World Scientific Research Journal Volume 11 Issue 5, 2025
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0017

Global Temperature Prediction Analysis Based on Random Forest and ARIMA Model

Zihan Yi[1, a]

[1]Xi'an International Studies University, Xi'an, Shaanxi, China

[a]13571648513@163.com

Abstract

This paper focuses on the study of global temperature change trends. By collecting historical temperature data from multiple sources, it conducts descriptive analysis and uses the Mann - Kendall mutation test to explore the temperature change characteristics. Two models, a random forest regression prediction model and an ARIMA - based time series model, are established to predict future global temperatures. The results show that the global temperature is on an upward trend, and the random forest regression prediction model has a higher fitting degree with the actual data. Additionally, the paper analyzes the spatiotemporal evolution of global temperature, the correlation between natural disasters and global temperature changes, and proposes measures to slow down global warming.

Keywords

Global temperature change; M - K mutation test; Random forest regression prediction model; ARIMA model; Correlation analysis.

1. INTRODUCTION

1.1. Background

In recent years, the global climate landscape has witnessed a disturbing surge in extreme weather events. Heatwaves roast continents, droughts parch vast regions, and floods inundate communities with increasing regularity. As this paper reveals, the Mann - Kendall mutation test indicates that global temperature mutations occurred in 2001 and 2005, with a marked upward trend since then. This makes understanding the temporal and spatial changes of global temperature crucial. By leveraging data from 1880 - 2022, and models like random forest regression and ARIMA as in this study, accurate future trend prediction becomes possible. Such predictions are the cornerstone for formulating climate - change adaptation strategies, safeguarding humanity's future.

1.2. Research Objectives

The main objectives of this study are to meticulously analyze the historical trend of global temperature. By comprehensively examining data from 1880 to 2022, we aim to uncover the underlying patterns of temperature variations over time. We will build advanced mathematical models, such as the random forest regression and ARIMA models, to accurately predict future temperature levels. Through spatial analysis using ArcGIS, we will explore the relationship between global temperature change, time, and location. Moreover, we will identify key influencing factors like greenhouse gases and solar radiation, and propose practical measures, including promoting clean energy and strengthening international cooperation, to effectively mitigate global warming.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0017

2. DATA COLLECTION AND PREPROCESSING

2.1. Data Sources

Global meteorological data spanning from 1880 to 2022 were sourced from the official website of the National Oceanic and Atmospheric Administration (NOAA). This dataset encompasses global average temperature values. Subsequently, these data were meticulously filtered and processed using Microsoft Excel software. The filtering process aimed to eliminate any potential errors or outliers in the raw data, ensuring data integrity. The processing in Excel involved tasks such as data cleaning, standardizing formats, and calculating relevant statistical metrics.

2.2. Missing Data Handling

Due to differences in data recording departments and cycles, missing values exist in the datasets. Two methods are used to handle these missing values: means imputation and interpolation based on series trend.

3. METHODOLOGY

3.1. Mann - Kendall Mutation Test

M-K mutation test method is a nonparametric statistical test method, that is, without assuming the distribution of random variables, the mean mutation start time and mutation region of a time series are found, which has been widely used in time series analysis and climate diagnosis of meteorological elements. The specific steps for model creation are as follows.

For a temperature time series containing samples, construct an ordinal sequence:

$$S_k = \sum_{i=1}^{k} p_{ij} \quad (k = 2,3, \dots, n)$$

Among the formula, we define:

$$p_{ij} = \begin{cases} 1, & t_i > t_j \\ 0, & t_i \le t_j \end{cases} \quad (j = 1,2, \dots, i)$$

And the defined statistics are:

$$UF_k = \frac{S_k - E(S_k)}{\sqrt{Var(S_k)}} \quad (k = 2,3, \dots, n)$$

Among them, $E(SK)$ is the mean of SK,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0017

