Time Series Analysis of Sunspot Activity

Swayam Hingu
202201207

Jenil Goswami
202201247

Abstract

This study presents a comprehensive time series analysis of monthly sunspot data to investigate the long-term behavior and periodic nature of solar activity. The analysis begins with an exploration of the data's distributional characteristics, including the assessment of skewness and kurtosis through histogram visualization, to evaluate deviations from normality. Key statistical methods are then employed to examine the stationarity of the series, identify autocorrelation patterns, and uncover underlying trends using smoothing techniques. The time series is further decomposed into trend, seasonal, and residual components to gain deeper insights into its structure. Additionally, cross-correlation analysis is conducted to explore potential relationships between solar activity and Earth's surface temperature, highlighting possible climatological connections. The results contribute to a deeper understanding of solar cycles and their temporal dynamics within the broader context of climate variability.

Introduction

Sunspots are temporary phenomena on the Sun's photosphere that appear as dark spots compared to surrounding regions. These areas of reduced surface temperature are caused by concentrations of magnetic field flux that inhibit convection. The systematic recording of sunspot observations dates back to 1749, making it one of the longest continuous scientific datasets available, providing invaluable insights into solar activity cycles and their potential impacts on Earth's systems.

Understanding these patterns through time series analysis enables deeper insight into the underlying mechanisms of solar dynamics. Moreover, as solar activity can modulate Earth's climate through changes in solar irradiance, it becomes increasingly important to examine possible connections between sunspot variability and global temperature trends.

In this project, we perform a detailed time series analysis on **monthly mean sunspot data,** aiming to uncover its statistical and temporal properties. The dataset spans multiple decades and captures substantial variations in solar activity. To characterize its distributional behavior, we analyze fundamental statistical metrics such as the **mean** (≈ 81.78), **variance** (≈ 4608.95), and **standard deviation** (≈ 67.89). The observed **positive skewness** (≈ 0.93) indicates a longer right tail in the distribution, suggesting more extreme sunspot counts during peak periods. The **kurtosis** value (≈ 0.34) further reveals that the data distribution is slightly more peaked than a normal distribution, indicating mild leptokurtosis.

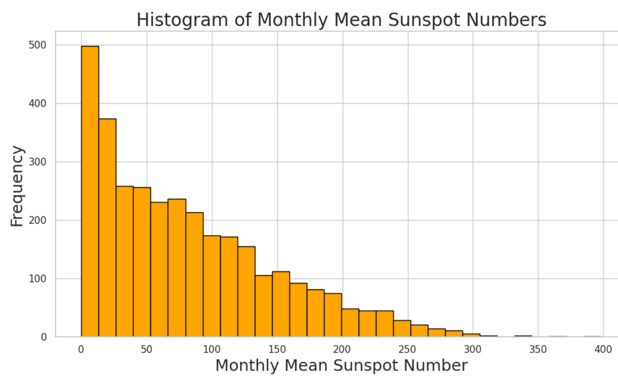

Fig. 1. Distribution of Monthly Sunspot Counts. This helps visualize skewness/kurtosis and support the deviation from normality.

To explore time-based patterns, we visualize the raw time series, highlighting both high-frequency fluctuations and long-term oscillations.

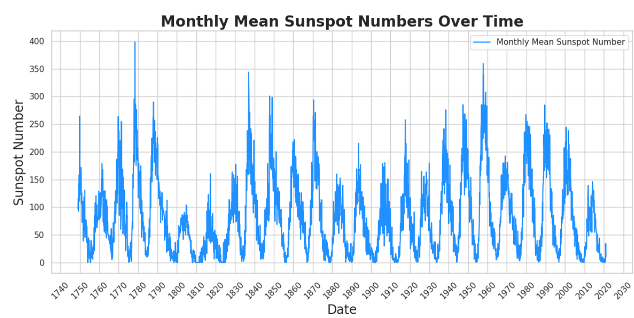

Fig. 2. Monthly Sunspot Data Over Time. This provides a visual introduction to cyclical trends.

