American University of Armenia (AUA)

Spring 2023, CSE DS207, Time Series Forecasting

Group Project: “Time Series Forecasting of Monthly Sunspot Data”

Students: Anahit Baghdasaryan, Lilit Galstyan, Mane Davtyan
Instructor: Aram Butavyan

Abstract:

This research project focuses on the analysis and forecasting of monthly sunspot data spanning from 1818 to 2019. The study aims to understand the sun's magnetic activity and its implications for Earth's climate and space weather. Time series forecasting techniques, specifically SARIMA and exponential smoothing models, are employed to predict monthly aggregated sunspot numbers based on historical observations. Additionally, a GARCH model is applied to analyze the forecasts of the SARIMA model. The research question addressed is the accuracy of forecasting monthly sunspot numbers using available dataset attributes and how the variability of sunspot numbers evolves over time. The findings of this study have implications for various stakeholders, including solar energy companies, climate scientists, space weather forecasters, and amateur astronomers, providing valuable insights into the behavior of the sun and its impacts on Earth and the space environment.

Introduction:

The study of sunspot data is essential in understanding the sun's magnetic activity and its effects on Earth's climate and space weather. Since the early 19th century, astronomers have recorded daily observations of sunspot activity, resulting in a vast collection of historical data. This dataset is now a significant area of research in solar physics, offering insights into long-term solar behavior and aiding in the development of forecasting models. Considerable progress has been made in the field of sunspot research, with numerous studies exploring the characteristics and behavior of sunspots. Existing literature has demonstrated the cyclical nature of sunspot activity, highlighting periodic fluctuations with an average period of around ten years, known as the solar cycle. This cyclicity, along with the sun's rotation and other factors, contributes to different dynamics of sunspot formation and evolution.

The primary motivation behind this research project is to leverage time series analysis and forecasting techniques to predict monthly aggregated sunspot numbers based on the available attributes in the daily sunspot dataset. By employing sophisticated statistical models, we aim to identify patterns and trends within the sunspot data, enabling accurate and reliable predictions of future activity. The outcomes of this research have broad implications for multiple stakeholders, including solar energy companies, climate scientists, space weather forecasters, and amateur astronomers.

In this research, our hypotheses revolve around the accuracy of forecasting monthly sunspot numbers using time series forecasting models. We hypothesize that employing SARIMA and exponential smoothing models will yield accurate predictions of monthly aggregated sunspot numbers. Additionally, we expect to observe changing patterns of sunspot variability over time, reflecting the evolving nature of the sun's magnetic activity.

The primary objectives of this research project are as follows:

- Apply SARIMA and exponential smoothing models to forecast monthly aggregated sunspot numbers based on daily observations.
- Assess the accuracy and reliability of the forecasting models by comparing the predicted values with actual data.
- Investigate the variability of sunspot numbers over time and analyze its changing patterns.
- Perform mean and volatility dynamics analysis by making a forecast of volatility using a GARCH model.

Literature Review:

Sunspot activity prediction has been the subject of extensive research in solar physics. Various techniques, including neural networks and time series analysis, have been explored to forecast sunspot activity. This literature review examines several notable studies that focus on predicting sunspot cycles using different modeling approaches. The objective is to gain insights into the effectiveness of these methods and their contributions to the field of solar physics.

V. Sandeep Kumar. (2018). "Sunspot Activity Prediction with Neural Networks" (Kaggle):
The work by V. Sandeep Kumar explores the use of neural networks for sunspot activity prediction. The author utilizes a dataset from Kaggle and employs a multilayer perceptron (MLP) neural network model to forecast the monthly average sunspot activity. The model achieved promising results, demonstrating the potential of neural networks in sunspot prediction. The work provides valuable insights into the application of neural networks in solar physics research.

M. Ali. (2018). "Sunspot Forecasting Using Time Series Analysis" (GitHub)[2]:
M. Ali's work focuses on sunspot forecasting using time series analysis. The author presents a comprehensive analysis of the sunspot dataset and applies various time

series techniques, including autoregressive integrated moving average (ARIMA) modeling and exponential smoothing methods. The study emphasizes the importance of data preprocessing and model selection in achieving accurate sunspot forecasts.

