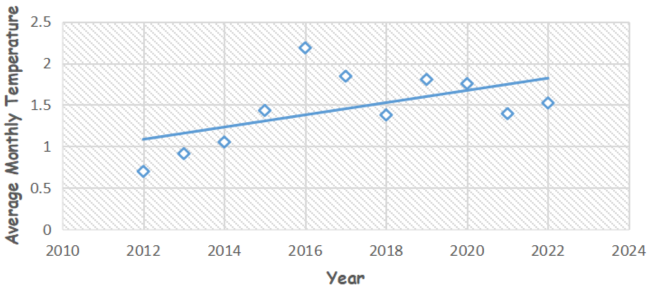



**Figure 1:** 2021-2022 global temperature (all land) average monthly temperature--March

Figure 1 shows the change in the global average monthly temperature from March 2012 to 2022 and the trend line. As can be seen from the figure, the global temperature is showing a rising trend.

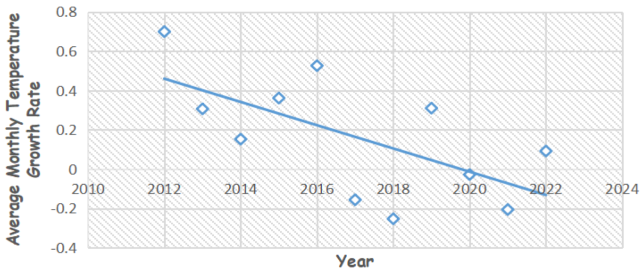



**Figure 2:** 2021-2022 global temperature (all land) average monthly temperature growth rate--March

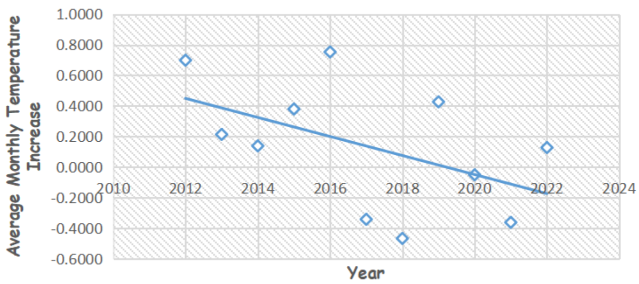



**Figure 3:** 2021-2022 global temperature (all land) average monthly temperature increase--March

Figures 2 and 3 show the growth rate and increase of the global average monthly temperature from 2012 to 2022. It can be seen that although the temperature from 2010 to 2022 is generally on the rise, the monthly average temperature growth and growth rate are declining. This is due to the base of calculating the temperature increase being large, so the temperature increase changes little, and the growth rate shows a downward trend. The growth rate in March 2022 did not exceed any March increase in the previous decade, let alone be larger than every period. Overall, global temperatures are rising, but growth rates are slowing down.

## 4. Construction of an ARIMA-OLS World Temperature Forecasting Mode

To predict future temperatures, the paper considers the periodicity and trend stability of the temperature ***data and*** consider time-series models and simple regression models.

### 4.1. ARIMA model construction and prediction

#### 4.1.1 The Establishment of ARIMA Model

To explain the past and forecast the future levels of global temperature based on historical data, the paper first establishes a time series model. By identifying the characteristics, trends, and development patterns of global temperature, effective forecasts for future global temperature levels can be made. The steps are as follows:

Step 1: Define the date as "Year, Month"

Step 2: Draw the time series diagram (some figures are shown below)

