Kappa Journal
Program Studi Pendidikan Fisika
FMIPA Universitas Hamzanwadi
http://e-journal.hamzanwadi.ac.id/index.php/kpj/index

April 2023. Vol. 7, No. 1
e-ISSN: 2549-2950
pp. 114-127

Analysis of Solar Flux and Sunspot Correlation Case Study: A Statistical Perspective

[1]Budiman Nasution, [2]Lulut Alfaris, [3]Ruben Cornelius Siagian, [4]Habibi Azka Nasution
[1,3]Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Indonesian, 20221
[2]Marine Technology Department, Politeknik Kelautan dan Perikanan Pangandaran, 46396. [4]Department of Information and Technology, Universitas Pembinaan Masyarakat Indonesia, Indonesia, 21201
Email Korespondensi: rubensiagian775@gmail.com

Article Info

Article History
Received: 12 March 2023
Revised: 25 April 2023
Published: 30 April 2023

Keywords
Solarflux; Sunspots; Radiation intensity; Analysis correlation; Impact solarflux and; sunspots

Abstract

Analysis of Solar Flux and Sunspot Correlation Case Study: A Statistical Perspective. *This analysis examines the relationship between the number of solar flares and the number of sunspots in 2005 using 11 observations in months 2 to 12. The number of solar currents measures the intensity of the radiation emitted by the Sun, while the number of sunspots measures the number of sunspots on the surface of the Sun. Multivariate linear regression analysis was used to analyze the relationship between Solar Current Rate and Number of Sunspots. The results of the analysis show that the coefficient of the Amount of Solar Current is 1.1239 with a significant t value of 2.510 (probability that there is no effect on the Number of Sunspots is 3.33%). The linear regression model has good results with an F-statistic value of 6.301 and a p-value of 0.0333, with an R-squared value of 0.4118 which indicates that 41.18% of the variation in the number of sunspots is influenced by variations in the amount of solar currents. The corrected R-squared value is 0.3464 indicating that there are still variations in the number of sunspots that cannot be explained by variations in the number of solar currents. ARIMA analysis results show an MA coefficient of 0.7351 with an average value of 45.9542 and a s.e value of 0.2590 and 6.1550 respectively. The AIC, AICc, and BIC values are 92.97, 96.4, and 94.16. The error results in the training set show that the ME value is 0.2615561, the RMSE value is 12.16969, the MAE value is 9.03306, the MPE value is -15.14689, the MAPE value is 30.42013, and the MASE value is 0.674109. The ACF1 value in the exercise set is 0.0808969.*

Informasi Artikel

Sejarah Artikel
Diterima: 12 Maret 2023
Direvisi: 25 April 2023
Dipublikasi: 30 April 2023

Kata kunci
Petunjuk penulisan; Jurnal prisma; template artikel

Abstrak

Analisis ini mengkaji hubungan jumlah semburan matahari dengan jumlah bintik matahari pada tahun 2005 dengan menggunakan 11 pengamatan pada bulan ke-2 sampai ke-12. Jumlah arus matahari mengukur intensitas radiasi yang dipancarkan Matahari, sedangkan jumlah bintik matahari mengukur jumlah bintik matahari di permukaan Matahari. Analisis regresi linier multivariat digunakan untuk menganalisis hubungan antara Solar Current Rate dan Jumlah Sunspots. Hasil analisis menunjukkan bahwa koefisien Arus Matahari sebesar 1,1239 dengan nilai t signifikan sebesar 2,510 (probabilitas tidak berpengaruh terhadap Jumlah Sunspot sebesar 3,33%). Model regresi linier memiliki hasil yang baik dengan nilai F-statistic 6,301 dan p-value 0,0333, dengan nilai R-squared 0,4118 yang menunjukkan bahwa 41,18% variasi jumlah bintik matahari dipengaruhi oleh variasi jumlah arus matahari. Nilai R-squared terkoreksi sebesar 0,3464 menunjukkan bahwa masih terdapat variasi jumlah bintik matahari yang tidak dapat dijelaskan oleh variasi jumlah arus matahari. Hasil analisis ARIMA menunjukkan koefisien MA sebesar 0,7351 dengan nilai rata-rata 45,9542 dan nilai se masing-masing sebesar 0,2590 dan 6,1550. Nilai AIC, AICc, dan BIC adalah 92,97, 96,4, dan 94,16. Hasil error pada training set menunjukkan nilai ME sebesar 0.2615561, nilai RMSE sebesar 12.16969, nilai MAE sebesar 9.03306, nilai MPE sebesar -15.14689, nilai MAPE sebesar 30.42013, dan nilai MASE sebesar 0.674109. Nilai ACF1 pada set latihan adalah 0,0808969.

Sitasi: Nasution, B., Alfaris, L., Siagian, R. C., & Nasution, H. A. (2023). Analysis of Solar Flux and Sunspot Correlation Case Study: A Statistical Perspective, *Kappa Journal*. Vol. 7 No.1, 114-127.

