ISSN: 2790-5209 Vol 2, No.4, 2024

Global Warming: Predictive Models and Correlation Analysis

Yichen Cai, Ziting Ding, Yixuan Xu

Millbrook school

15618517686@163.com

Abstract

In recent years, the Earth's rising temperatures and the shrinking Antarctic ice cover have become increasingly evident. Understanding the factors driving global warming and developing strategies to mitigate it are critical. By applying the Pettitt change-point detection method, we analyzed March temperatures from the past decade and identified a significant shift occurring in March 2015. Forecasting models, including ARIMA and LSTM, predict similar temperature trends, with the LSTM model demonstrating slightly higher accuracy. Through grey correlation analysis and multiple linear regression, we identified CO2 as the most influential factor affecting global temperatures. Therefore, reducing CO2 emissions is crucial to mitigating global warming.

Keywords

Global Warming, LSTM, ARIMA, The Gray Correlation Analysis, The Multiple Linear Regression Model, the Pearson correlation, Pettitt Mutation Detection.

1. Introduction

1.1. Background

Global warming, defined as the ongoing increase in Earth's average temperature, has emerged as a significant challenge for our generation. This phenomenon is primarily caused by the accumulation of greenhouse gases due to human activities like deforestation, burning of fossil fuels, and industrial processes. Over recent years, global warming has sparked various discussions and interpretations within the scientific community. With greenhouse gas emissions continuing to rise, researchers caution that the effects of global warming may intensify, leading to more frequent extreme heat events, as observed in recent years, and potentially even sharper temperature increases projected for 2022 and beyond. According to the IPCC and studies from organizations such as Berkeley Earth, global temperatures are projected to rise by around 1.5 to 2 degrees Celsius by 2050, and could increase by as much as 3.5 degrees Celsius by 2100.

The IPCC's Special Report on Global Warming of 1.5°C highlights the urgent need for comprehensive transformations across energy, land use, urban planning, and industrial sectors to limit global warming to 1.5°C above pre-industrial levels. Meeting this goal requires a significant cut in CO2 emissions—around 45% below 2010 levels by 2030—and achieving net zero by 2050. Predicting the trajectory of global warming is not only a scientific challenge but also vital for policy-making and future planning. Accurate climate modeling can contribute to better understanding of potential scenarios, effective policy development, informed decision-making, and increased public awareness.

1.2. Research objective

ISSN: 2790-5209 Vol 2, No.4, 2024

The main objectives of this paper are as follows:

To analyze global temperature trends from 2013 to 2022 and assess whether 2022 experienced a more significant rise in temperatures compared to previous years within this period.

To develop global temperature projection models to estimate the potential increase in average global temperatures by 2050 and 2100, assuming the absence of climate governance measures, and to predict when global temperatures might reach 20 degrees Celsius.

To investigate the key factors influencing global temperature fluctuations and explore the interrelationships among these factors.

1.3. Assumptions and Justifications

Considering that practical problems often involve numerous complex factors, it is necessary to make reasonable assumptions to simplify the model. Each assumption is accompanied by a corresponding explanation:

Assumption 1: The data used in this study are accurate and reliable.

Justification: The validity of the analysis hinges on the reliability of the data. Therefore, the results are meaningful only when the underlying data is trustworthy.

Assumption 2: Recent climate patterns will persist in the future.

Justification: Although advancements in technology and changes in government regulations may eventually lead to reduced greenhouse gas emissions, the time required for these initiatives to transition from research to full implementation and ecological impact means that no substantial shifts in climate patterns are expected in the near term.

Assumption 3: Data sources are credible and precise.

Justification: Given that the data is sourced from recognized international databases, we assume that it is credible. Based on this assumption, applying the data to build our model should yield objective and accurate results.

Additional assumptions are made to facilitate analysis in specific sections, and these will be discussed as needed in the relevant parts of the study.

1.4. Notation

Some important mathematical notations used in this paper are listed in Table 1.

