Global Temperature Over Time: Analyzing the temperature time series data

Group Members

Sujal Mavdiya *202201040*
Parth Prajapati *202201250*
Harsh Mangukiya *202201363*

Professor Mukesh Tiwari

Abstract—This project focuses on analyzing climate change using time series analysis techniques to study historical climate data here we have global temperatures data. By applying concepts such as stationarity and other data analysis techniques, the goal is to understand trends and patterns in the data rather than perform forecasting. The data, sourced from online resources, provides insights into how climate variables have changed over time. This analysis highlights the value of time series methods in understanding temperature changes and supporting awareness of climate change progression.

1. Introduction

Global temperature is a critical indicator of climate variability. In this project, we analyze NASA's global temperature anomaly data, which includes the Source, Date, and Mean columns. The Date is in the YYYY-MM format, and the Mean represents the monthly mean temperature anomalies in degrees Celsius relative to a **Base period** (1951–1980 for GISTEMP or the 20th-century average for GCAG). A positive anomaly indicates a warmer-than-average month, while a negative anomaly indicates a cooler-than-average month.

Using this dataset, we apply time series analysis techniques to study both yearly and monthly temperature trends. Methods such as rolling averages, Mean Squared Error (MSE), and stationarity tests like Augmented Dickey-Fuller (ADF) is used to examine data stability. We also use regression models, polynomial fitting, and differencing to analyze and remove trends and seasonality. Autocorrelation analysis is included to further explore temporal patterns. Throughout the analysis, visualizations support the interpretation of how global temperatures have changed over time.

2. Temperature Data Analysis

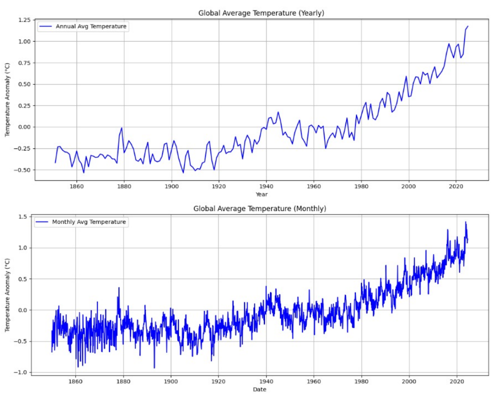


Figure 1. Annual and Monthly Global Average Temperature

The dataset considered in this study consists of global average temperature records collected from NASA's Global Climate Portal, covering the period from 1880 to 2024. The dataset provides both yearly and monthly temperature values, allowing a detailed analysis of long-term trends and seasonal fluctuations.

The above figure represents the yearly and monthly temperature variation plotted over time. It shows how the global temperature has changed in comparison to a baseline average, known as temperature anomalies.

Fig. 1 shows the Annual and Monthly Global Average Temperature variations between 1880 and 2024. The graph clearly highlights the steady and consistent increase in global temperatures over the last century, especially accelerating after the 1950s.

3. Analysis of Trend

Trend analysis is performed to study the long-term behavior and variation of temperature data. This section elaborates on three different methods used to identify the underlying patterns in the data.

3.1. Rolling Average Method (Visual method)

A rolling mean (moving average) technique was utilized to observe the underlying trend by calculating the average of temperature values within a specified window size.

- In this study,we have used a window size of 5 unit.
- This method makes the pattern more prominent
- The Mean Squared Error (MSE) value was observed to be 0.0062, indicating a very close fit to the original data with minimal error.

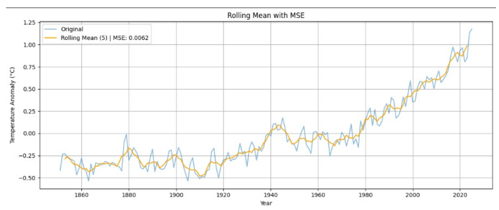


Figure 2. Rolling Average Trend of Temperature Data

3.2. Linear Trend Line

The linear trend line was fitted using the least squares method to determine the overall increasing or decreasing nature of the temperature data.

