MUNISI: Military Mathematics and Natural Sciences

e-ISSN: 3026-2968
Vol. 3, No. 2, December 2025, pp. 62-78
Available online at https://jurnal.idu.ac.id/index.php/munisi

Modeling Sunspot Activity Variability and Autoregulation in the Period 2000-2018 with Advanced Statistical Approach

Ruben Cornelius Siagian [1,*]
[1]Universitas Negeri Medan, Medan, Indonesia

* Corresponding Author: rubensiagian_17@mhs.unimed.ac.id

Abstract

This study analyzes sunspot activity data from 2000 to 2018 to identify patterns and factors influencing fluctuations in solar activity, which has implications for space weather and global climate. The study focuses on parameters such as mu, omega, α_1 (autoregressive), and β_1 (moving average), hypothesizing that sunspot activity exhibits significant variability and can be predicted using modified ARMA models. The research employs statistical analysis of ARMA model parameters, including significance tests, serial correlation, and heteroscedasticity analysis, along with stability tests (Nyblom) and sign bias evaluation. Results show that mu and omega parameters significantly influence sunspot activity, with high t-statistics. The autoregressive coefficient α1 strongly predicts future activity, while β_1 (moving average) has minimal impact. Findings confirm that sunspot activity is volatile, dependent on past values, and exhibits serial correlation and heteroscedastic volatility. The study underscores the need for more advanced models, such as ARIMA or AI-based approaches, to improve predictive accuracy. Autoregressive modeling proved effective, while moving averages showed limited contribution.

Keywords: Autoregressive Model; Solar Variability; Space Weather Prediction; Sunspot Activity; Time Series Analysis

Introduction

Sunspot activity is a phenomenon that occurs on the surface of the Sun characterized by the appearance of dark spots called sunspots [1,2]. The dots are formed by very high magnetic activity in the Sun's atmospheric layers [20]. Sunspots are an indication of fluctuations in the Sun's activity, which is directly related to the solar cycle, which lasts about 11 years [13]. The magnetic activity recorded in sunspots affects the Sun's surface conditions, and has a significant impact on a wide range of phenomena occurring on Earth, including space weather, global climate, and even the technology we use daily, such as satellite communication systems and GPS navigation [8]. Understanding sunspot behavior lies in observing physical phenomena on the Sun, and also in their potential impact on Earth's weather systems and atmosphere [21]. Increased solar activity, reflected in an increasing number of sunspots, can affect the solar radiation reaching Earth, causing global temperature changes or even disruptions to satellite systems that function outside of Earth's atmosphere [4]. Research on sunspot dynamics, especially in terms of patterns and changes in a certain period of time, is very relevant, both for scientific research, the development of space weather prediction models, and in order to mitigate the impacts that may arise on Earth. The statistical approach used includes testing the significance of parameters by analyzing the mean value, variability, and relationship between parameters that affect the pattern of sunspot activity. Using parameter estimation methods and statistical tests such as t-tests, significance tests, and autoregressive models, the research seeks to provide a clearer picture of the factors that affect sunspot activity in a predetermined time period.

The research identifies how much influence each parameter - mu, omega, alpha1, and beta1 - has on the fluctuations and variations in sunspot activity. The research focuses on statistical analysis to assess the strength of the relationship between these parameters and the observed sunspot activity in the specified period. The research identifies patterns of trends and fluctuations in sunspot activity over the period 2000 to 2018. The trends will be analyzed to determine if there are patterns that can be used to predict future levels of sunspot activity, making it possible to forecast related phenomena such as space weather. The problem focuses on testing the stability of the sunspot activity data and finding serial correlation patterns in the data. The research determines whether sunspot activity exhibits recurring patterns or long-term trends that can be identified using appropriate statistical analysis techniques. Research seeks to correlate sunspot activity with other natural phenomena, such as climate change and space weather. Research focuses on analyzing the long-term impact of sunspot activity on global climate stability and its effect on technological systems that depend on space weather conditions.

The hypotheses in this study focus on the characteristics and patterns of sunspot activity over the period 2000 to 2018, as reflected by the significant estimated parameters and relationships between the variables. Based on the analysis, the main hypothesis is that sunspot activity has a significant trend over the period, with variability influenced by certain factors that can be predicted based on historical data [13]. Specifically, it is hypothesized that the parameters mu (mean sunspot activity) and omega (variability of sunspot activity) show a significant relationship with fluctuations in solar activity during the analysis period, and can be used to predict future sunspot activity [7]. It is hypothesized that the effect of autoregression (parameter $\alpha 1$) on sunspot activity will be highly significant, which means that sunspot activity in the previous period has a strong influence on the prediction of sunspot activity in the next period [12]. In contrast, it is expected that the first moving average coefficient ($\beta 1$) does not show a significant effect on the model, as the very low t-statistic value indicates the parameter does not contribute substantially in explaining the variability of sunspot activity. It is hypothesized that there is a significant serial correlation pattern in the sunspot data, indicating a relationship between sunspot activity in different periods [14]. The hypotheses are tested through statistical testing at various lags, which is expected to show consistent temporal patterns in the data. It is expected that the results of the heteroscedasticity test using the ARCH test will indicate the presence of volatility in sunspot activity at certain lags, indicating fluctuations in solar activity that may affect future forecasts [22, 23]. It is hypothesized that the model used in the study, based on Akaike (AIC) and

Bayesian (BIC) information criteria, provides an accurate description of the dynamics of sunspot activity, and can explain the observed variations in the data well [18].