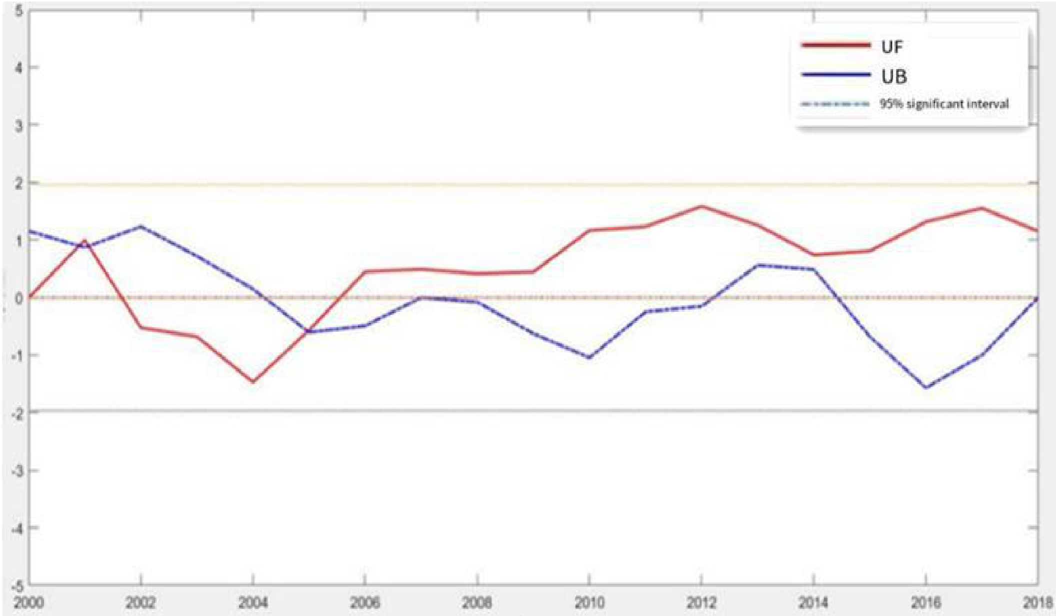

Figure 1. M-K mutation test of global average temperature in March

As shown in the figure above, UF and UB have 2 distinct intersections within the confidence interval, indicating that mutations occurred in these two years, and according to the location of the intersection point between UF and UB, the two years were determined to be 2001 and 2005. UF<0 between 2001 and 2005, indicating a downward trend in temperature; From 2001 to 2004, the average temperature characteristic curve UF fluctuated and decreased. After 2005, UF was always greater than 0, and the temperature gradually increased, and the warming trend became obvious; Volatility began to rise between 2005 and 2018, and according to the curve forecast, the global average temperature in March 2022 will continue to rise, and the increase will significantly exceed the temperature increase in the past 10 years.

Taken together, the rise in global temperatures in March 2022 was greater than observed in any previous decade.

3.2. Random Forest Regression Prediction Model

Random forest regression is a model that ensembles multiple CART trees for attribution prediction through the idea of ensemble learning. The specific modeling process is as follows.

For arbitrary division factor A, corresponding to any division point s, the parent node splits into two child nodes (hypothetical), the data sets of the two nodes are D1 and D2, and the factors and factor values corresponding to the sum of the mean squared deviations of D1 and D2 are minimized, and the factors and factor values corresponding to the sum of the mean squared deviations of D1 and D2 are used as the division points. The expression is:

$$\min_{A,s}\left[\min_{c_1}\sum_{x_i\in D_1(A,s)}\left(y_i-c_1\right)^2+\min_{c_2}\sum_{x_i\in D_2(A,s)}\left(y_i-c_2\right)^2\right]$$

In the formula, c1 represents the mean of the first node and c2 represents the mean of the second node. The above formula is used as the principle to determine the factor division point, and finally the temperature prediction regression model based on each factor is established.

Next, we tune the parameters of the model. We select parameters on the training set, and the result is shown in the following figure.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0017

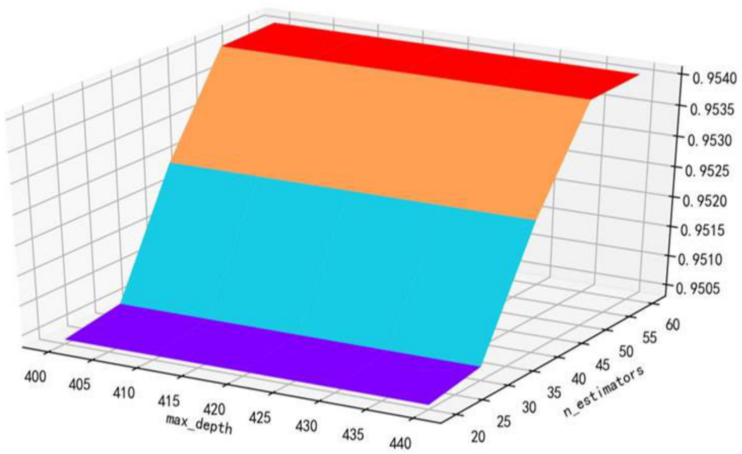

Figure 2. The best parameter selection under the random forest algorithm

As can be seen from the above figure, when the best parameter is selected as shown in the following table, the model converges and the accuracy converges to 0.954, achieving the best modeling effect.