- The linear line indicated a slight upward trend over the observed period.
- This method provides a simple and direct estimation of the trend behavior.

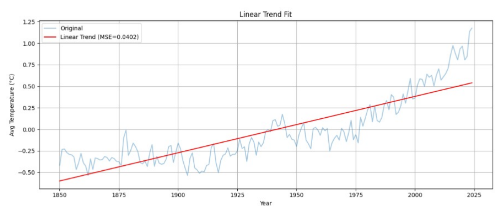


Figure 3. Linear Trend Line of Temperature Data

3.3. Polynomial Trend Line

A polynomial trend line was applied to capture non-linear patterns present in the data.

- A polynomial of degree 2 was selected for better flexibility in capturing variations.
- This method effectively captured the nonlinear rise in global temperature anomalies observed in the dataset.

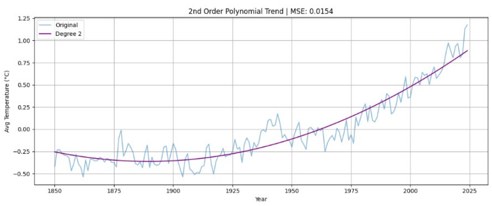


Figure 4. Polynomial Trend Line of Temperature Data

As per the above observation , the polynomial(polynomial of degree 2)trend fitting is little bit more accurate than the linear.As it is giving the efficient curve fitting with less MSE.

4. Detrending using Second Order Differencing

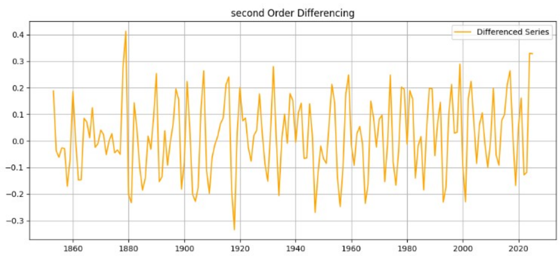


Figure 5. Detrending of Temperature Data using Second Order Differencing

4.1. Second Order Differencing Equation

The second-order differencing is a technique used to eliminate both linear and nonlinear trends from a time series. The general formula is given by:

$$y''_t = y_t - 2y_{t-1} + y_{t-2} \tag{1}$$

This transformation helps stabilize the mean of the series, making it more stationary and suitable for further analysis.

Above figure presents the second-order differenced global temperature anomaly series, aimed at eliminating both linear trends and long-term variations. This technique helps in making the series more stationary, meaning the statistical properties such as mean and variance remain stable over time.

The fluctuations are centered around zero, with no clear upward or downward trend. This suggests that second-order differencing effectively removes the underlying growth pattern present in the original temperature series.

To confirm the stationarity of the differenced data, the Augmented Dickey-Fuller (ADF) Test was performed. The obtained test statistic value is -4.3425 and the corresponding p-value is 0.0003745. Since the p-value is very small (less than 0.05), it confirms that the data is stationary after second order differencing.

5. Seasonal Decomposition of Average Temperature (Using Inbuilt functionality)

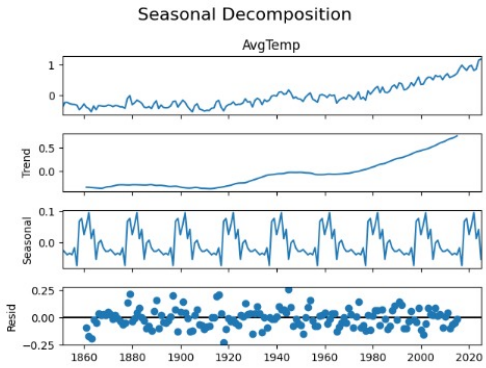


Figure 6. Seasonal decomposition of average global temperature time series

Above figure presents the seasonal decomposition of the global average temperature data. The decomposition is broken down into four components:

- **Observed (AvgTemp):** The original time series showing a clear increasing trend from 1850 to 2020, indicating global warming.
- **Trend:** A smoothed version of the original series that captures the long-term upward movement, especially accelerating post-1970.
- **Seasonal:** A repeating pattern that reflects seasonal temperature variations across years.
- **Residual (Resid):** The remaining noise after removing both trend and seasonal components. These appear randomly scattered, indicating effective decomposition.