C. Y. Liu, S. Q. Cao, L. H. Lu, & X. Y. Zhang. (2013). "Forecasting Sunspot Cycles Using ARIMA Models" (Advances in Astronomy)[3]:
In this work, Liu et al. investigate the use of ARIMA models for sunspot cycle forecasting. The authors propose a methodology to predict the amplitude and timing of future sunspot cycles based on historical sunspot data. They demonstrate the effectiveness of ARIMA models in capturing the cyclic nature of sunspot activity and forecasting future cycles. The study contributes to the understanding of long-term sunspot behavior.

M. T. Ullah, M. Z. Uddin, & M. A. Matin. (2015). "A Hybrid ARIMA and Artificial Neural Network Model for Sunspot Forecasting" (International Journal of Scientific & Engineering Research)[4]:
Ullah et al. propose a hybrid model that combines ARIMA and artificial neural network (ANN) techniques for sunspot forecasting. The authors integrate the strengths of both approaches, utilizing ARIMA for short-term prediction and an ANN for long-term forecasting. The hybrid model demonstrates improved accuracy compared to individual ARIMA or ANN models. This study highlights the potential benefits of combining different forecasting methods for sunspot activity prediction.

In conclusion, the reviewed literature highlights the diverse approaches used in sunspot activity prediction. Neural networks, such as the multilayer perceptron model, show promise in capturing the complexity of sunspot data and forecasting monthly average activity. Time series analysis techniques, including ARIMA modeling and exponential smoothing methods, provide valuable insights into the cyclic nature of sunspot cycles and the importance of appropriate data preprocessing and model selection.

The studies underscore the significance of combining forecasting methods, as demonstrated in the hybrid ARIMA and artificial neural network model. This hybrid approach leverages the strengths of individual techniques to improve the accuracy of sunspot forecasts, enabling both short-term and long-term predictions.

Overall, these literature findings contribute to the field of solar physics by providing valuable insights and methodologies for predicting sunspot activity. Continued research and exploration of innovative modeling techniques hold the potential to advance our understanding of sunspot behavior further and improve the accuracy of sunspot forecasts.

Materials and Methods:

The primary material used in this research project is the daily sunspot dataset spanning from 1818 to 2019. The dataset was collected by the Royal Observatory of Belgium, making it a reliable and well-documented source of sunspot data. It consists of 73,583 recorded observations, with each one representing a single day's sunspot count, along with attributes such as date-related columns, daily sunspot count, and other relevant variables. The dataset provides a comprehensive historical record of sunspot activity and forms the foundation for our time series analysis and forecasting.

The subjects of this study are the daily sunspot observations recorded by astronomers over the course of two centuries. The dataset encompasses the entire solar cycle along with the daily sunspot numbers, allowing for a comprehensive examination of sunspot behavior and variability.

The research design employed in this study is a time series analysis and forecasting approach. We utilize statistical models, namely SARIMA (Seasonal Autoregressive Integrated Moving Average), exponential smoothing, and GARCH (Generalized Autoregressive Conditional Heteroscedasticity), after cleaning and preprocessing the data. These models are well-established and widely used in time series analysis to capture and forecast complex patterns in data.

The research procedure can be divided into the following steps:

1. Data preprocessing:
   The daily sunspot dataset was carefully curated and cleaned to ensure data quality and consistency. This involved handling missing values and transforming the data as necessary for analysis. Additionally, for this research particularly, a subset of 10 years of the dataset was captured due to the cyclical nature of the sunspots to be able to forecast for a short period of time and observe more details in the predictions.
2. Model selection and training:
   As mentioned previously, two primary models were employed for forecasting: SARIMA and exponential smoothing. These models were selected due to their effectiveness in capturing temporal dependencies and seasonality in time series data. The models were trained using historical daily sunspot observations and fine-tuned to achieve optimal performance.
3. Forecast and evaluation:

After the models were trained, they were used to forecast monthly aggregated sunspot numbers. The forecasted values were plotted along with the actual values of the data to evaluate the reliability of the forecasts visually and then were compared to each other based on their mean squared error values to determine which one performed better in this research.