We investigate the **stationarity** of the series and study autocorrelation structures with **ACF** plots, followed by **decomposing** the data into its **trend**, **seasonal**, and **residual** components. Finally, we assess the **relationship** between sunspot activity and **Earth's surface temperature** through **cross-correlation** analysis, identifying potential lags or lead-lag relationships between solar cycles and climatic shifts. This project aims to describe the statistical behavior of sunspot data while uncovering meaningful patterns that could inform future climate studies.

Stationarity Analysis

We analyzed the stationarity of the **monthly mean sunspot** time series through visual inspection and statistical testing, as stationarity is essential for accurate modeling and valid inferences.

A. Rolling Statistics on Original Series

To assess the stationarity of the sunspot time series, we computed the **rolling mean** and **rolling standard deviation** using a fixed-size moving window of **132 months** (11 years), which aligns with the average solar cycle length. // For a given time series $\{y_t\}$, the rolling (or moving) mean μ_t and rolling standard deviation σ_t over a window of size w are defined as:

$$\mu_t = \frac{1}{w}\sum_{i=0}^{w-1} y_{t-i} \qquad \sigma_t = \sqrt{\frac{1}{w-1}\sum_{i=0}^{w-1}(y_{t-i}-\mu_t)^2}$$

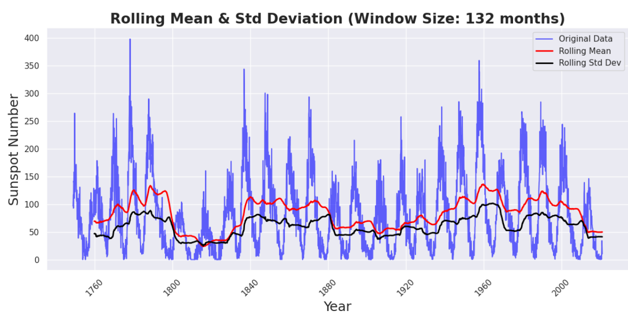

Fig. 3. **Original sunspot time series with rolling mean and rolling standard deviation.**
Rolling mean and standard deviation of the original series exhibit substantial fluctuation over time, indicating non-stationarity likely caused by trend and seasonality in solar activity.

B. Differencing and Transformed Series

To stabilize the mean and remove temporal dependencies, first-order differencing was applied:

$$y'_t = y_t - y_{t-1}$$

This transformation, commonly used to eliminate trend-induced non-stationarity, re-expresses the data in terms of changes between consecutive observations.

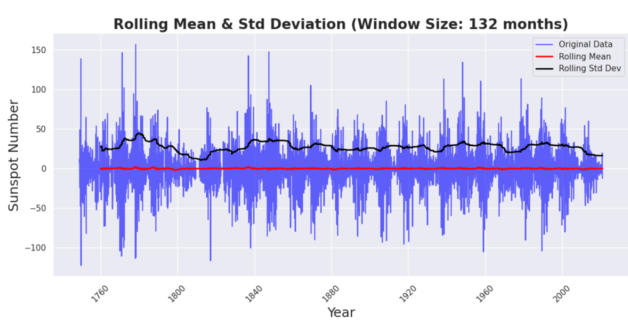

Fig. 4. **First-order differenced sunspot series with corresponding rolling mean and standard deviation**
Rolling statistics of the differenced series, showing a stabilized rolling mean and significantly reduced variability in the rolling standard deviation. These indicate that the series has become approximately stationary after differencing.

The initial time series exhibited non-stationary behavior as indicated by its fluctuating rolling statistics. Upon applying first-order differencing, both visual and statistical indicators aligned to support the assumption of stationarity. This transformation was essential to meet the foundational requirements of further time series modeling and analysis.

Autocorrelation Function (ACF) Analysis

In time series analysis, the Autocorrelation Function (ACF) serves as a critical diagnostic tool to quantify the degree of linear dependence between observations separated by a given lag. For a stationary stochastic process $\{y_t\}$, the autocorrelation at lag k is defined as:

$$\rho(k) = \frac{\mathbb{E}[(y_t-\mu)(y_{t-k}-\mu)]}{\sigma^2} \tag{1}$$

where μ and σ^2 denote the mean and variance of the process, respectively.
In practical applications, the empirical autocorrelation is computed using the sample autocovariance function normalized by the sample variance:

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^{T}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2} \tag{2}$$

where $\bar{y}$ is the sample mean and T is the number of observations.