6. Stationarity After Detrending and Deseasonalizing

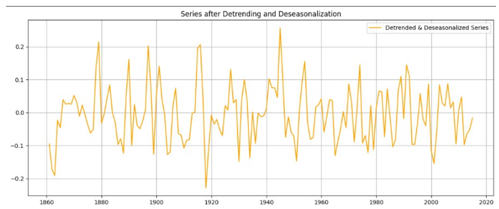


Figure 7. Time series after removing trend and seasonality, with ADF test results

After removing the trend and seasonal components from the time series, we tested the stationarity of the resulting residuals using the Augmented Dickey-Fuller (ADF) test. The results are:

- **ADF Statistic:** -7.3948
- **p-value:** 7.83×10^{-11}

Since the ADF statistic is significantly negative and the p-value is far below 0.05, we reject the null hypothesis that the series has a unit root. Therefore, the residual series is **stationary**, which is a crucial prerequisite for many time series Analysis.

Monthly Temperature Time Series Analysis (Using Inbuilt functionality)

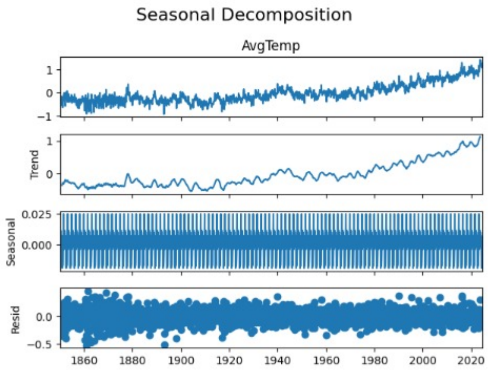


Figure 8. Seasonal decomposition of monthly average temperature

The monthly average temperature series was decomposed into trend, seasonal, and residual components. A clear seasonal pattern is visible with a clear trend. The residuals appear to be random fluctuations after decomposition.

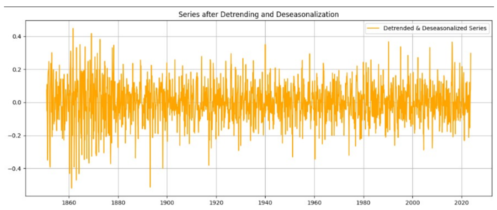


Figure 9. Monthly series after detrending and deseasonalizing, with ADF test results

After removing trend and seasonality, the Augmented Dickey-Fuller test yielded the following results:

- **ADF Statistic:** -14.112348
- **p-value:** 2.506×10^{-11}

These values indicate strong evidence against the presence of a unit root. Thus, the transformed series is stationary and suitable for further statistical analysis and interpretation.

7. Equation: Autocorrelation Function (ACF)

The autocorrelation function measures the linear relationship between lagged values of a time series. The formula for the sample autocorrelation at lag k is:

$$\rho_k = \frac{\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

Where:

- Y_t is the value of the series at time t
- $\bar{Y}$ is the mean of the series
- k is the lag
- T is the total number of observations

In this report, ACF plots are used to evaluate the presence of trend and seasonality. High values of ρ_k for large k suggest non-stationarity and potential seasonality. After transformations like differencing or detrending, we expect the autocorrelations to drop within the confidence bands, indicating a stationary series.

8. Autocorrelation Analysis of Monthly Average Temperature

Autocorrelation Function (ACF) plots are used to analyze the correlation of a time series with its past values (lags). The vertical blue bars (spikes) represent the autocorrelation at different lags, while the shaded blue area represents the 95% confidence interval. If a spike falls outside the shaded region, it is statistically significant, suggesting a meaningful correlation at that lag.