4. Mean and volatility dynamics analysis:
   By making a forecast of volatility using a GARCH model, we projected the future volatility levels in the sunspot data. It helps stakeholders assess and prepare for potential fluctuations or changes in the level of volatility.

Overall, this methodology allowed us to analyze the historical sunspot data, apply forecasting models, and assess their accuracy, enabling us to gain a deeper understanding of sunspot behavior and provide valuable predictions for stakeholders in various fields.

Results:

Initially, both manual and automated SARIMA models were estimated and evaluated based on their RMSE (root mean squared error); the automated one had a lower RMSE of 3.79, and the manual one had 3.83. Therefore, the automated one performed better for this data than the manual one and was chosen as the best SARIMA model.

In Figure 1 the original data is plotted along with the forecast of the best SARIMA model.

In Figure 2 the zoomed-in version of Figure 1 is presented (one-year period).

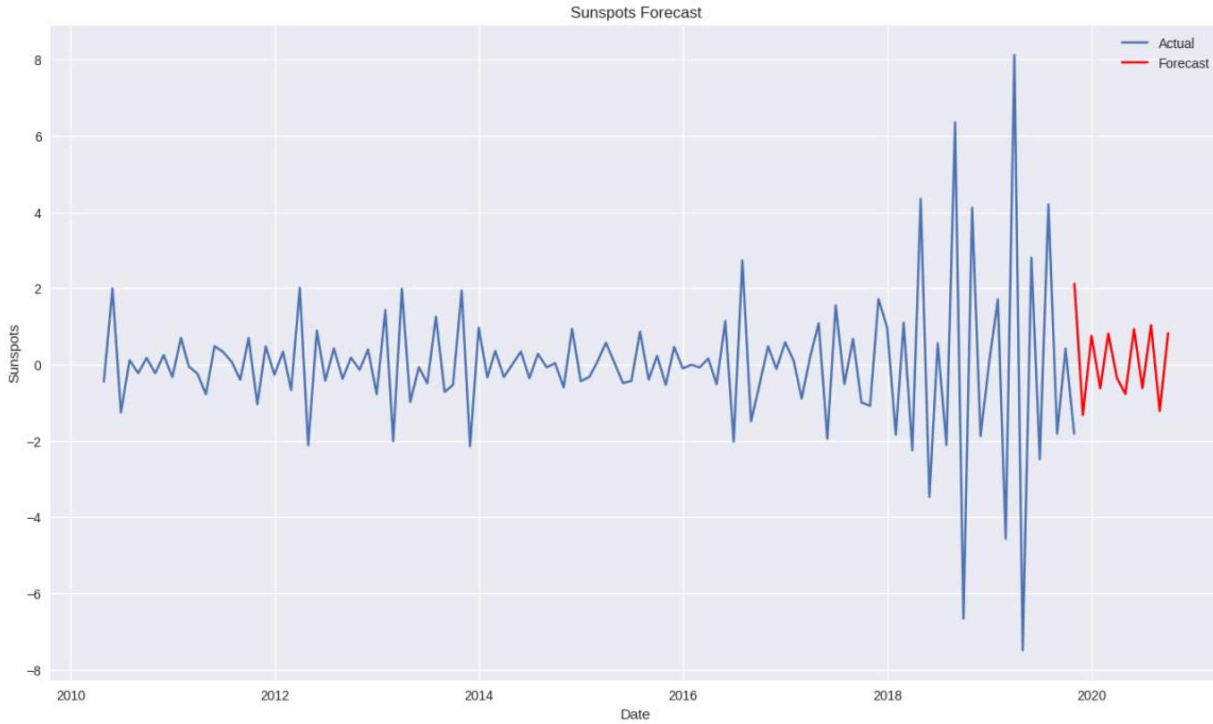

Figure 1.

Figure 2.

Holt Winter's exponential smoothing model was also fitted on the data with the additive trend and seasonality. The seasonal patterns in the data are relatively stable over time and do not exhibit any significant changes in variance over time. Additionally, the level of sunspot activity has remained relatively stable over time, with only minor fluctuations. Therefore, both trend and seasonality were set to additive for the exponential smoothing model. The model yielded a RMSE (root mean squared error) of 3.77, which was compared to that of the best SARIMA model, which had a RMSE of 3.79. As a result of the comparison, the RMSE value of the exponential smoothing model was less; therefore, it performed better for this data and was referred to as the best model for this research.