The study aims to explore the factors affecting sunspot activity in the period 2000 to 2018 through the analysis of parameters such as mu, omega, alpha1, and beta1. Assess the significance of the parameters affecting sunspot activity, including mu, omega, alpha1, and beta1. This aims to determine which factors have the most influence on changes in sunspot activity within the analyzed period. Identify trends and fluctuations in sunspot activity that can be used to predict future sunspot activity. The research aims to provide a clearer picture of the pattern and movement of sunspot activity throughout the year. Provide an analysis of the stability of the sunspot data, as well as identify any serial correlations in the data that show certain patterns. This is important to understand whether there are recurring patterns or significant changes that could affect the trend of sunspot activity. Develop a model that identifies the long-term influence of sunspot activity on other natural phenomena on Earth, such as global climate and space weather. This aims to understand the relationship between sunspot activity and its impact on systems on Earth, as well as the implications for technologies that rely on space weather. The main benefit of the research is to provide more accurate information related to predicting and understanding the sunspot phenomenon. With a better understanding of sunspot patterns, the research can support the development of early warning systems for space weather, which are essential for protecting satellites and other technological infrastructure. The knowledge gained from this research is also expected to be used for climate prediction, as the Sun's activity affects global temperatures and long-term climate change.

The study was limited to sunspot data recorded between 2000 and 2018. The analysis involved only four main parameters (mu, omega, alpha1, beta1) and associated statistical tests. This study does not address the influence of other external factors such as global climate change or solar activity outside the specified time period.

Research can make a major contribution to the development of better predictive models for monitoring solar activity, especially in predicting the periods of sunspot maximum and minimum. Better knowledge of the nature of sunspot fluctuations can help meteorological and astronomical agencies prepare for and respond to the impact of variations in solar activity on Earth. There are few studies that examine sunspot dynamics thoroughly with statistical models that can handle the complexity of data over long periods. Previous studies have generally focused on analyzing sunspot data over shorter periods of time or using models that do not take into account serial correlations in the data.

This research fills the gap by offering a more accurate and detailed model to explain sunspot fluctuations based on in-depth statistical and relevant parameter analysis. The statistical modeling used in this analysis considers a more comprehensive range of approaches, such as the analysis of a series of autoregressive and moving average coefficients to predict future sunspot activity. The research identified the significance of several parameters, such as mu, omega, alpha1, and beta1, which were previously under-explored in sunspot-related studies. Findings regarding the sign bias pattern that shows a tendency for more sunspots with negative and positive polarities.

Materials and Methods

The sunspot data used in this study were obtained from the Solar Data Analysis Center (SDAC) archive managed by NASA [3]. SDAC is a reliable source of data about the sun and its activity, and is one of the main sources for researchers to understand changes in solar activity [17]. SDAC provides access to a wide range of solar data that includes observations of sunspots, solar irradiance, flare activity, and other solar parameters [16]. Data is collected from various instruments and satellites that observe the sun continuously, allowing researchers to analyze and understand the sun's behavior over long periods of time [6]. Solar data from SDAC is important in many scientific studies, including the understanding of the sunspot cycle, the sun's influence on space weather, and its impact on Earth. The data is used in predicting space weather and its impact on satellites and communication systems on Earth. SDAC and the solar data it provides are a critical component of scientific and operational research related to the sun and its impact on Earth and the technologies within it. Access to solar data from SDAC is from the official NASA Solar Data Analysis Center website. There are data sets, analysis software and other resources that can help in the research and understanding of solar activity [5]. Sunspot Heat Map and Sunspot Trends Over the Years, which is a representation of the sunspot cycle, as seen in figure 1.