Table 1: Notations Used in This Paper

Symbol	Meaning
$U_{t,n}$	Mutation point calculation
$sgn(.)$	Symbolic functions
K_t	The most likely mutation point in the sequence
ε	Random error
α	Fitting coefficient

β	Fitting coefficient
L	Extremely large likelihood function
e	Residual
p	ARIMA model order
q	ARIMA model order
MSE	Mean squared error
$SMAPE$	Symmetric mean absolute percentage error
dw	Calculated value of Durbin-Watson test

2. Data Acquisition

To ensure that evaluation data is both authoritative and consistent, we utilized data published by reputable international organizations and official statistical agencies, with each metric's source clearly labeled.

Table 2: Data Source Table

Database Names	Database Websites Data
NOAA	https://www.esrl.noaa.gov/gmd/ccgg/trends/
NASA GISTEMP	https://data.giss.nasa.gov/gistemp/
HadCRUT	https://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html
Berkeley Earth	http://berkeleyearth.org/data/
Google Scholar	https://scholar.google.com/

3. Analysis of Global Temperature Change

3.1. Pettitt mutation detection

The Pettitt test is a nonparametric method used to detect whether there is an unknown point of change within a time series and assess if this change is statistically significant. The steps for implementing the Pettitt mutation detection are as follows:

(1) Formulating the hypothesis test.

In Pettitt's test, the null hypothesis H_0 posits that there is no change point in the time series at time t, while the alternative hypothesis H_1 suggests that there is a change at that point.

(2) Identifying Potential Change Points.

For a time series X = ($x_1, x_2, \ldots, x_n$) with an assumed change at point x_t, the series is divided into two segments: $x_1, x_2, \ldots, x_t$ and $x_{t+1}, x_{t+2}, \ldots, x_n$. The statistic $U_{t,n}$ is calculated as follows:

$$U_{t,n} = U_{t-1,n} + \sum_{j=1}^{n} sgn(x_t - x_j) \quad t = 2, \ldots, n \tag{1}$$

The symbolic function sgn is defined as:

$$sgn(x_t - x_j) = \begin{cases} 1 & x_t - x_j > 0 \\ 0 & x_t - x_j = 0 \\ -1 & x_t - x_j < 0 \end{cases} \tag{2}$$

Then, the statistic Kt is used to locate the most probable change point.

$$K_t = \max_{1 \le t \le n} \left| U_{t,n} \right| \quad (3)$$

(3) Significance testing.

After identifying the optimal change point through Formula 3, its significance level Pt is computed with the following formula:

$$P_t = 2e^{\frac{-6K_t^2}{n^3+n^2}} \quad (4)$$

For a specified confidence level, if $Pt > \alpha$ we accept the null hypothesis, indicating no significant change at time t; Conversely, if $Pt < \alpha$, we reject the null hypothesis, signifying a significant change at t.

In this study, we applied the Pettitt test separately to four time series: Northern Hemisphere land mean temperatures, Southern Hemisphere land mean temperatures, global land mean temperature change, and global land-ocean mean temperature change, specifically for March over the last ten years. This allowed us to examine whether there was an accelerated increase in global temperatures in March 2022.

3.1.1. ARIMA Model Results

Using global average land temperature data from 1800 to 2021 as the independent variable, we applied the ARIMA model to make predictions, shown in Figure 11.

As depicted in Figure 11, the ARIMA model predictions closely match the actual data for known values, with the predicted trend aligning with the original data. To objectively evaluate the model's accuracy, we calculated the MSE and SMAPE, yielding values of 0.2901 and 2.6436%, respectively.

Finally, we trained the ARIMA model on global average land temperature data from 1800 to 2021 and used it to forecast future temperatures. The ARIMA model coefficients are presented in Table 1.

Table 1: The coefficients of the ARIMA model

Coefficient	α_0	α_1	β_1
Value	0.0063591	0.33684	-0.84058

Using this ARIMA model, we projected global temperatures for 2050 and 2100, with results shown in Figure 12. According to Figure 12, global temperatures are not expected to reach 20°C in 2050 or 2100. The ARIMA model's temperature projections for these years are approximately 10.05°C and 10.5°C, respectively.