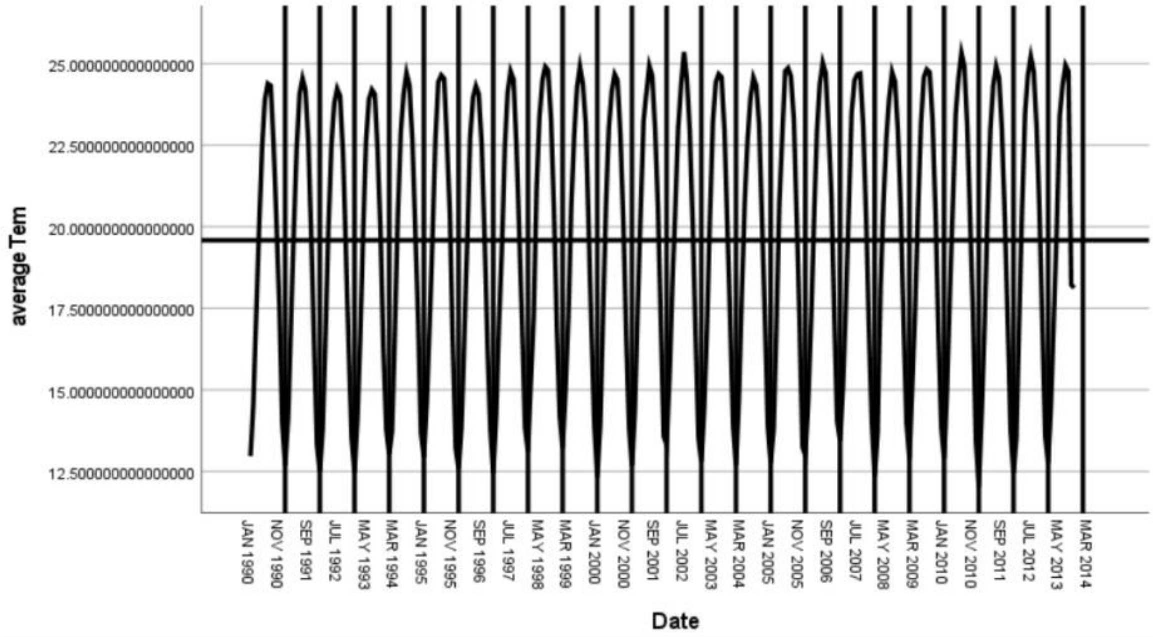



**Figure 4:** Time series diagram

As can be seen from Figure 4, the time series is relatively stable and does not require additional differential processing.

Step 3: Draw the autocorrelation and partial autocorrelation figures to further confirm the preliminary observations of the sequence diagram, test whether they are white noise sequences, and further modify the model.

Autoregression Integrated Moving Average Model (ARIMA) is one of the time series forecasting analysis methods. In ARIMA(p, d, q), p is the number of autoregressive items; q is the number of moving average items, and d is the number of differences made to make it a stationary series. Next, the paper will process data by numeric difference method and establish ARIMA model.

ARIMA (1,1,1) can be represented as:

$$\mathrm{ARIMA}(1,1,1):\quad Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + e_t - \theta_1 e_{t-1} \tag{1}$$

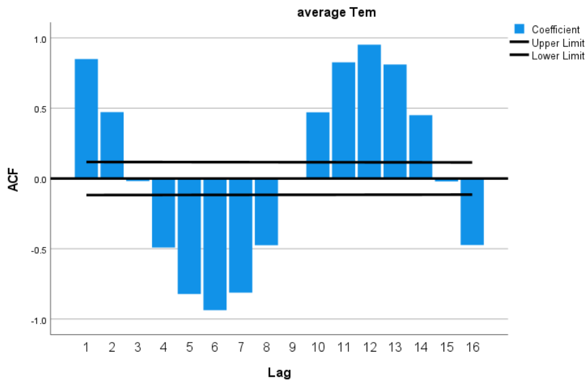
The results shown in Figure 5 and Figure 6:

For the observed time series $Y_1, Y_2, \ldots, Y_n$, the ACF is defined as:

$$ACF_k = \frac{\sum_{t=k+1}^{n}(Y_t - \overline{Y})(Y_{t-k} - Y)}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2} \tag{2}$$

Due to the fact the ACF dies off rather than cuts off in AR models, the paper uses PACF to determine the order of AR models, which is represented as:

$$\begin{gathered} PACF_k = Corr(Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \cdots - \beta_{k-1} Y_{t-k+1}, \\ Y_{t-k} - \beta_1 Y_{t-k+1} - \beta_2 Y_{t-k+2} - \cdots - \beta_{k-1} Y_{t-1}) \end{gathered} \tag{3}$$