Table 1. Model optimal parameters

parameter	optimal value
n_estimators	400
max_depth	60

Python is used to establish a random forest regression-based prediction model to describe the history data and predict the future data. And the fitting and prediction results of the model are shown in the figure below.

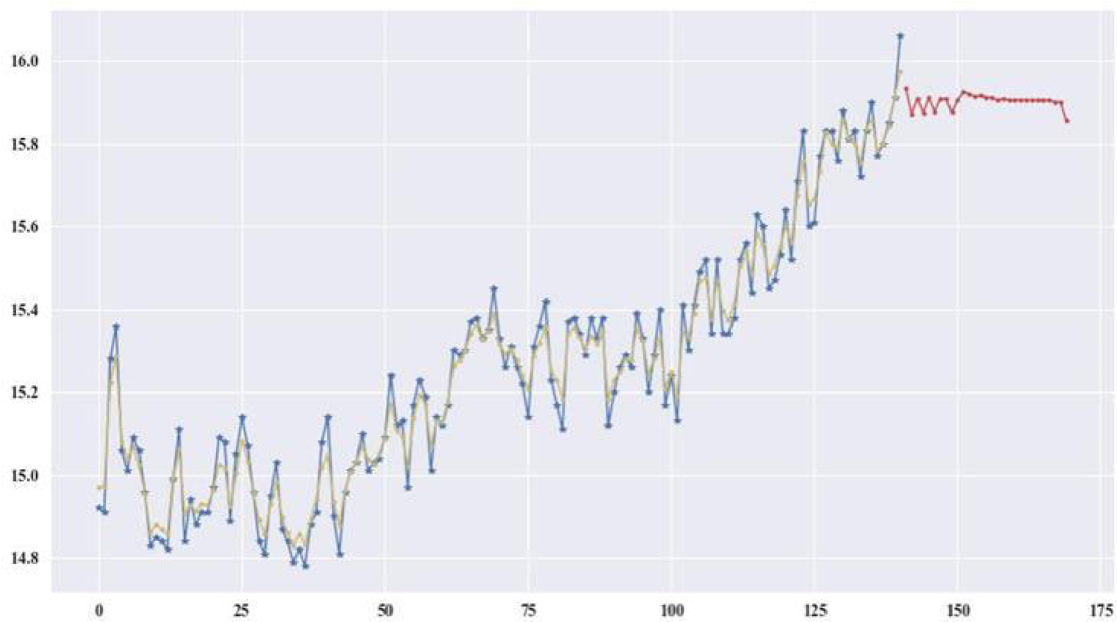

Figure 3. Fitting and prediction results of a random forest prediction model

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0017

Table 2. Results of random forest prediction model for future global temperature predictions

year	2020	2021	2022	2023	2024
Global Temperature Anomaly (C)	1.78	2.03	2.27	2.47	2.66
The average global temperature (C)	16.77	16.95	17.16	17.32	17.44
year	2025	2026	2027	2028	2029
Global Temperature Anomaly (C)	2.91	3.07	3.26	3.40	3.56
The average global temperature (C)	17.65	17.83	17.99	18.15	18.33
year	2030	2031	2035	2040	2050
Global Temperature Anomaly (C)	3.81	4.05	4.33	4.57	4.77
The average global temperature (C)	18.54	18.75	18.97	19.17	19.38
year	2060	2070	2080	2090	2100
Global Temperature Anomaly (C)	5.11	5.31	5.49	5.79	6.00
The average global temperature (C)	19.67	19.88	20.11	20.35	20.60

From the above table, the use of random forest regression prediction model predicts that the global average temperature will reach 19.38°C in 2050, 20.60°C in 2100, and 20.00°C in 2070-2080. This prediction results (i.e. reach 20.00°C between 2070 and 2080) is consistent with the requirement stated in the question that the global temperature will reach 20°C between2050 and 2100.