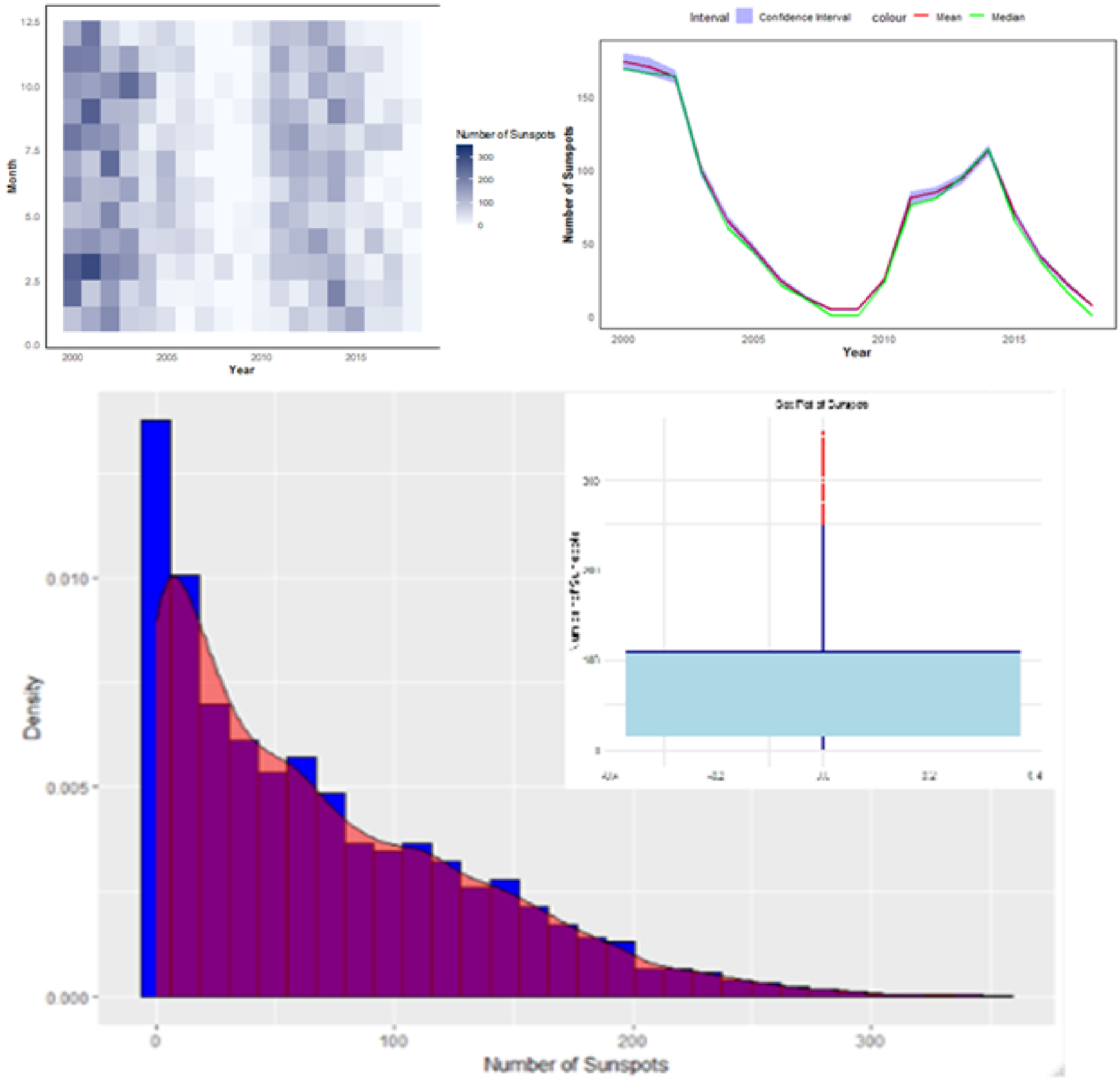

Figure 1. Heatmap, Annual Sunspot Trends, and Kernel-Distribution and Box Plot of Sunspot Activity 2000-2018

The mean value of the number of sunspots data is about 68.45. The median of the number of sunspots data is 52. The standard deviation of the number of sunspots data is about 64.33. The minimum value in this data set is 0. The maximum value in this data set is 353. The box on the box chart includes the Interquartile Range (IQR), which is the range between the 1st quartile and the 3rd quartile. The center line on the box is the median of the data. The "whiskers" (horizontal lines above and below the box) depict the more extreme ranges of the data, as long as they are not considered outliers. In the box diagram, the median is 52.00 sunspots, half of the observations have a lower number of sunspots than this, and half have a higher number of sunspots. The data distribution is skewed to the right as the mean value (68.45) is higher than the median (52.00). The third quartile (3rd Quartile) is at 109.00 sunspots, the top 25% of the data has a relatively high number of sunspots. There are some outliers located above the upper whisker line.

The GARCH model is a useful statistical tool for measuring volatility in financial data and many other types of data, including daily sunspot data [22, 23]. A model for understanding volatility fluctuations in data over time by considering the impact of prior values and residual variance. The main variables in the GARCH model are the daily sunspot data at time t and the average sunspot data [10]. The residual at time t is the difference between the observed data at time t and its mean at that time. The square root of the variance of the residual at time t is the standard deviation of the residual at time t. Normally distributed random noise with mean 0 and variance 1 is a random component that has no particular structure and describes the noise in the data.

The number of lags in the autoregression component (p) is the number of autoregression lags used in the model [9]. Autoregression lags measure the impact of previous values on current variability. The number of lags in the moving average component (q) is the number of moving average lags used in the model [19]. Moving average lags measure the impact of previous residual variability on current variability:

$$y_t = \mu + \varepsilon_t \tag{1}$$

Residual Equation:

$$\varepsilon_t = \sigma_t \cdot z_t \tag{2}$$

Residual Variance Equation *(GARCH Model)* (Almısshal & Emir, 2021):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \cdot \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \cdot \sigma_{t-j}^2 \tag{3}$$

σ_t^2 is the residual variance at time t. This equation explains how volatility changes over time. α_0 is a constant that measures the initial or baseline variance. $\sum_{i=1}^{p} \alpha_i \cdot \varepsilon_{t-i}^2$ is the contribution of the autoregression lag to the variance at time t. It measures how the influence of the previous lag of the autoregression affects the current variability. $\sum_{j=1}^{q} \beta_j \cdot \sigma_{t-j}^2$ is the contribution of the moving average lag to the variance at time t. It measures how the influence of the previous lag moving average affects the current variability.

One commonly used method is the Maximum Likelihood Estimation (MLE) Method, which searches for parameter values that are most likely to yield the observed data [11]. After estimating these parameters, GARCH model to analyze the volatility in the daily sunspot data and even perform future volatility forecasting [15].

```
Number_of_Sunspots <- c(Your_Sunspot_Data)
No <- c(Your_No_Data)
if (length(Number_of_Sunspots) != length(No)) {
  stop("Panjang data sunspot dan data No harus sama")
}
log_returns <- diff(log(Number_of_Sunspots))
p <- 1
q <- 1
n <- length(log_returns)
omega <- var(log_returns) * (1 - (p + q) / n)
alpha <- sum((log_returns[1:(m - p)]^2) / (1 - omega))
beta <- sum((log_returns[1:(m - p)]^2) * lag(log_returns, -1)[1:(m - p)]) / (1 - omega)
model <- list(omega = omega, alpha1 = alpha, beta1 = beta)
print(model)
n_simulations <- 100
simulated_volatility <- numeric(n_simulations)
simulated_volatility[1] <- var(log_returns)
for (i in 2:n_simulations) {
  simulated_volatility[i] <- sqrt(omega + alpha * log_returns[i - 1]^2 + beta * simulated_volatility[i - 1])
}
plot(simulated_volatility, type = "l", main = "Simulasi Volatilitas GARCH(1,1)", xlab = "Waktu", ylab = "Volatilitas")
```