8.1. ACF Before and After Simple Differencing

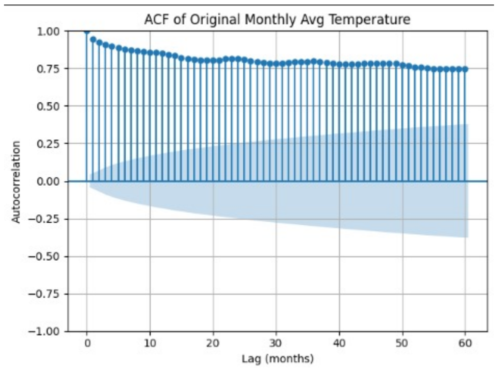


Figure 10. ACF of Original Monthly Average Temperature

In the ACF of the original monthly temperature series, we observe:

- High autocorrelation at all lags up to 60 months, slowly tapering off.
- Strong persistence and trend, which suggests the presence of non-stationarity.
- Seasonal cycles visible in a smooth, gradual decay of autocorrelations.

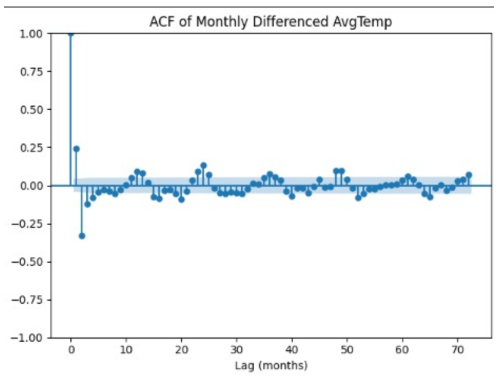


Figure 11. ACF of Monthly Detrended data series

After applying second-order differencing (only detrended data):

- The autocorrelation at lag 1 is still strong, but subsequent spikes rapidly fall within the confidence bounds.
- This indicates that differencing has successfully removed the polynomial (degree 2) trend.
- However, some minor structure remains, hinting at possible seasonal effects still present.

8.2. ACF After Detrending and Seasonal Differencing

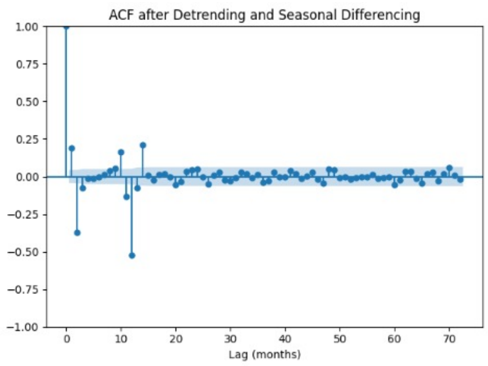


Figure 12. ACF After Detrending and Seasonal Differencing

Upon applying both trend and seasonal differencing:

- Most autocorrelation spikes lie within the blue confidence band.
- This suggests the residual series is now approximately white noise and thus stationary.
- The absence of significant spikes supports the success of preprocessing in removing temporal structure.

Conclusion: The original series is highly autocorrelated due to trend and seasonality. By differencing addresses the trend, and further seasonal differencing removes periodic patterns, resulting in a stationary series ready for modeling.

9. Final Detrended and Deseasonalized Series Analysis

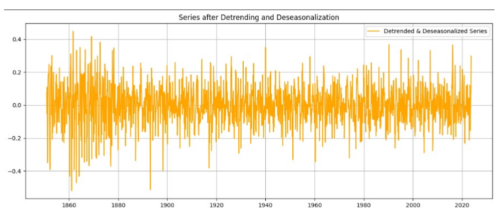


Figure 13. Final detrended and deseasonalized series

The above plot,the final series that is detrended and deseasonalized using the inbuilt functionality.