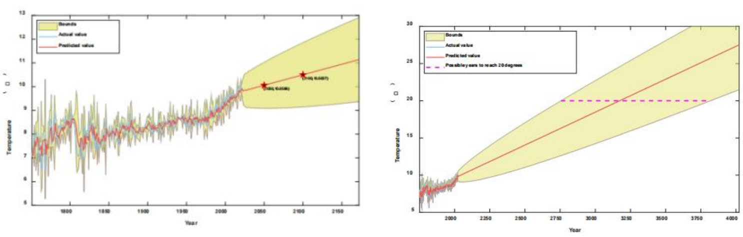


Figure 12: The results of ARIMA model prediction.

ISSN: 2790-5209 Vol 2, No.4, 2024

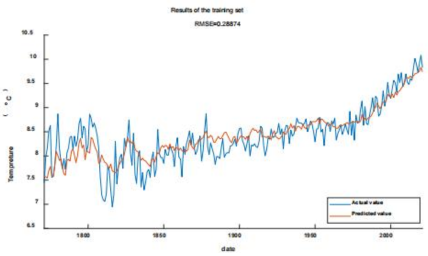
Figure 13: The results of ARIMA model prediction (2022-4021).

We extended the LSTM model to predict global temperature changes from 2022 to 4021, estimating the years when temperatures might reach 20°C. Given that model predictions can include some error, the actual year could vary within a range. Using a 95% confidence interval, this estimated range is shown in Figure 13.

Our projections indicate that global temperatures could reach 20°C by 3081, with a 95% confidence interval suggesting this might happen between the years 2756 and 3819.

3.1.2. LSTM Model Results

Using the previously trained LSTM model, we evaluated the entire dataset (1750–2021) to predict temperature changes, as seen in Figure 14. The model shows a good fit with the observed data, and the MSE and SMAPE between the actual and predicted values were calculated, resulting in an MSE of 0.2888 and a SMAPE of 2.6302%. These results indicate that the model is well-trained and can accurately predict future temperature trends.



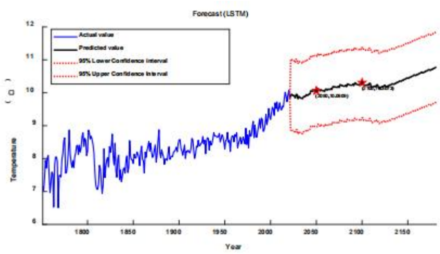


Figure 14: Results of prediction of the full set of data using the trained model.

Figure 15: The results of LSTM model prediction.

The LSTM model also generated temperature predictions for 2050 and 2100, shown in Figure 15. Like the ARIMA model, it suggests that global temperatures will remain below 20°C for both years. Specifically, the LSTM model projects temperatures of 10.06°C in 2050 and 10.31°C in 2100.

A further forecast using the LSTM model over the next 2100 years is displayed in Figure 16. In this figure, the purple line marks the range where the global average land temperature reaches 20°C. Given potential errors in model predictions, the estimated years are presented within a 95% confidence interval.

As seen in Figure 16, the global average land temperature is predicted to reach 20°C by the year 3094. Within the 95% confidence interval, this temperature level is expected to occur between 2988 and 3199.

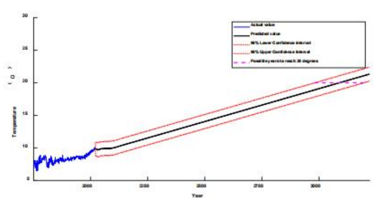


Figure 16: LSTM Model Long-Term Predictions for Global Average Temperature

3.1.3. Model Comparison

Key differences between the ARIMA and LSTM models are summarized in Table 2.

Table 2: Comparison Between ARIMA and LSTM Models

Models	ARIMA	LSTM
MSE	0.29006	0.28874
SMAPE	2.6436%	2.6302%
Forecast interval	[2756, 3819]	[2988, 3199]

(1) MSE and SMAPE Comparison

The ARIMA model's MSE and SMAPE values for the observed data are 0.2901 and 2.6436%, respectively, whereas the LSTM model yields smaller MSE and SMAPE values of 0.2888 and 2.6302%, suggesting that the LSTM model provides a slightly better fit.