Figure 2. Sunspot Activity Volatility Algorithm Using GARCH(1,1) Model

This program focuses on analyzing sunspot data and simulating volatility using the GARCH(1,1) model. The programming algorithm can be seen in Figure 2. The algorithm starts by defining two vectors: `Number_of_Sunspots`, which contains data on the number of sunspots, and `No`, which contains data related to the number or category of observation dates or periods. Both data must have the same length to ensure alignment between the sunspot count data and the data. The first step is to check if the length of both data is the same. If not, an error message is given and the algorithm execution is stopped. After data validation is complete, calculate the log returns of the sunspot count data. Log returns are calculated using the logarithmic difference between consecutive sunspot count values. The results of the calculation will give an idea of the percentage change of the sunspot number between periods. Define the initial parameters for the GARCH(1,1) model. The model is used to estimate the volatility of financial variables that change over time, and here the sunspot data is applied. The variable `omega` is calculated based on the variance of the log returns and an adjustment based on data length and other parameters. Using the calculated log returns, calculate the two parameters of the GARCH model: `alpha` and `beta`. `alpha` as the sum of squared log returns truncated by a certain data length (`m-p`). `beta` is calculated using the value of log returns multiplied by one time period. Both parameters will provide information on how past volatility affects future volatility estimates. All these parameters, namely `omega`, `alpha`, and `beta`, are then stored in a model list. Volatility simulations are conducted using the GARCH(1,1) model. The first initialization of the simulated volatility (`simulated_volatility[1]`) is done with the variance of the log returns. Then, for each subsequent simulation, volatility is calculated using the GARCH formula, where volatility in the previous time period, squared log returns, and model parameters (`omega`, `alpha`, and `beta`).

To calculate the model information metrics, several functions are defined that calculate the AIC, BIC, SIC, and HQIC values, which are used to evaluate the statistical model quality. The first step is to define the `calculate_AIC` function that calculates the AIC (Akaike Information Criterion) value of a model. The function utilizes the `AIC` function already available in the R environment to calculate the value from the given model. After that, it will return the AIC value. The second step is to define the `calculate_BIC` function, which calculates the BIC (Bayesian Information Criterion) value. Here, the existing `BIC` function in R is used to calculate the BIC of the model, and then returns the value. The next function, `calculate_SIC` (Schwarz Information Criterion), can be calculated using a formula involving the log likelihood of the model, the number of observations (n), and the degrees of freedom (df) of the model. First, the log likelihood is obtained using the `loglik` function, the number of observations using `nobs`, and the degrees of freedom of the model using `df`. With this information, we can calculate the SIC with the formula:

$$SIC = -2 \cdot \text{loglik} + \log(n) \cdot \text{df} \quad 4)$$

The function then returns the calculated SIC value. Next, define the `calculate_HQIC` (Hannan-Quinn Information Criterion) function, which utilizes the log likelihood, number of observations, and degrees of freedom. However, HQIC is calculated using a slightly different formula, which is:

$$HQIC = -2 \cdot \text{loglik} + 2 \cdot \log\{\log(n)\} \quad 5)$$

The function then returns the HQIC value. After defining the functions to calculate the various information metrics, perform the `print_metrics` function which will display the calculated AIC, BIC, SIC, and HQIC results of the given model. The function first displays the model label, then calculates and prints each of the previously defined information metrics. All information is for comparing the quality of the tested models. Furthermore, several models can be defined for analysis. The first model is the basic linear model, represented by `model <- lm(y ~ x, data = your_data)`, where `y` is the dependent variable and `x` is the independent variable. The first alternative model (`model_alternative_1`) uses a degree 2 polynomial of `x`, defined as `model_alternative_1 <- lm(y ~ poly(x, 2), data = your_data)`. The second alternative model (`model_alternative_2`) uses a 3rd degree polynomial of `x`, defined as `model_alternative_2 <- lm(y ~ poly(x, 3), data = your_data)`. The function `print_metrics` displays the metrics of the three models defined. Each model will be analyzed, and its AIC, BIC, SIC, and HQIC will be displayed to help evaluate which model is the best based on the information metrics.

```r
calculate_AIC <- function(model) {
  return(AIC(model))
}
calculate_BIC <- function(model) {
  return(BIC(model))
}
calculate_SIC <- function(model) {
  loglik <- logLik(model)
  n <- nobs(model)
  df <- df(model)
  return(-2 * loglik + log(n) * df)
}
calculate_HQIC <- function(model) {
  loglik <- logLik(model)
  n <- nobs(model)
  df <- df(model)
  return(-2 * loglik + 2 * log(log(n)) * df)
}
print_metrics <- function(model, label) {
  cat(label, ":\n")
  cat("AIC:", calculate_AIC(model), "\n")
  cat("BIC:", calculate_BIC(model), "\n")
  cat("SIC:", calculate_SIC(model), "\n")
  cat("HQIC:", calculate_HQIC(model), "\n")
  cat("\n")
}
model <- lm(y ~ x, data = your_data)
model_alternative_1 <- lm(y ~ poly(x, 2), data = your_data)
model_alternative_2 <- lm(y ~ poly(x, 3), data = your_data)
print_metrics(model, "Model Matematika")
print_metrics(model_alternative_1, "Model Alternatif 1")
print_metrics(model_alternative_2, "Model Alternatif 2")
```