4. MEASURES TO SLOW DOWN GLOBAL WARMING

4.1. Strengthen International Cooperation

In the face of global warming, international cooperation is crucial. We should build a global community with a shared future and actively promote the goals of the Paris Agreement. Developed countries, with more resources and advanced technologies, have a greater responsibility. They should not only fulfill their own emission reduction commitments but also strengthen cooperation and assistance with developing countries. This includes providing financial support, sharing low - carbon technologies, and helping developing countries improve their energy - efficiency. Such efforts can enhance the overall global ability to combat climate change and benefit the global ecological environment.

4.2. Save Energy and Reduce Emissions

Saving energy and reducing emissions is a key step in curbing global warming. We need to actively develop clean and environmentally - friendly new energy sources like wind energy, solar energy, and tidal energy. These renewable energy sources can replace traditional fossil fuels, which are the main cause of greenhouse gas emissions. Additionally, improving energy efficiency in various industries can significantly reduce energy waste. By optimizing the industrial structure and phasing out high - energy - consuming industries, we can further cut down on carbon emissions and move towards a more sustainable development model.

4.3. Restore the Ecological Environment

Restoring the ecological environment is essential for mitigating the effects of global warming. Implementing measures such as returning farmland to forest and returning farmland to lakes can increase the area of green spaces and water bodies. Forests act as carbon sinks, absorbing carbon dioxide from the atmosphere, while water bodies help regulate the climate. Moreover, effectively managing industrial, agricultural, and animal husbandry pollution is necessary. This can restore the ecological balance, enhance the ecosystem's resilience, and contribute to a more stable climate.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0017

4.4. Improve the Adaptability of Poor Regions

Poor regions are often the most vulnerable to the impacts of global warming. Strengthening South - South cooperation can help improve the economic level and climate change response capacity of these areas. By sharing experiences and technologies, countries in the global South can jointly develop solutions. For example, promoting water - saving and efficient agriculture, and introducing renewable energy technologies. These efforts can not only enhance the energy - saving and emission - reduction capabilities of poor regions but also improve their overall adaptability to climate change.

5. CONCLUSION

This study conducts a comprehensive analysis of the global temperature change trend. By leveraging historical data from 1880 - 2022, we employed the Mann - Kendall mutation test to understand temperature change characteristics. Two models, the random forest regression prediction model and the ARIMA - based time series model, were established to forecast future temperatures. The results clearly show that the global temperature is on the rise, and human activities, such as greenhouse gas emissions, play a major role.

To mitigate global warming, measures like strengthening international cooperation, promoting energy conservation and emission reduction, restoring the ecological environment, and improving the adaptability of poor regions are proposed. For future research, enhancing the accuracy of prediction models and intensifying international cooperation are crucial. This will help us better address climate change and ensure the sustainable development of the global ecosystem.

REFERENCES

[1] SHU Zhangkang, LI Wenxin, ZHANG Jianyun, JIN Junliang, XUE Qing, WANG Yintang, WANG Guoqing. Historical changes and future trends of extreme precipitation and high temperature in China [J]. Strategic Study of CAE,2022,24(05):116 - 125.

[2] Reilai Kader, Ibraim Abduwup, Chen Gang. Research progress on global climate change and its influencing factors[J]. Agricultural Development & Equipment,2020(09):81 - 82.

[3] Wang Yuanhao. Global temperature prediction analysis based on ARIMA model and LSTM neural network [J]. Scientific and Technological Innovation,2021(35):166 - 170.

[4] XIANG Xudong, HU Xueping, YU Xiaomei, DENG Xin. Prediction of monthly mean temperature of Antarctic surface based on ARIMA model and LSTM model[J].Journal of Science of Teachers' College and University,2021,41(06):33 - 37.

[5] Kou Luyan, Liao Jing, Li Xuejun, Wu Changshu, Xiong Jianhua. Climate change prediction in Canada based on VAR model[J].Computer and Modernization,2022(10):13 - 18.

[6] HOU Huiqing. Global climate change prediction model based on BP neural network[J]. Science Technology and Innovation,2021(09):10 - 11.DOI:10.15913/j.cnki.kjycx.2021.09.005.

[7] XIAO Xuan, YANG Xinkai. Temporal and spatial variation trend of Canadian temperature based on ARIMA - wavelet analysis[J].Computer & Digital Engineering,2022,50(06):1183 - 1189.