A. Role of ACF in Time Series Analysis

The ACF provides insights into the temporal dependence structure of the data. It aids in:

- **Identifying persistence or memory**, which is essential for model selection and forecasting.
- **Detecting seasonality or cyclicity**, as periodic structures manifest through regular peaks in the ACF.
- **Assessing stationarity**, since autocorrelations of non-stationary processes often decay slowly or do not decay at all.

Understanding the ACF is particularly crucial when building models such as AR, MA, or ARIMA, where lagged correlations directly influence the model structure.

B. ACF Analysis in Sunspot Data

The behavior of the Autocorrelation Function (ACF) indicates two key properties of the sunspot time series:

- The **damped oscillation** in the ACF plot suggests that the correlation between observations decreases gradually as the lag increases. This pattern implies the presence of a **cyclical component** in the data, where values are influenced by their past values in a repeating, yet weakening, manner over time.
- The recurring peaks every 132 lags align with the well-established **11-year solar cycle** (132 months), further confirming the periodic nature of sunspot activity. The regularity of these peaks also suggests **strong seasonal components** in the series.

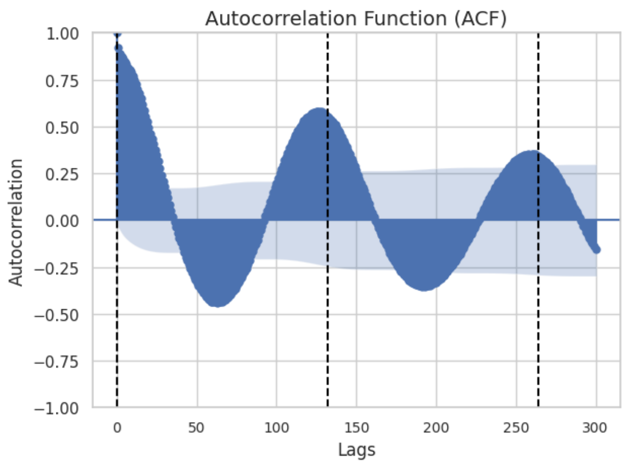

Fig. 5. **Autocorrelation Function (ACF) of the original sunspot time series** The observed damped sinusoidal pattern with regular peaks at 132-lag intervals reflects strong cyclic behavior consistent with the solar cycle.

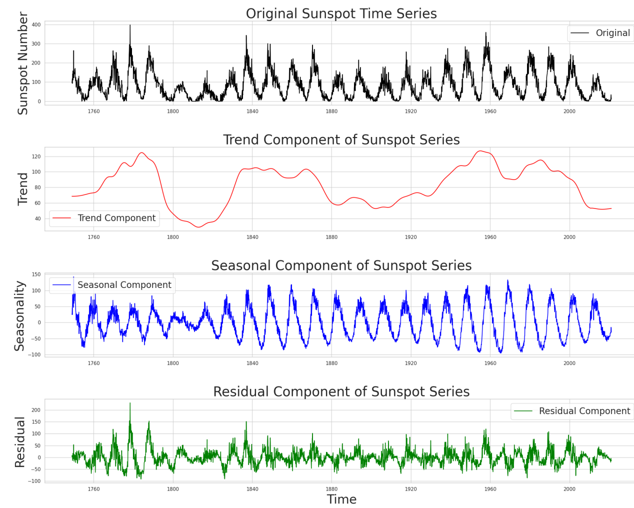

Fig. 6. **Seasonal decomposition of monthly sunspot numbers into observed, trend, seasonal, and residual components.**

Time Series Decomposition of Monthly Sunspot Data

To gain deeper insights into the structure of the sunspot time series, we performed a classical seasonal decomposition, separating the observed data into three fundamental components: trend, seasonal, and residual. This decomposition allows us to isolate long-term behavior, cyclic variations, and irregular fluctuations, each of which contributes uniquely to the overall time series pattern.