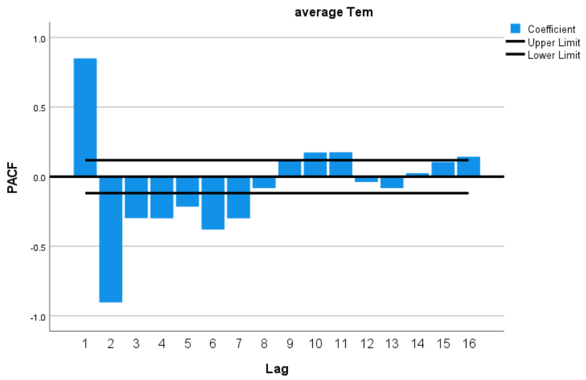




**Figure 5:** Autocorrelation figures **Figure 6:** Partial autocorrelation

Since both the ACF and PACF figures are in a trailing state, the ARIMA model is applied for prediction.

Step 4: Since the data period is too long, the paper decomposes the data seasonally to ensure the integrity of the periodic information. Use the Periodic Decomposition process to remove any systematic seasonal variations. A trend analysis is then performed on the seasonally adjusted series. Since the seasonal fluctuations of the data remain constant over time, the cyclical factors are modeled using additive. The comparison diagram is shown as below:

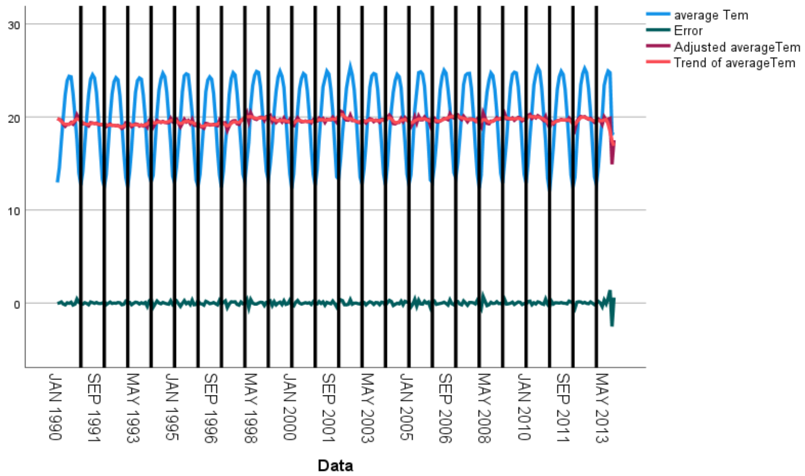



**Figure 7:** Comparison diagram

Step 5: Add the seasonal factors to plot the time series.

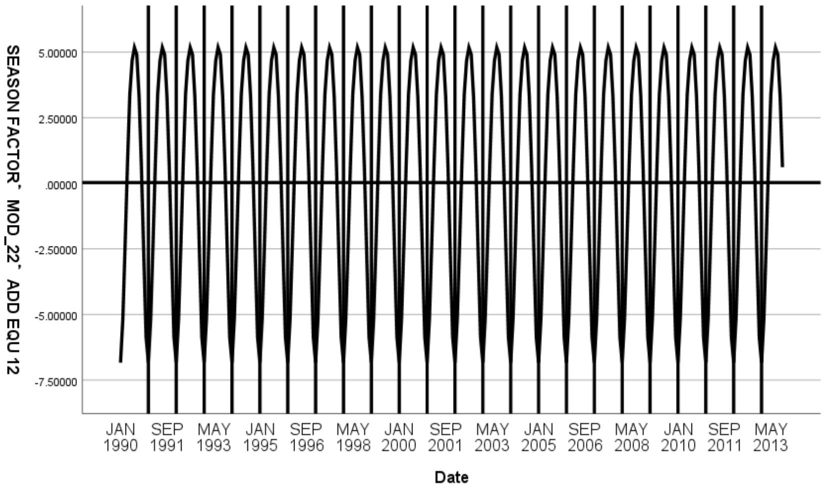



**Figure 8:** Time series after adding the seasonal factors.

The figure 8 shows periodic cycles of 12 months, with the temperature rising from January on.