In Figure 3 the plot of the original data and forecasts of both the SARIMA and the exponential smoothing models are presented.

In Figure 4 the zoomed-in version of Figure 3 is presented (one-year period again).

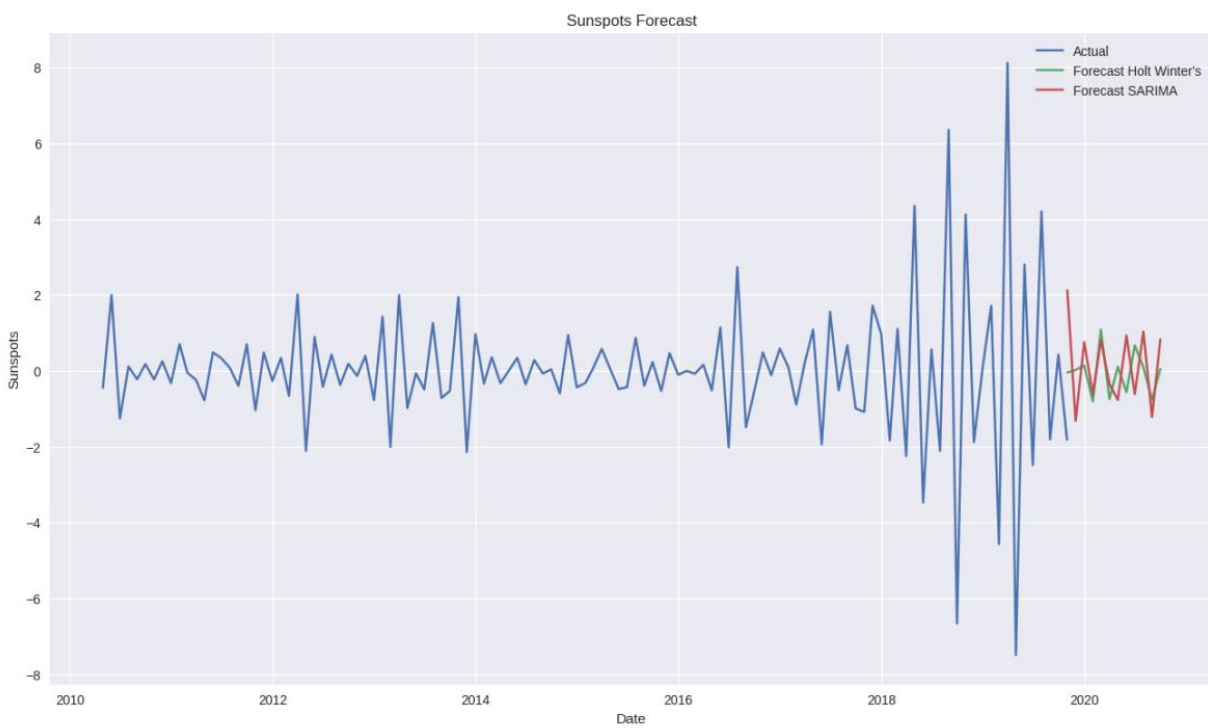

Figure 3.