Given the nature of the data—monthly averaged sunspot counts—we used a seasonal period of 132 months, approximately corresponding to the average duration of the solar cycle (around 11 years). An additive decomposition model was employed, which is appropriate when the amplitude of seasonal effects remains relatively constant over time. Mathematically, the additive model is represented as:

$$Y_t = T_t + S_t + R_t$$

where Y_t is the observed sunspot count at time t, T_t is the long-term trend, S_t is the seasonal effect, and R_t is the residual or random component.

Trend Component

The trend component highlights the gradual changes in sunspot activity over the long term. It captures the **cyclical rise and fall** in solar activity, aligning closely with the known 11-year solar cycle. Peaks in the trend component match the sunspot maxima, while troughs align with solar minima. This confirms that sunspot activity follows a broad cyclic trend over decadal scales.

Seasonal Component

The seasonal component reveals the **repeating periodic structure** in the data. This part of the decomposition successfully captures the regular oscillations associated with the solar cycle. The stability in the pattern and amplitude across cycles suggests that sunspot numbers exhibit strong seasonality, reaffirming earlier observations from the autocorrelation function (ACF), where peaks were observed at lag intervals of approximately 132.

Residual Component

The residual component represents the **random or unexplained variation** after removing both trend and seasonality. The residuals appear to be centered around zero but show bursts of volatility. These irregular fluctuations may arise from short-term solar phenomena or measurement noise, and they are crucial when modeling the series with stochastic processes.

The decomposition of the sunspot time series clearly demonstrates that the data exhibits **a strong seasonal structure and long-term cyclical** trend, consistent with known solar physics. The decomposition not only validates prior conclusions drawn from rolling statistics and autocorrelations but also provides a cleaner view of the stochastic residuals, which are essential for building predictive models or conducting further statistical tests.

Cross-Correlation Analysis Between Sunspot Activity and Global Temperature

To examine the possible linkage between solar variability and Earth's climate, we perform a cross-correlation analysis between monthly sunspot numbers and global temperature anomalies. We also incorporate atmospheric CO_2 concentration and temperature uncertainty to broaden our understanding of intervariable dynamics. The objective is to investigate whether sunspot activity exhibits any statistically meaningful association—either contemporaneous or lagged—with global surface temperatures or related climate indicators.

A. Bivariate Scatter Plot Observation

The analysis begins with aligning and merging monthly sunspot counts and global temperature anomalies to explore potential relationships between solar activity and climate.

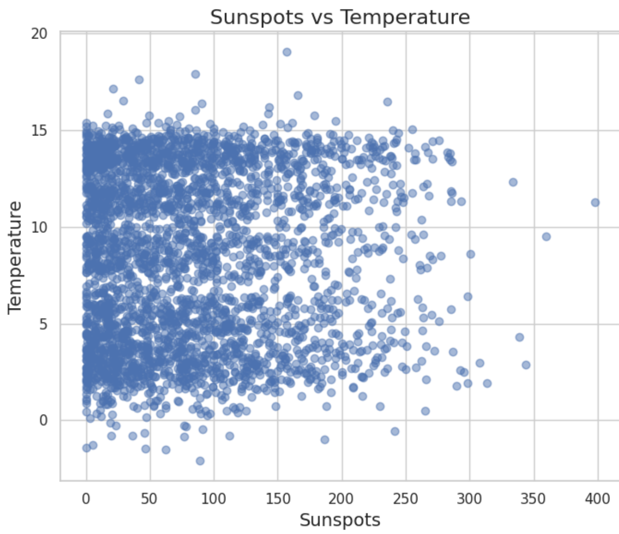

Fig. 7. **Scatter plot:Sunspot numbers vs AvgTemperature**
Scatter plot showing the relationship between sunspot number and global temperature anomaly. The plot reveals no discernible trend or clustering, suggesting no direct linear dependency

B. Correlation Matrix and Heatmap Analysis

To further evaluate potential dependencies, we computed the Pearson correlation coefficients among four variables: *sunspot number*, *global temperature anomaly*, *atmospheric CO_2 concentration*, and *temperature uncertainty*. The results are visualized using a correlation heatmap.