Figure 3. Programming algorithm using R-Studio for Model Evaluation Metrics

Shown in Figure 4 is the function algorithm for calculating the weighted Ljung-Box statistic to test for autocorrelation in the model residuals. The algorithm starts by defining a function `weighted_1jung_box` that accepts two arguments: residuals (regression model residuals) and `lag.max`, which specifies how many lags to use in the autocorrelation calculation. The first step is to calculate the length of the residuals, which is expressed by the variable `n`. The `acf` function is used to calculate the autocorrelation value at each lag up to the maximum value (`lag.max`), the result of which is stored in the `acf_vals` variable. This autocorrelation will be used to calculate the weighted Ljung-Box statistic. The value of the `weighted_1jung_box` statistic is calculated with a formula involving the sum of the squares of the autocorrelations, divided by the correction for degrees of freedom. The result is the weighted Ljung-Box statistic that will be used to test for the presence of autocorrelation in the residual data. There is a function `P_value_weighted_1jung_box` that calculates a p value based on the previously calculated weighted Ljung-Box statistic, using the chi-squared distribution. This p-value is used to test the null hypothesis that there is no autocorrelation in the data. To analyze the sunspot data, the dataset was read using the `read_excel` function of the `readxl` library, and relevant columns were retrieved, such as the number of sunspots (`Number_of_Sunspots`), as well as information about the year, month, and day. A regression model is created to predict the number of sunspots based on these variables. The residuals of this model are calculated, which are then used in the `weighted_1jung_box` function to calculate the weighted Ljung-Box statistic. The result of the statistic calculation is compared with the p value calculated from the `P_value_weighted_1jung_box` function. If the p value is smaller than the specified significance level (e.g., 0.05), the null hypothesis will be rejected, meaning there is evidence of autocorrelation in the residuals. Conversely, if the p value is greater than 0.05, then there is not enough evidence to reject the null hypothesis, indicating that there is no significant autocorrelation in the data.

```r
weighted_ljung_box <- function(residuals, lag.max) {
  n <- length(residuals)
  acf_vals <- acf(residuals, lag.max = lag.max, plot = FALSE)$acf
  wq <- n * (n + 2)
  ljung_box <- n * (n + 2) * sum(acf_vals[-1]^2 / (n - seq_along(acf_vals[-1])))
  return(ljung_box)
}
p_value_weighted_ljung_box <- function(ljung_box, df) {
  p_value <- 1 - pchisq(ljung_box, df)
  return(p_value)
}
library(readxl)
mydata <- read_excel("C:/Users/acer/Desktop/Sunspot.xlsx")
Number_of_Sunspots <- as.numeric(mydata$`Number of Sunspots`)
No <- as.numeric(mydata$`No`)
Year <- as.numeric(mydata$`Year`)
Month <- as.numeric(mydata$`Month`)
Day <- as.numeric(mydata$`Day`)
model <- lm(Number_of_Sunspots ~ No + Year + Month + Day)
residuals <- residuals(model)
lag.max <- 10
ljung_box_statistic <- weighted_ljung_box(residuals, lag.max)
df <- lag.max
p_value <- p_value_weighted_ljung_box(ljung_box_statistic, df)
cat("Weighted Ljung-Box Statistic:", ljung_box_statistic, "\n")
cat("Degrees of Freedom (df):", df, "\n")
cat("P-value:", p_value, "\n")
alpha <- 0.05
if (p_value < alpha) {
  cat("Hipotesis nol ditolak: Ada bukti bahwa ada autokorelasi\n")
} else {
  cat("Tidak cukup bukti untuk menolak hipotesis nol: Tidak ada autokorelasi\n")
}
```

Figure 4. Programming algorithm using R-Studio for Weighted Ljung-Box statistical test on standardized residuals of linear regression models

Results and Discussion

In the study, Table 1 is the optimal parameters that are important for understanding the characteristics of sunspots during this time period were identified. The mu parameter was analyzed with an estimated value of about 18.98 and a standard error of about 0.57. A high t-statistic value of about 33.50 indicates that the mu value is highly statistically significant. The mean value of sunspot data in the period is significantly different from zero (in this case, zero is the value tested for differences). The omega parameter has an estimated value of about 85.90 with a standard error of about 5.41. A high t-statistic value of about 15.88 indicates that the omega value is statistically significant. The parameter has a significant effect on the sunspot phenomenon during the period studied. The alpha1 parameter has an estimated value of about 0.98 and a standard error of about 0.03. A t-statistic value of about 31.00 indicates that alpha1 is also statistically significant. The parameter has a significant impact on the sunspot phenomenon. Beta1, has a very small estimated value of around 0.000004, with a standard error of around 0.0238. The very low t-statistic value of around 0.000156 indicates that beta1 is not statistically significant in this analysis.