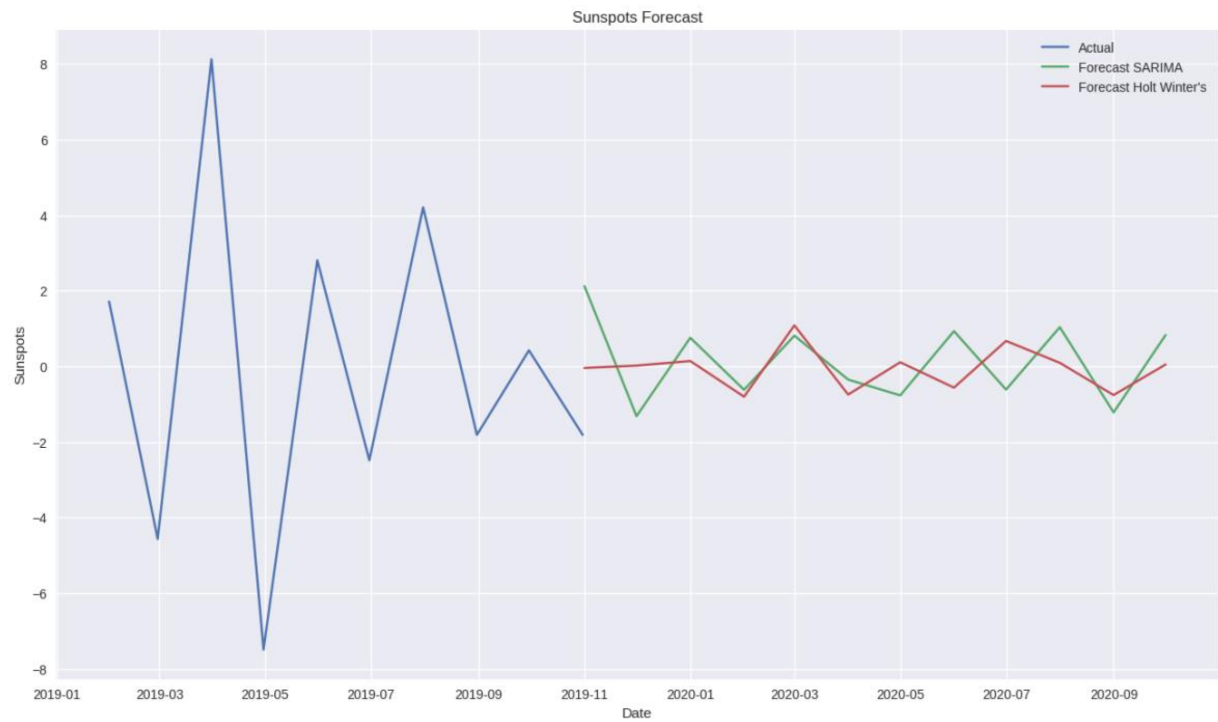

Figure 4.

After the SARIMA model was fitted, the returns of its forecasted values were calculated. For those returns, an ARMA model was identified for the mean dynamic, and then a GARCH model was fitted to make volatility forecasts, which yielded increasing values.

Figure 5 is the plot of the variance forecasts using the GARCH model.