(2) Model Forecast Comparison

According to Section 5.4.3, the ARIMA model predicts that global mean land temperature will reach 20°C by 3288, with a 95% confidence interval spanning the years 2756 to 3819. In contrast, the LSTM model estimates that this temperature threshold will be reached by 3094, with a 95% confidence interval of 2988 to 3199. Since the forecast intervals overlap in the range of 2988 to 3199, it is likely that global land temperatures will reach 20°C within this period.

4. Analysis of Global Temperature Impact Factors

4.1. Impact of Temporal and Spatial Factors

4.1.1. Data Overview

Since north-south latitudes and east-west longitudes correspond to distinct geographic areas, we assign negative values to data for western longitudes to differentiate these locations. Temperature increases were minimal prior to 1900, so we focus on data from January 1900 to August 2013 for this analysis. Temperatures from a sample of 15 cities are displayed in Figure 17.

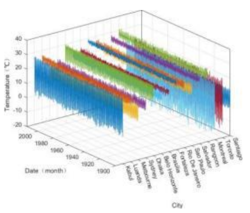


Figure 17: Temperatures by city.

In Figure 17, temperatures vary significantly across cities; for instance, Kabul has temperatures around 20°C with considerable fluctuations, while Fortaleza's temperatures are generally around 30°C with less variance. A gradual temperature increase over time is observed in each city, suggesting potential correlations with latitude, longitude, and time. We apply grey correlation analysis along with multiple linear regression to explore how these factors are related to temperature.

4.1.2. Grey correlation analysis

To assess the relationships between temperature, time, and location, we use grey correlation analysis to determine the correlation between temperature and the factors of time, latitude, and longitude. This process involves the following steps:

(1) Data Standardization

Since temperature, time, and location values are on different scales, standardizing the data is necessary to eliminate scale effects and normalize the range.

$$x_{ij} = \frac{x_{ij}^{'}}{mean\left(x_j^{'}\right)} \tag{14}$$

(2) Calculating grey correlation

The grey correlation coefficient formula is used to calculate the correlations:

$$\zeta_i(k) = \frac{min_i min_k |x_0(k)-x_i(k)| + \rho \cdot max_i max_k |x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)| + \rho \cdot max_i max_k |x_0(k)-x_i(k)|} \tag{15}$$

Here, ρ is the discrimination coefficient, typically ranging from 0 to 1, controlling the level of discrimination. A smaller ρ increases discrimination, while a larger ρ reduces it. It is often set to 0.5.

4.1.3. Multiple linear regression model

While grey correlation analysis identifies correlation strength between factors, it does not define the nature of their relationships. To address this, we build a multiple linear regression model, following these steps:

(1) Data Standardization

Like in grey correlation analysis, we start by standardizing the data.

(2) Formulating the Regression Equation

The multiple linear regression equation is expressed as follows:

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_{p-1} X_{1,p-1} + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_{p-1} X_{2,p-1} + \varepsilon_2 \\ \vdots \\ Y_3 = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_{p-1} X_{n,p-1} + \varepsilon_n \end{cases} \tag{16}$$

Here, ε represents random error, assumed to follow a normal distribution, $\varepsilon \sim N(0, \sigma$^2). The values of β are estimated using the least squares method, calculated as:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{17}$$

The coefficients (β) reflect the influence of each factor on temperature.

(3) Significance testing

To validate the multiple linear regression model, we calculated F-values with a significance level (P) for analysis.

4.1.4. Results

The grey correlation analysis results, as shown in Table 3, provide the correlation values between temperature, time, latitude, and longitude.

Table 3: Grey correlation analysis.

	Time	Latitude	Longitude
Temperature	0.7919	0.7514	0.6818

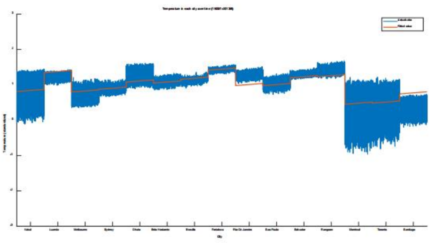
The results in Table 3 indicate a strong correlation of temperature with time and latitude, while the correlation with longitude is comparatively lower.