The fluctuations are centered around zero with no visible trend or strong seasonal patterns, indicating that the series is approximately stationary.

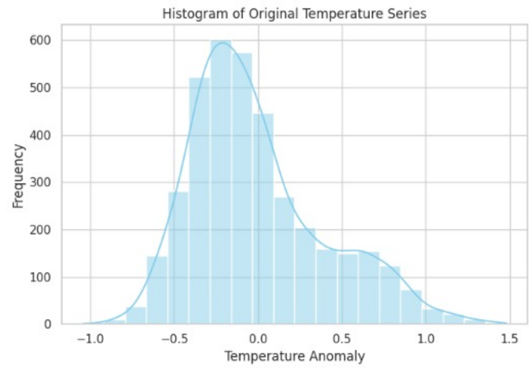
The Augmented Dickey-Fuller (ADF) test confirms stationarity:

- **ADF Statistic:** -14.17
- **p-value:** 1.99×10^{-26}

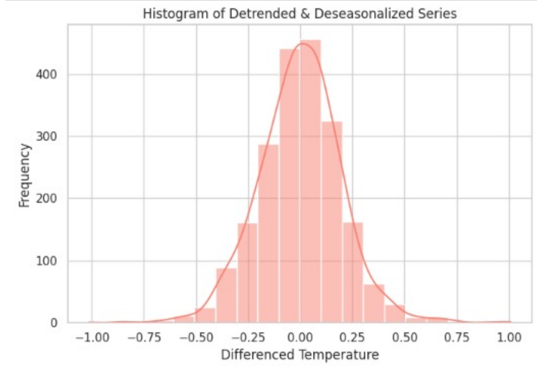
Since the p-value is significantly less than 0.05, we reject the null hypothesis of a unit root, confirming that the final series is stationary.

10. Histogram Analysis of Temperature Series

To evaluate the distributional characteristics of the temperature data before and after preprocessing, we examine two histograms: one for the original temperature anomalies and one for the detrended and deseasonalized series.



(a) Original Temperature Series



(b) Detrended & Deseasonalized Series

Figure 14. Comparison of Histogram Distributions Before and After Preprocessing

10.1. Histogram of Original Temperature Series

- The original temperature series exhibits a slightly **right-skewed** distribution.
- Values range from approximately **-1.0 to +1.5** degrees Celsius.
- The mode lies around **-0.2 to -0.1**, indicating the most frequently observed anomalies.
- The distribution has **heavier tails**, particularly on the right, indicating occasional extreme high temperature anomalies.
- The skewness suggests the presence of a **warming trend** and **seasonal effects**, both of which contribute to the observed non-normal distribution.

10.2. Histogram of Detrended and Deseasonalized Series

- After applying differencing and seasonal differencing, the distribution becomes nearly **symmetric and bell-shaped**, resembling a **normal distribution**.
- The range narrows to approximately **-0.75 to +1.0**, reflecting reduced variance.
- The mean of the differenced series is centered near **zero**, as expected from differencing transformations.
- The tails are thinner, with fewer outliers, indicating the removal of structured patterns such as seasonality and trend.
- This transformation leaves behind mostly **white noise**, making the series suitable for further time series analysis like model fitting.

The comparison between the original and processed histograms highlights the effectiveness of preprocessing steps. The original series displayed prominent temporal and structural patterns, indicating non-stationarity. In contrast, the processed series shows characteristics of a **stationary** time series such as randomness, constant variance, and a stable mean makes it more suitable for further statistical analysis.

References

[1] Class notes from Time Series Analysis course (SC475) DAIICT, 2025.

[2] Python Tutor, “Visualize code execution,” [Online]. Available: https://pythontutor.com/.

[3] Google Gemini, [Online]. Available: https://gemini.google.com/?hl=en-IN.

Google Colab Link

https://colab.research.google.com/drive/1iX2NbsVlBbkg7DyELiUB8NhygZq3raCR?usp=sharing#scrollTo=SOW6xXupplxS.