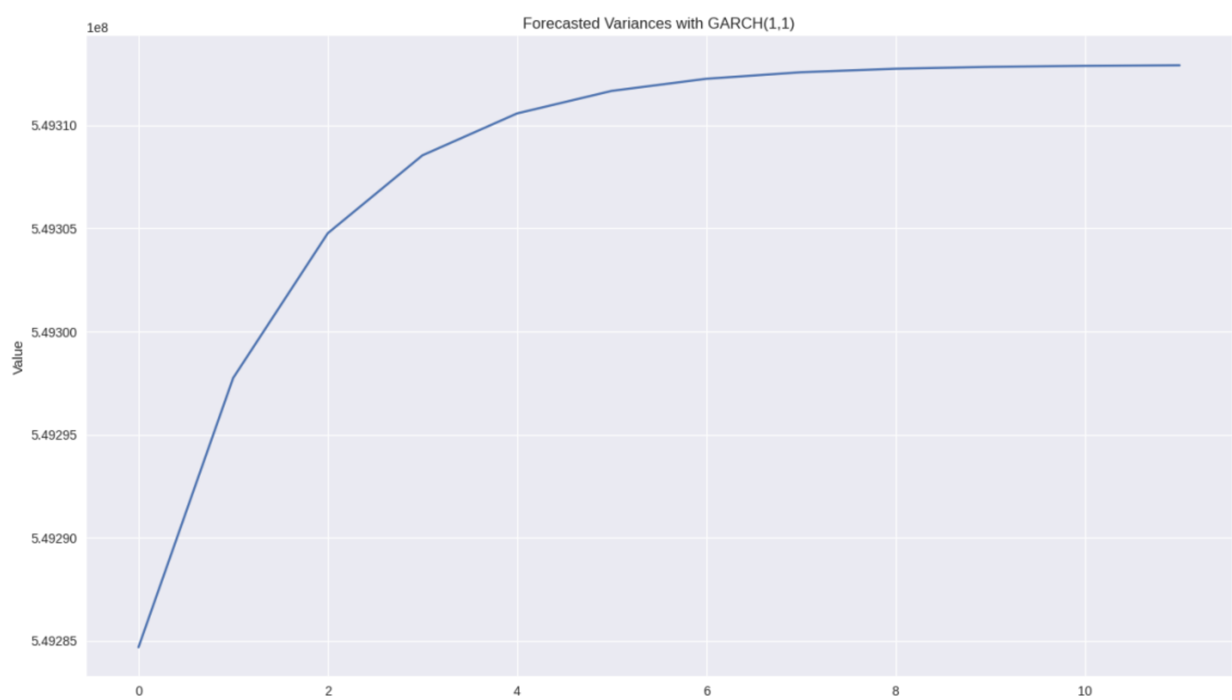

Figure 5.

Overall, the models used on the data were SARIMA, both manual and automated, and exponential smoothing. RMSE was taken as a comparison measure for these models, and the lower one indicated the best model, which was exponential smoothing. Just for the analysis and forecasts of the mean and variation, the ARMA and GARCH models were applied to the returns of the SARIMA forecasts.

Conclusion:

In conclusion, this study employed SARIMA, Exponential Smoothing, ARMA, and GARCH models to forecast sunspots and analyze the mean and volatility dynamics of the return series. The forecasts obtained through these models provide valuable insights for stakeholders interested in researching the sun's behavior and the broader space environment.

The SARIMA and Exponential Smoothing models successfully generated forecasts for sunspots over the next year. These forecasts can assist stakeholders, including scientists, researchers, and policymakers, in enhancing their understanding of sunspot activity and its implications. By having access to reliable predictions, stakeholders can plan and conduct further research, design experiments, and develop strategies to explore and analyze the behavior of the sun.

Furthermore, the ARMA and GARCH models allowed for the prediction of the mean and volatility dynamics of the return series derived from the SARIMA predictions. This information is crucial for stakeholders interested in assessing the stability and fluctuations in sunspot activity. Understanding the mean and volatility of sunspot returns enables stakeholders to evaluate potential risks, make informed decisions, and adapt their approaches accordingly.

Overall, the forecasts provided by the SARIMA, Exponential Smoothing, ARMA, and GARCH models have significant implications for the scientific community and other stakeholders involved in studying the sun and its impact on the space environment. By utilizing these forecasting models, stakeholders can enhance their research efforts, improve the accuracy of their predictions, and ultimately contribute to a deeper understanding of solar physics. The findings presented in this paper pave the way for further exploration and investigation into the dynamic nature of sunspots, fostering advancements in space science and related fields.

References:

[1] Abhinand, A. (2019). Daily Sunspot Data 1818-2019 [Dataset]. Kaggle. https://www.kaggle.com/datasets/abhinand05/daily-sun-spot-data-1818-to-2019

[2] V. Sandeep Kumar. (2018). Sunspot Activity Prediction with Neural Networks. Kaggle. https://www.kaggle.com/sandeepkumar121995/sunspot-activity-prediction-with-neural-networks

[3] M. Ali. (2018). Sunspot Forecasting Using Time Series Analysis. GitHub. https://github.com/mirali361/Sunspot-Forecasting-Using-Time-Series-Analysis

[4] C. Y. Liu, S. Q. Cao, L. H. Lu, & X. Y. Zhang. (2013). Forecasting Sunspot Cycles Using ARIMA Models. Advances in Astronomy, 2013, 1-8.

[5] M. T. Ullah, M. Z. Uddin, & M. A. Matin. (2015). A Hybrid ARIMA and Artificial Neural Network Model for Sunspot Forecasting. International Journal of Scientific & Engineering Research, 6(4), 1283-1291. https://www.ijser.org/researchpaper/A-Hybrid-ARIMA-and-Artificial-Neural-Network-Model-fo r-Sunspot-Forecasting.pdf