Using multiple linear regression, we also analyzed the relationship between these factors and derived regression coefficients and relevant statistical values, presented in Table 4.

Table 4: Multiple linear regression analysis.

Item	β_0	β_1	β_2	β_3	F	P
Value	1.5345	0.0308	-0.5996	0.0343	26648.5	0

As shown in Table 4, the model's F-value is 26648.5 with a p-value close to zero, indicating a good model fit. To illustrate the model fit and the effects of each factor on temperature, we plotted temperature against time, latitude, and longitude in Figures 18 and 19.



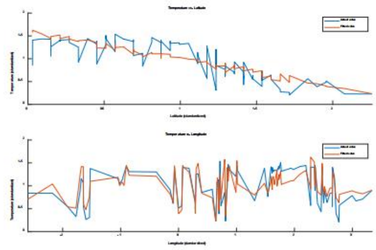


Figure 18: Multiple linear regression analysis (Temperature vs. Time).

Figure 19: Multiple linear regression analysis (Temperature vs. Latitude & Temperature vs. Latitude).

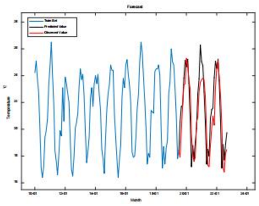
In Figure 18, we observe that the average temperature in each city increases over time, with both observed and predicted trends showing a similar upward pattern. In Figure 19, temperature decreases as latitude increases from lower to higher values.

4.2. Influence of Natural Disaster Factors on Temperature

Since the LSTM model outperforms the ARIMA model in both MSE and SMAPE metrics, the subsequent analyses utilize the LSTM model.

(1) Influence of Forest Fires on Temperature

By examining the discrepancy between forecasted and observed values, we can assess how Australian forest fires influence both local and global temperatures. As illustrated in Figure 20, the forecasted temperatures exceed the actual measurements, indicating that without the impact of forest fires, future temperatures would likely be higher. This suggests that forest fires can contribute to a reduction in temperature.



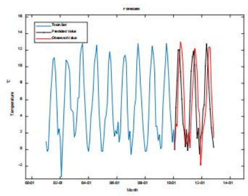


Figure 20: Forecasted local temperatures in Australia.

Figure 21: Forecasted temperatures post-Icelandic volcanic eruption.

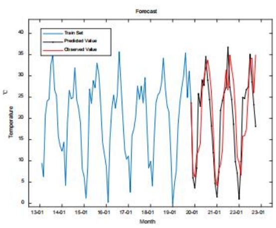
(2) Influence of Volcanic Eruptions on Temperature

In March 2010, an Icelandic volcano erupted. For analysis, we selected monthly temperature data from Iceland from January 2001 to December 2012, using data before March 2010 as the training set for the model. The trained model was then used to predict future temperatures and compare them with observed values, as shown in Figure 21.

The results reveal that the predicted temperatures are slightly higher than the observed data, suggesting that local temperatures in Iceland dropped following the volcanic eruption. Thus, volcanic eruptions appear to lower temperatures in the subsequent period.

(3) Impact of COVID-19 on Temperature

The global COVID-19 pandemic emerged in December 2019, triggered by the SARS-CoV-2 virus. To analyze its effects on temperature, we used monthly average temperatures from Wuhan between February 2013 and November 2022. Data prior to December 2019 was used as the training set, with comparisons between predicted and actual values shown in Figure 22.



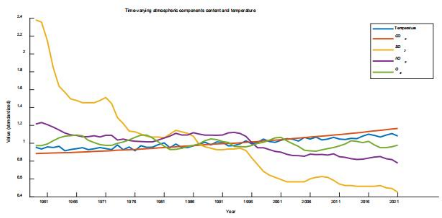


Figure 22: Predicted results for temperature after COVID-19 outbreak.

Figure 23: Time-varying atmospheric components content and temperature

Observing Figure 22, we see that the predicted temperatures are generally higher than actual recorded values, suggesting a temperature decline in Wuhan following the COVID-19 outbreak. This indicates a potential cooling effect associated with COVID-19.