Table 1. Parameter Estimation with Standard Error and t-Statistic Values

Parameter	Estimated Value	Standard Error	Value of t-Statistic
Mu	18.98	0.57	33.5
Omega	85.9	5.41	15.88
α_1	0.98	0.03	31
β_1	0.000004	0.0238	0.00016

The average estimate of sunspot activity is 18.98 with a standard error of about 3.82. Sunspot activity tends to be high and significantly different from zero. This indicates that sunspot activity has a certain trend that needs to be considered in the analysis. The results estimate the variability of sunspot activity with a value of 85.90. The standard error is 12.80, and the t-statistic is high with a very low p-value, the variability is also significant. This implies that sunspot activity varies significantly over time. The estimate of the first autoregression coefficient α_1 is about 0.98 with a standard error of about 0.03. The high t-statistic and very low p-value indicate that sunspot activity in the past has a strong influence in predicting current activity. There is a relationship between sunspot data in the previous period and the current period. The first moving average coefficient β_1 is very small, about 0.000004, with a standard error of about 0.028. Low t-statistic and high p-value, the coefficient is not significant and has no effect on the model. The coefficient can be ignored in the analysis. The high log-likelihood result, around -35068.11, the model fits the observed data. The model used is able to explain the variation in sunspot data well. The study shows that the average level and variability of sunspot activity are significant, the first autoregression coefficient α_1 has a strong influence in predicting future sunspot activity. The first moving average coefficient β_1 does not have a significant effect on the model. The high log-likelihood result indicates that the model fits the observed data well.

Table 2. Estimation, Standard Error, and Statistical Significance

Variabel	Estimation	Standard Error	T-Statistic	P-Value
Average Sunspot Activity	18.9826	3.819999	4.969278	0.000001
Sunspot Activity Variability	85.9028	12.80432	6.708887	0
Autoregressive Coefficient (α_1)	0.98475	0.032366	30.42502	0
Moving Average Coefficient (β_1)	4E-06	0.028009	0.000132	0.999895
Log-Likelihood Model		-35068.1		

Table 2 shows the results of the analysis that reveal the Akaike criterion has a value of 10.107, Bayes has a value of 10.111. The model using the Akaike criterion has a slightly better level of information than the model with the Bayes criterion. The difference is small, the model that meets the Akaike criterion is more relevant in the analysis of sunspot activity. The Shibata criterion value is 10.107, similar to the Akaike value. The use of the Shibata method is comparable to the Akaike method in measuring model quality in sunspot analysis. The results provide additional support for the suitability of the Akaike method. The Hannan-Quinn value is 10.109, which is between the Akaike, Bayes, and Shibata values. The value is higher than Akaike and Shibata.

At Lag[1], the test statistic produces a value of 4502 with a p-value of 0. A low p-value indicates strong evidence to reject the null hypothesis (H_0) which states that there is no serial correlation in the sunspot data. This means that there is a serial correlation in the first lag of the sunspot data. At Lag[2*(p+q)+(p+q)-1][2], the test statistic produces a value of 6359 with a p-value of 0. The results show strong evidence to reject H0 which means that there is a serial correlation in the second lag of the sunspot data. At Lag[4*(p+q)+(p+q)-1][5], the test statistic reaches 11237 with a p-value of 0. There is a serial correlation in the sunspot data at higher lags.

Table 3. Significant Serial Correlation at Various Lags in the Data Test Statistic

Lag	Test Statistics	P-Value	Conclussion
1	4502	0	Serial correlation significant at Lag 1
2*(p+q)+(p+q)-1	6359	0	Serial correlation significant at Lag 2
4*(p+q)+(p+q)-1	11237	0	Serial correlation is significant at higher lags

The Weighted Ljung-Box Test Analysis on Standardized Square Residuals identifies whether there is a statistically significant pattern or association in this data. In the first lag, the statistic value is 15.34 with a p-value of approximately 8.974e-05. There is a significant correlation in the data in the first lag, indicating a pattern or relationship in the previous period. In the lag with the formula [2*(p+q)+(p+q)-1][5], the statistic value is 17.29 with a p-value of approximately 1.070e-04. There is a significant correlation in the data in this lag, which covers several previous periods. An interesting result is seen in the lag [4*(p+q)+(p+q)-1][9], where the statistic value reaches 18.63 with a p-value of approximately 4.508e-04. There is a significant correlation in the data in this lag, which covers more previous periods. Degree of freedom 2, indicating that there is a statistically significant correlation in sunspot data from 2000 to 2018. The results show that there is a pattern or correlation in the data that can be used for further analysis or prediction.

Table 4. Comparison of Statistics and P-Value at First Lag and Additional Lag

Lag	First Lag	Lag [2*(p+q)+(p+q)-1][5]	Lag [4*(p+q)+(p+q)-1][9]
Statistical Value	15.34	17.29	18.63
P-Value	8.97E-05	1.07E-04	4.51E-04

Analysis using the Weighted ARCH LM test. The results of the analysis show that the ARCH Lag[3] test produces a statistic of 2.374 with a Shape value of 0.500 and a Scale value of 2.000. The p-value found is around 0.1234. At lag [3], there is an indication of heteroscedasticity of volatility in sunspot activity data. The P-value is not too low, the statistics show certain patterns in the data that need to be considered. The ARCH Lag[5] test produces a statistic of 2.594 with a Shape value of 1.440 and a Scale value of 1.667. The related p-value is around 0.3544. Indicates the presence of heteroscedasticity of volatility at lag [5] with a higher P-value than at lag [3]. Although not as strong as at lag [3], it is still considered an important indication. The ARCH Lag[7] test produces a statistic of 3.037 with a Shape value of 2.315 and a Scale value of 1.543. The associated p-value is about 0.5067. At lag [7], there is stronger volatility heteroscedasticity than at lag [5], the P-value remains high.