### 4.1.2 The Solution of ARIMA Model

**ARIMA models** are built by spss to draw the residual ACF and the residual PACF figures, as shown in figure 9:

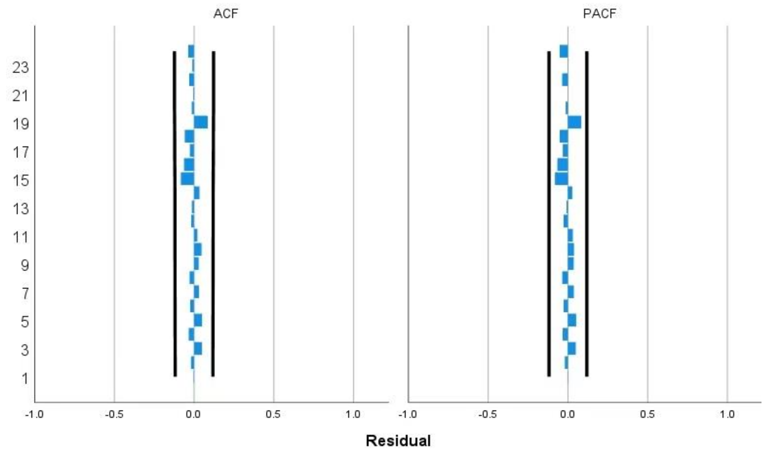



**Figure 9:** Residual ACF and residual PACF

Specific prediction results are shown in the Appendix.

Then the paper selects the predicted values from 2005 to 2100 and draw the figure 10:

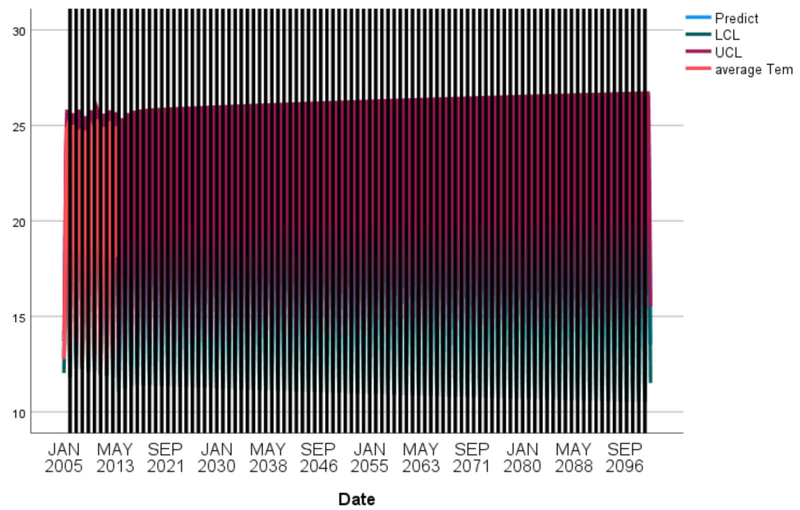



**Figure 10:** Predicted values from 2005 to 2100

### 4.1.3 The Robustness Test of ARIMA Model

The comparison figure of the predicted and measured values from 1990 to 2013 is drawn as follows:

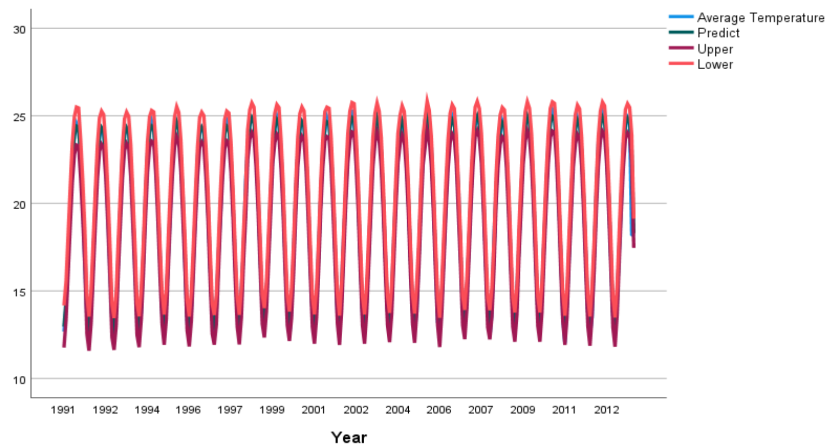



**Figure 11:** Comparison figure

As depicted in Figure 11, the measured values from 1990 to 2013 demonstrate a strong alignment with the predicted values, indicating a good fit. Moreover, the predicted value graph exhibits clear adherence to the periodic characteristics. These observations contribute to obtaining more realistic prediction results, thereby reinforcing the model's robustness.

### 4.1.4 Further Prediction Based on ARIMA Model

Based on the ARIMA Model, the paper calculates and sorts out the forecast results for 2050 and 2100.

The 2050-2100 prediction data in the time series model are drawn as follows:

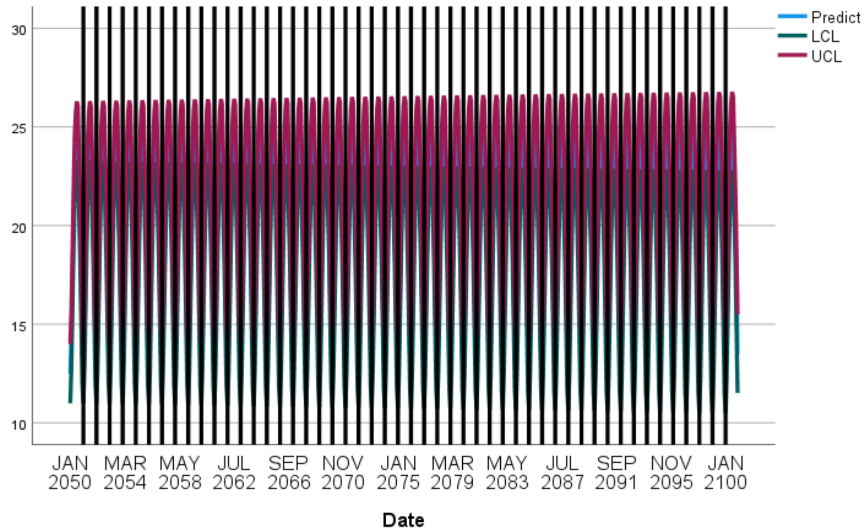



**Figure 12:** 2050-2100 prediction data

The specific prediction data for the years 2050 and 2100 are shown in table 2:

**Table 2:** Specific prediction data for the years 2050 and 2100

| Year | Month | Predict | Year | Month | Predict |
|---|---|---|---|---|---|
| 2050 | 1 | 12.4949 | 2100 | 1 | 12.4949 |
| 2050 | 2 | 14.2575 | 2100 | 2 | 14.2575 |
| 2050 | 3 | 17.2466 | 2100 | 3 | 17.2466 |
| 2050 | 4 | **20.2461** | 2100 | 4 | **20.2461** |
| 2050 | 5 | **22.9917** | 2100 | 5 | **22.9917** |
| 2050 | 6 | **24.2414** | 2100 | 6 | **24.2414** |
| 2050 | 7 | **24.7918** | 2100 | 7 | **24.7918** |
| 2050 | 8 | **24.5250** | 2100 | 8 | **24.5250** |
| 2050 | 9 | **22.2096** | 2100 | 9 | **22.2096** |
| 2050 | 10 | 19.9145 | 2100 | 10 | 19.9145 |
| 2050 | 11 | 16.7919 | 2100 | 11 | 16.7919 |
| 2050 | 12 | 13.5119 | 2100 | 12 | 13.5119 |
| Average | | 19.4352 | Average | | 19.4352 |

Based on the forecast results, the forecast value of future temperature from April 2050 to September 2050 and from April 2100 to September 2100 will reach 20.00 ℃.