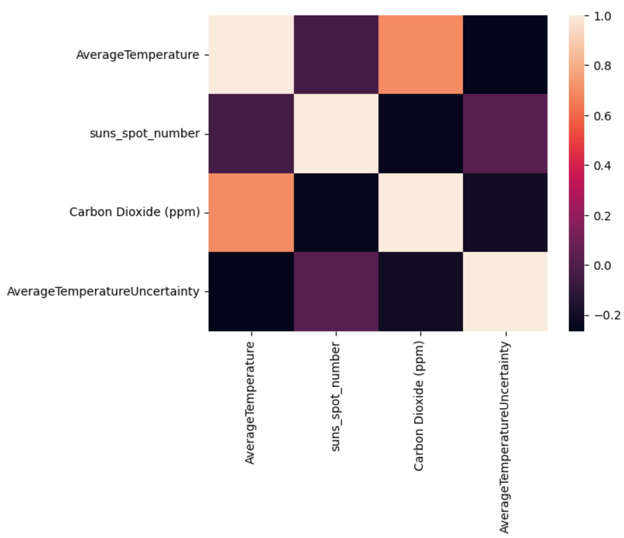

Fig. 8. **Heatmap of Pearson correlation coefficients between Sunspot Number, Global Temperature Anomaly, Atmospheric CO_2, and Average Temperature Uncertainty.**

The correlation between sunspot number and both temperature and temperature uncertainty is very weak, while a mild negative association is observed between sunspot number and atmospheric CO_2 concentration. In stark contrast, the correlation between CO_2 and temperature is moderate to strong. These results emphasize that solar activity, as quantified by sunspot numbers, does not appear to be a dominant driver of temperature anomalies or uncertainty in the observed period. Instead, greenhouse gas concentrations—particularly CO_2—exhibit a far stronger influence on global temperature patterns.

C. Cross-Correlation Function (CCF) Interpretation

To investigate whether the impact of sunspot activity on temperature anomalies might manifest after a temporal delay, we apply the Cross-Correlation Function (CCF). This method examines correlations at various time lags, helping to detect any delayed cause-effect dynamics.

Mathematically, for stationary time series x_t (sunspot numbers) and y_t (temperature anomalies), the cross-correlation at lag k is given by:

$$\rho_{xy}(k) = \frac{\mathrm{Cov}(x_t, y_{t+k})}{\sigma_x \sigma_y}$$

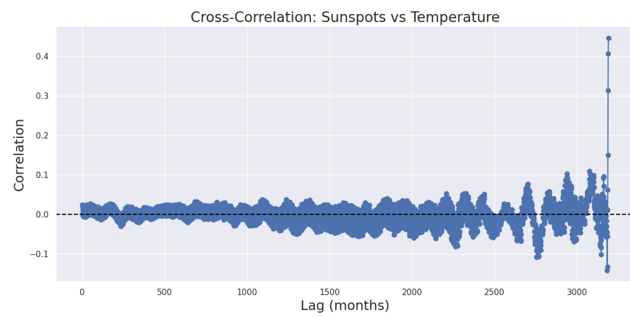

Fig. 9. **Cross-correlation function between sunspot numbers and temperature anomalies**
CCF plot shows values fluctuating around zero with no statistically significant peaks, indicating the absence of strong lagged correlations between the series.

Overall, the cross-correlation analysis provides compelling evidence that sunspot activity is not significantly correlated—either instantaneously or over time lags—with global surface temperature anomalies or their uncertainty. While it is scientifically plausible that solar activity influences Earth's climate over longer or more complex timescales (e.g., through multi-decadal solar cycles), our analysis shows that on a month-to-month basis, other factors such as atmospheric CO_2 exhibit far stronger and more consistent correlations with temperature variations.

Conclusion

This study explored the monthly sunspot data through time series analysis. The series was initially non-stationary, with clear trends and seasonality, which became stationary after first differencing, as confirmed by rolling statistics. A strong seasonal pattern of approximately 132 months was observed, consistent with the known solar cycle. While various statistical tools like ACF, decomposition, and correlation analysis were applied, no significant external relationship with temperature was found. Overall, the analysis highlights the periodic and stochastic nature of sunspot activity.