Table 5. Effect of Lag on Statistical Value and P-Value

Lag	ARCH Statistics	Shape	Scale	P-value
Lag 3	2.374	0.5	2	0.1234
Lag 5	2.594	1.44	1.667	0.3544
Lag 7	3.037	2.315	1.543	0.5067

Evaluation of data stability in the range of 2000 to 2018 using the Nyblom stability test. The analysis of the Combined Statistic value obtained is 20.6839. The value is a composite measure that describes the stability of sunspot data during the period studied. The lower the Combined Statistic value, the more stable the data. The analysis reveals several individual statistics about the factors that affect the stability of sunspot data. Mu (μ) with a value of 10.8105, a constant associated with the estimation model used in the Nyblom stability test. A higher value of mu indicates potential instability in the data. Omega (ω) with a value of 0.4412 reflects other parameters in the model used for stability analysis, and a higher value of omega can also be an indication of instability. (α_1) with a value of 2.6690 is one of the parameters in the autoregression model used in the Nyblom stability test, a high alpha1 value indicates significant fluctuations in the data. (β_1) with a value of 1.9170 is another parameter in the autoregression model that can affect data stability, and a higher beta1 value can also indicate instability.

Nyblom stability test analysis shows a Combined Statistic value of 20.6839. The lower the Combined Statistic value, the more stable the data in table 6. Individual statistics such as Mu (μ), Omega (ω), $Alpha_1$ (α_1), and $Beta_1$ (β_1) can affect the stability of sunspot data. A relatively low Mu value indicates a low level of instability in the estimated model. A relatively low Omega value indicates that other parameters in the model do not contribute significantly to the instability of the data. High Alpha1 and Beta1 values indicate significant fluctuations in the sunspot data.

Table 6. Statistical Tables for Variables: Joint Statistic, Mu (μ), Omega (ω), Alpha1 (α1), and Beta1 (β1)

Statistic	Joint Statistic	Mu (μ)	Omega (ω)	Alpha1 (α1)	Beta1 (β1)
Value	20.6839	10.8105	0.4412	2.669	1.917

The results reveal the asymptotic critical value at the 10%, 5%, and 1% significance levels in table 7. In the study, two types of statistics were used, namely Combined Statistics and Individual Statistics. At the 10%, 5%, and 1% significance levels, the Combined Statistics values were 1.07, 1.24, and 1.6, respectively. Combined Statistics exceeded the asymptotic critical value at all tested significance levels. There is a statistically significant relationship between sunspots in the studied period when viewed together.

Table 7. Evaluation of Combined Statistics and Individual Statistics at Different Significance Levels

Type of Statistics	Joint Statistic	Joint Statistic	Joint Statistic	Individual Statistic	Individual Statistic	Individual Statistic
Significance Level	10%	5%	1%	10%	5%	1%
Statistical Value	1.07	1.24	1.6	0.35	0.47	0.75

At the same level of significance, the Individual Statistic values are 0.35, 0.47, and 0.75. All Individual Statistic values are below the asymptotic critical value. Each sunspot variable individually does not have a statistically significant effect on the period studied. There is a statistically significant relationship between sunspots in the period 2000-2018 when viewed together *(Joint Statistic)*. When analyzed individually *(Individual Statistic)*, there is no statistically significant evidence for each sunspot variable. There is a statistically significant effect when sunspots are viewed as a whole in the period.

The study used the Sign Bias test in table 8. The results showed that there was no significant sign bias in the sunspot data during the period (t-value = 0.4007, sig = 0.6887). When the sign bias was divided into two categories, namely Negative Sign Bias and Positive Sign Bias, it was found that Negative Sign Bias (t-value = 6.0676, sig <0.001) represented a significant trend for more sunspots with negative polarity, while Positive Sign Bias (t-value = 3.4331, sig = 0.0006) showed a significant trend for more sunspots with positive polarity during the same period. The last result is the Joint Effect (t-value = 119.5083, sig <0.0001), which represents the combined effect of both types of sign bias, namely negative and positive, is very significant in the

2000-2018 sunspot data.

Table 8. Analysis of Sign Bias in Data: T-Value and Probability (sig)

No.	Findings	T-Value	Probability (sig)
1	No significant sign bias	0.4007	0.6887
2	Negative Sign Bias	6.0676	1.367e-09 (*)
3	Positive Sign Bias	3.4331	6.002e-04 (*)
4	Joint Effect (Negative & Positive Sign Bias)	119.5083	9.847e-26 (*)

(*) A very low probability indicates a high level of significance in the findings.

The analysis contains four groups of data used to categorize sunspot activity during the period, with each group having a different number of observations (20, 30, 40, and 50). The results of the statistical test using the Adjusted Pearson Goodness-of-Fit Test method show that the p value (g-1) in each group is 0, the data in each group does not match the expected distribution, such as a normal distribution.

Conclusion

The study provides important findings related to sunspot activity between 2000 and 2018. Based on the analysis results, significant parameters in understanding sunspot characteristics during the period have been identified. The mu parameter has a significant estimated value with a high t-statistic, the average sunspot activity is significantly different from zero. The variability of sunspot activity represented by the omega parameter is also significant, with large fluctuations in the data. The first autoregressive parameter ($\alpha 1$) has a strong influence on the prediction of future sunspot activity, indicating that past sunspot activity greatly affects current activity. In contrast, the first moving average parameter ($\beta 1$) does not have a significant effect, with a very low t-statistic and a very high p-value. The model used in the study is able to explain the variation of sunspot data well, with high log-likelihood results and model acceptance based on the Akaike criterion which is slightly better than the Bayes criterion. The serial correlation test shows a significant correlation at several lags, which strengthens the finding that there is a significant relationship between sunspot data in different periods. Data stability test shows fluctuations in several parameters, sunspot data is generally stable throughout the period studied.