## 4.2. OLS Model Construction and Prediction

### 4.2.1 Establishment of Simple Regression Model

The monthly average temperature changes from 1743 to 2012 are drawn as follows:

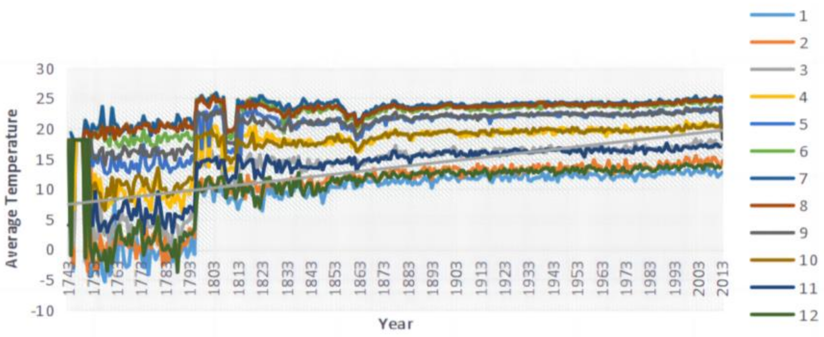



**Figure 13:** 1743-2013 monthly average temperature changes trends

As shown in figure 13, the global average monthly temperature shows a slow upward trend with the year. The figure shows that since 1833, the global monthly average temperature rise trend is stable, and presents a linear characteristic. Therefore, a simple linear regression model is considered for temperature prediction.

To make the trend clearer, the paper chooses the monthly average temperature from 1873 to 2013 to redraw the figure as follows:

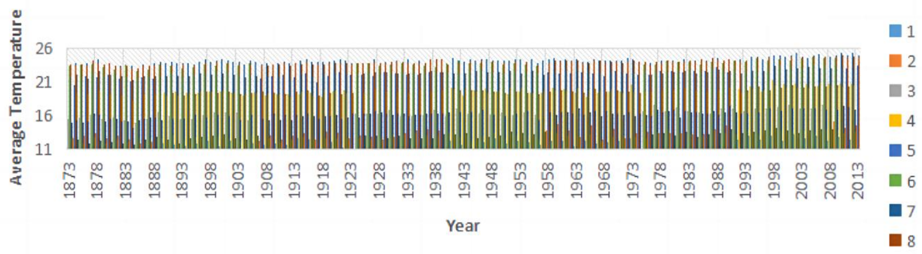



**Figure 14:** 1873-2013 Monthly average temperature changing trends.

Next, the paper establishes a simple regression model with monthly average temperature as the dependent variable and year as the dependent variable:

$$D_i = k_i Y + b_i \tag{4}$$

Taking January as an example to draw the regression prediction and the fitted figure:

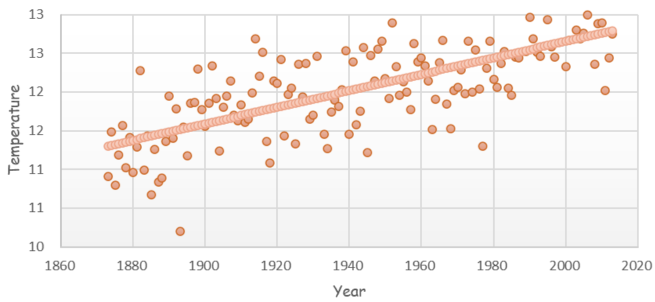



**Figure 15:** Fit plot for January

Residual distribution figure is shown in figure 16:

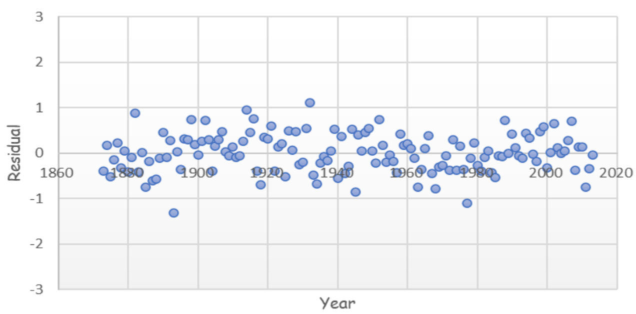



**Figure 16:** Residual Plot for January

### 4.2.2 Solutions of Simple Regression Model

The monthly average temperature data of 12 months are successively regressed, and the regression results are presented in table 3:

**Table 3:** Monthly average temperature data of 12 months

| | $k_i$ | $b_i$ |
|---|---|---|
| 1 | 0.01065 | -8.64043 |
| 2 | 0.01267 | -11.14780 |
| 3 | 0.01162 | -6.12857 |
| 4 | 0.01029 | -0.42708 |
| 5 | 0.00995 | 2.89832 |
| 6 | 0.00781 | 8.46858 |
| 7 | 0.00712 | 10.33923 |
| 8 | 0.00793 | 8.47735 |
| 9 | 0.00643 | 9.77642 |
| 10 | 0.00842 | 3.2385 |
| 11 | 0.01073 | -4.64346 |
| 12 | 0.01022 | -6.82008 |

According to the prediction results table, the regression coefficients are all integers, which represent the increasing trend of the air temperature changing with the year. Inputting historical data into equation $D_i = k_i Y + b_i$ can predict future temperature levels.

### 4.2.3 Further Prediction Based on Simple Regression Model

Based on the regression formula $D_i = k_i Y + b_i$ and the coefficient regression results, the prediction results for 2050 and 2100 are calculated in table 4:

**Table 4:** prediction results for 2050 and 2100

| Year | Month | Predict | Year | Month | Predict |
|---|---|---|---|---|---|
| 2050 | 1 | 13.1867 | 2100 | 1 | 13.7191 |
| 2050 | 2 | 14.8319 | 2100 | 2 | 15.4655 |
| 2050 | 3 | 17.6904 | 2100 | 3 | 18.2713 |
| 2050 | 4 | **20.6715** | 2100 | 4 | **21.1861** |
| 2050 | 5 | **23.3040** | 2100 | 5 | **23.8017** |
| 2050 | 6 | **24.4811** | 2100 | 6 | **24.8717** |
| 2050 | 7 | **24.9373** | 2100 | 7 | **25.2933** |
| 2050 | 8 | **24.7421** | 2100 | 8 | **25.1388** |
| 2050 | 9 | **22.9661** | 2100 | 9 | **23.2878** |
| 2050 | 10 | **20.5016** | 2100 | 10 | **20.9227** |
| 2050 | 11 | 17.3530 | 2100 | 11 | 17.8895 |
| 2050 | 12 | 14.1227 | 2100 | 12 | 14.6335 |
| Average | | 19.8990 | Average | | 20.3734 |

As can be seen from the forecast results, the temperatures from April 2020 to October 2050 and from April 2100 to October 2100 will exceed 20℃. Meanwhile, the simple regression model predicts that the average annual temperature in 2100 will exceed 20℃. To predict future temperatures, the simple regression model demonstrates superior accuracy compared to the ARIMA model considered. This can be attributed to two key factors: Independence of Monthly Temperatures: Unlike the time series model, the simple regression model treats temperatures between months as independent variables. This approach acknowledges that the influence of one month's temperature on another is not significant. By assuming independence, the model can focus on capturing the overall trend rather than being influenced by interdependencies. Amplifying Global Temperature Trend: The simple regression model effectively amplifies the rising trend of global temperatures through the use of regression coefficients. By incorporating these coefficients, the model strengthens the representation of the increasing temperature trend associated with global warming. As a result, the predictions generated by the simple regression model align more accurately with the observed patterns. In summary, the simple regression model surpasses other models due to its treatment of independent monthly temperatures and the amplification of the global temperature trend. These factors contribute to the model's improved accuracy in predicting future temperature changes within the context of global warming.

## 5. Model Evaluation

### 5.1. Advantages

The time-series prediction model fully considers the periodicity of the data, which enhances the accuracy and reliability of the predictions. By comparing the predicted values with the actual values, the model's robustness is demonstrated. The simple regression model employed in the paper leverages the stable trend of temperature rise during the same period. It amplifies the rising trend of data, thereby capturing important information that may have been ignored by the time series model. This approach improves the overall data analysis. The paper utilizes data visualization techniques effectively, presenting charts with different types and angles. This visual representation aids in making the arguments clear and enhances the understanding of the data patterns.