4.3. Influence of Atmospheric Components on Temperature

We analyzed atmospheric composition and temperature data from 1959 to 2021, with the findings displayed in Figure 23. From this graph, we observe that trends in CO_2, O_3, and NO_2 closely align with temperature trends, while SO_2 trends show less similarity. To objectively assess each factor's association with temperature, we conducted grey correlation analysis, ranking the correlations by strength, as shown in Table 5.

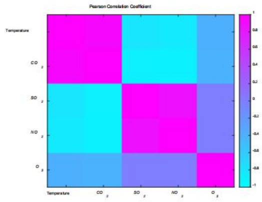
Table 5: Grey correlation analysis（Atmospheric composition and temperature）.

Rank	Component	Gray correlation value
1	CO_2	0.9592
2	O_3	0.9169
3	NO_2	0.8300

4	SO_2	0.6794

Table 5 shows that CO_2 and O_3 have the highest correlations with temperature, followed by NO_2, while SO_2 has the weakest correlation. This ranking is consistent with the observed trends in Figure 23.

Next, we calculated Pearson correlation coefficients to determine the nature of the correlations between atmospheric components and temperature, as well as inter-component relationships. These results are shown in Figure 24.



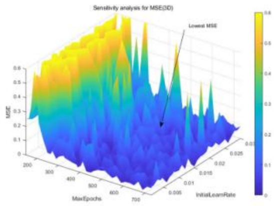


Figure 24: Pearson Correlation Coefficient
Figure 25: Sensitivity analysis for MSE(3D).

In Figure 24, CO_2 shows the highest correlation with temperature, while other components display weaker associations. From these findings, it appears that CO_2 is the primary driver of global temperature changes.

Conclusion

Our analysis identifies carbon dioxide as one of the most impactful factors on global warming, underscoring the need to prioritize CO_2 reduction strategies. To address this, advanced technologies should be developed and implemented to cut carbon emissions, including expanded use of renewable energy sources such as solar, wind, biomass, and hydropower. Additionally, developing innovative clean coal technologies, fuel cell advancements, nuclear energy options, advanced natural gas generation methods, unconventional energy applications, synthetic fuels, and carbon capture and storage techniques is essential.

Moreover, raising public awareness about environmental conservation is crucial so that individuals can contribute by adopting environmentally-friendly habits. Understanding climate change's effects should be encouraged across industries, business sectors, and everyday life. Practical steps such as choosing energy-efficient appliances, turning off electronics instead of using standby mode, and maximizing solar energy use can collectively make a substantial impact on energy savings and emissions reductions.

References

[1] Tedesco Marco,Keenan Jesse M.,Hultquist Carolynne. Measuring, mapping, and anticipating climate gentrification in Florida: Miami and Tampa case studies[J]. Cities,2022,131.

[2] Li Han,Grant Richard J.. Climate gentrification in Miami: A real climate change-minded investment practice?[J]. Cities,2022,131.

[3] Wang Liang. Exploring a knowledge map for urban resilience to climate change[J]. Cities,2022,131.

ISSN: 2790-5209 Vol 2, No.4, 2024
[4] Reu Junqueira Juliana,Serrao-Neumann Silvia,White Iain. Using green infrastructure as a social equity approach to reduce flood risks and address climate change impacts: A comparison of performance between cities and towns[J]. Cities,2022,131.

[5] Almeida Bernardo,Jacobs Carolien. Land expropriation – The hidden danger of climate change response in Mozambique[J]. Land Use Policy,2022,123.

[6] Athy Ariana E.,Milojev Petar,Gray Nathan Hoturoa,Osborne Danny,Sibley Chris G.,Milfont Taciano L.. Clarifying longitudinal relations between individuals' support for human rights and climate change beliefs[J]. Journal of Environmental Psychology,2022,84.

[7] Alfonso S, Gesto M, Sadoul B. Temperature increase and its effects on fish stress physiology in the context of global warming[J]. Journal of Fish Biology, 2021, 98(6): 1496-1508.

[8] Albouy C, Delattre V, Donati G, et al. Global vulnerability of marine mammals to global warming[J]. Scientific Reports, 2020, 10(1): 548.