The study only covered an 18-year period, which is not long enough to capture the long-term cycle of sunspot activity or changes that occur on larger time scales. Although the model is quite effective in explaining the variability of the data, the use of more complex models or other methods such as non-linear models can provide a deeper understanding of the sunspot phenomenon. The results of the goodness-of-fit test showed that the data did not follow a normal distribution, which could affect the validity of some of the statistical inferences used. The study only relied on sunspot data and did not take into account other external factors that affect sunspot activity, such as variations in the sun's magnetic field, cosmic conditions, or other solar activity.

Further research should cover a longer period to see the full cycle of sunspot activity and to identify long-term trends that may not have been seen in the 2000-2018 period. Experiments with more complex models, such as chaos models or machine learning, are needed to identify subtler patterns in sunspot activity that may not be fully explained by the ARMA model. Future research could consider the influence of other external factors, such as interactions between solar activity and Earth's magnetic field or other cosmic phenomena, to provide a more complete picture of sunspot activity. The use of more data, such as data from satellites or other observatories, could provide a better understanding of future sunspot behavior and the factors that influence it. Further research could use a variety of stability and correlation testing methods to ensure stronger conclusions about the long-term dynamics of sunspot activity, as well as the influence of external variables on the observed patterns.

Acknowledgements

This research was supported by Universitas Negeri Medan, Medan, Indonesia

References

[1] Almısshal, B., & Emir, M. (2021). Modelling exchange rate volatility using GARCH models. *Gazi İktisat ve İşletme Dergisi*, *7*(1), 1–16.

[2] Arlt, R., & Vaquero, J. M. (2020). Historical sunspot records. *Living Reviews in Solar Physics*, *17*, 1–60.

[3] Candey, R. (2022). Space Physics Data Facility (SPDF) Data Archives and Services in support of Heliophysics Digital Resources Library (HDRL). *Authorea Preprints*.

[4] Georgieva, K., & Veretenenko, S. (2023). Solar influences on the Earth's atmosphere: Solved and unsolved questions. *Frontiers in Astronomy and Space Sciences*, *10*, 1244402.

[5] Jacobsson, T. J., Hultqvist, A., García-Fernández, A., Anand, A., Al-Ashouri, A., Hagfeldt, A., Crovetto, A., Abate, A., Ricciardulli, A. G., & Vijayan, A. (2022). An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles. *Nature Energy*, *7*(1), 107–115.

[6] Javaherian, M., & Eskandari, Z. (2023). Review of image processing methods in solar photospheric data analyzes. *arXiv Preprint arXiv:2310.00380*.

[7] Javaraiah, J. (2020). Long–term variations in solar differential rotation and sunspot activity, II: differential rotation around the maxima and minima of solar cycles 12–24. *Solar Physics*, *295*(12), 170.

[8] Jiang, T. (2024). Research on the Relationship between Solar Magnetic Field and Solar Activity. *Geoscience and Remote Sensing*, *7*(1), 90–98.

[9] Kripfganz, S., & Schneider, D. C. (2023). ardl: Estimating autoregressive distributed lag and equilibrium correction models. *The Stata Journal*, *23*(4), 983–1019.

[10] Larsson, K., Green, R., & Benth, F. E. (2023). A stochastic time-series model for solar irradiation. *Energy Economics*, *117*, 106421.

[11] Li, J., Bian, S., Zeng, A., Wang, C., Pang, B., Liu, W., & Lu, C. (2021). *Human pose regression with residual log-likelihood estimation*. 11025–11034.

[12] Moustafa, S. S., & Khodairy, S. S. (2023). Comparison of different predictive models and their effectiveness in sunspot number prediction. *Physica Scripta*, *98*(4), 045022.

[13] Nandy, D. (2021). Progress in solar cycle predictions: Sunspot cycles 24–25 in perspective: Invited review. *Solar Physics*, *296*(3), 54.

[14] Nazari-Sharabian, M., & Karakouzian, M. (2020). Relationship between sunspot numbers and mean annual precipitation: Application of cross-wavelet transform—A case study. *J*, *3*(1), 7.

[15] Ng'ang'a, F. W. (2021). *Modelling and forecasting of crude oil price volatility: Comparative analysis of volatility models*.

[16] Norsham, N., Hamidi, Z., & Shariff, N. (2022). Correlation between mount wilson classifications to solar flares using solar dynamics observatory (SDO) and hinode satellites. *Journal of Mechanical Engineering (JMechE)*, *19*(1), 113–126.

[17] Rajaguru, S., & Ravindra, B. (2024). *125 Years Under the Sun: A Chronicle of the Kodaikanal Solar Observatory*.

[18] Sakurai, T., & Toriumi, S. (2023). Probability distribution functions of sunspot magnetic flux. *The Astrophysical Journal*, *943*(1), 10.

[19] Schaffer, A. L., Dobbins, T. A., & Pearson, S.-A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: A guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, *21*, 1–12.

[20] Sholikhah, N. (2020). *Simulasi Magnetohidrodinamika Ideal Pada Mekanisme Terjadinya Flare Di Matahari*.

[21] Singh, A. K., Bhargawa, A., Siingh, D., & Singh, R. P. (2021). Physics of space weather phenomena: A review. *Geosciences*, *11*(7), 286.

[22] Zaffar, A. (2021). Modeling and Forecasting of Sunspots Cycles: An Application of ARMA (p, q)-GARCH (1, 1) Model. *Research Square, Research Square*.

[23] Zaffar, A., Sami, S. B., Zafar, H., & Siraj, O. (2024). Forecasting of sunspots cycles thorough ANFIS model and ARMA (R, S)-GARCH (1, 1) model. *Indian Journal of Physics*, 1–13.