### 5.2. Disadvantages

The time series model in this paper fails to fully use the information on the slow rise of global temperature and fails to highlight the importance of this information in the periodic large temperature changes. The prediction models in this paper only make predictions based on the temperature data itself, without considering natural disasters, geographic locations, or oceanic special climates.

## 6. Conclusion

In conclusion, this paper strives to provide reliable predictions for the future trajectory of global warming. To achieve this objective, a meticulous process is followed, commencing with the comprehensive collection of global monthly average temperature data from Berkeley Earth. The dataset encompasses major land areas worldwide and includes monthly sea surface temperatures spanning from 2012 to 2022. To analyze and forecast future temperature patterns, two models are employed: a time series model and a simple regression model. These models are specifically tailored to accommodate the cyclical and stable growth characteristics observed in the collected data, enhancing their accuracy and effectiveness. Subsequently, the established time series and simple regression models are utilized to generate temperature forecasts for two pivotal timeframes: 2050 and 2100. By employing these distinct modeling approaches, the study allows for a thorough comparison and contrast of the predictions made by each model, enabling a comprehensive assessment of their respective performance. The findings of this analysis yield valuable insights. The predictions generated by the time series model and simple regression model contribute to our understanding of future temperature trends. The comparison between the models helps in selecting a more accurate prediction model.

## References

[1] Su B, Gao X J, Xiao C D. (2019). IPCC "Global Warming of 1.5℃" Cryosphere changes and the interpretation of their effects. Progress in climate change research (04), 395-404.

[2] Zhao Z C, Luo Y, Huang J B. (2018). Review IPCC 30 years (1988-2018). Climate Change Research (05), 540-546.

[3] Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences of the United States of America, 102(23), 8245–8250.

[4] Tamma A. Carleton & Solomon M. Hsiang. (2016). Social and economic impacts of climate. Science (6304).

[5] Wei, F. Y., & Cao, H. X. (1995). Analysis of temperature abrupt changes and trend prediction in China, the Northern Hemisphere, and globally. Atmospheric Science, (02).

[6] Sha, W. Y., Shao, X. M., & Huang, M. (2002). Climate warming in China since the 1980s and its impact on natural regional boundaries. Science in China (D Earth Sciences), (04), 317-326.

[7] Tian, H. F., Qin, Y. C., Li, G. D., Peng, J. F., & Liu, Y. R. (2013). Multi-timescale analysis of temperature changes in the eastern Henan farming area based on wavelet and M-K methods. Chinese Agricultural Science Bulletin, (35), 329-338.

[8] Lan, Y. C., Hu, X. L., Ding, H. W., La, C. F., Song, J., Zou, S. B., & Yin, Z. L. (2014). Multi-scale characteristics and abrupt change analysis of temperature in the mountainous areas of the inland river basin in Hexi region over the past 50 years. Journal of Mountain Science, (02), 163-170.

[9] Ren, J. Y., Peng, S. Z., Cao, Y., Huo, X. Y., & Chen, Y. M. (2018). Spatiotemporal distribution characteristics of climate change on the Loess Plateau during 1901-2014. Journal of Natural Resources, (04), 621-633.

[10] Sévellec, F., & Drijfhout, S. S. (2018). A novel probabilistic forecast system predicting.

[11] Tierney Jessica E, Poulsen Christopher J, Montañez Isabel P, Bhattacharya Tripti, Feng Ran, Ford Heather L... & Zhang Yi Ge. (2020). Past climates inform our future. Science (New York, N.Y.) (6517).

[12] Clara Deser, Adam Phillips, Vincent Bourdette & Haiyan Teng. (2012). Uncertainty in climate change projections: the role of internal variability. Climate Dynamics (3-4).

[13] Gulick Sean P S, Shevenell Amelia E, Montelli Aleksandr... & Blankenship Donald D. (2017). Initiation and long-term instability of the East Antarctic Ice Sheet.. Nature (7684).