April 2023. Vol. 7, No. 1
e-ISSN: 2549-2950
pp. 114-127

INTRODUCTION

The sun is an integral part of our solar system and plays an important role in shaping our planetary environment (Kane et al., 2021). Understanding the sun's behavior and patterns is critical to predicting impacts on our planet and its climate (Davies et al., 2021). One of the much-studied aspects of the sun is sunspots, which are dark spots that appear on the surface of the sun (Kjelseth, 2020). The number of sunspots has been shown to fluctuate over time and is related to solar activity, including solar flares and coronal mass ejections (Ansor et al., 2019). Solar flux, on the other hand, refers to the total amount of electromagnetic radiation emitted by the sun (Widén and Munkhammar, 2019). Like sunspots, solar flux also shows a correlation with solar activity and is an important indicator of solar energy output. By studying the relationship between sunspots and solar flux, we can gain a better understanding of the behavior of the sun and its impact on our planet.

In this study, we aim to analyze the statistical relationship between sunspots and solar flux from 2005. Using advanced statistical methods, we will examine the correlation between the two variables and explore the underlying patterns. By investigating the relationship between sunspots and solar flux, we hope to provide a deeper understanding of the behavior of the sun and its impact on our planet (Shiokawa and Georgieva, 2021). This study will provide insight into the correlation between sunspots and solar flux and the importance of this relationship for understanding solar behavior. The findings of this research will be important for the scientific community and for those interested in predicting the behavior of the sun and its impact on our planet

METHODS

2.1 Data collection

Data collection in this study is an important part to ensure that the analysis results are valid and can be accounted for. In this case, the data taken is based on open access data, namely from SILSO (*SILSO data/image, Royal Observatory of Belgium, Brussels*): https://www.sidc.be/silso/datafiles, and the website of the Canadian government: https://www.asc-csa.gc.ca/eng/sciences/sunspots2.asp. The data taken from these two sources is absolute real data from 2005, which contains information about solar flux and sunspots. Data collection conducted from these sources includes data relating to sunlight intensity, solar activity, and sunspot behavior. The following data samples were analyzed to test the correlation.

Table 1. Samples solar flux and sunspot data 2005

Parameter waktu			Data solar flux	Sunspot
No	**Year**	**Month**	**Solar flux number**	**sunspot number**
1	2005	2	85.45	43.5
2	2005	3	80.22	39.6
3	2005	4	77.9	38.7
4	2005	5	93.5	61.9
5	2005	6	87.36	56.8
6	2005	7	93.54	62.4
7	2005	8	90.6	60.5
8	2005	9	99.69	37.2
9	2005	10	68.6	13.2
10	2005	11	76.02	27.5

11	2005	12	79.21	59.3

Source: Goverment of canada and silso (2023)

The data collected is ensured that it meets the requirements needed to answer the research questions. This data will be analyzed and interpreted so that it can help explain the characteristics of the sun in 2005. After the data is collected, the data will be validated to ensure that the data is complete and accurate. This is done by cross-checking with other sources and conducting data completeness tests. After that, the data will be prepared to be used in statistical analysis

2.2 Descriptive statistics

Descriptive Statistics is a part of data analysis that aims to describe or provide information about the distribution of data and other properties of the data (Washington et al., 2020). In this study, Descriptive Statistics will be used to analyze solar flux data and the number of sunspots collected from 2005 to 2022. The Descriptive Statistics method that will be used in this study includes measuring data properties such as mean, median, mode, range, standard deviation, and skewness (Kaliyadan and Kulkarni, 2019). In addition, data visualization will also be carried out through histograms, box plots, and scatter plots to help understand data distribution and see any outliers (Avraam et al., 2021). The results of this Descriptive Statistics analysis will help answer research questions about the relationship between solar flux and sunspots (Maliniemi et al., 2019). Information obtained from Descriptive Statistics will be used as a basis for further analysis such as correlation and regression analysis (Mishra et al., 2020). By carrying out Descriptive Statistics, it is expected to provide an overview of the distribution and properties of solar flux and sunspot data for the period 2005, so that it can help understand the relationship between the two variables and obtain better conclusions.

2.3 Pearson Correlation Coefficient

Pearson's Correlation Coefficient is a statistical analysis method used to measure the strength of a linear relationship between two variables (Armstrong, 2019). Pearson's Correlation Coefficient values range from -1 to 1, with a value of 1 indicating a very strong positive linear relationship, a value of -1 indicating a very strong negative linear relationship, and a value of 0 indicating no linear relationship between the two variables (Tian et al., 2022). The equations is:

$$r=\frac{\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum\left(x_i-\bar{x}\right)^2\sum\left(y_i-\bar{y}\right)^2}} \tag{2.1}$$

Where r is correlation coefficient x_i, is values of the x-variable in a sample, $\bar{x}$ is mean of the values of the x-variable, y_i is values of the y-variable in a sample, $\bar{y}$ is mean of the values of the y-variable.

In this study, the Pearson's Correlation Coefficient will be used to measure the strength of the relationship between solar flux and the number of sunspots during the period 2005 (Mukhtar et al., 2021). By analyzing the Pearson's Correlation Coefficient, it is expected to obtain information about how strong the relationship is between the two variables and whether there is a linear relationship between solar flux and number of sunspots (Cao et al., 2022).

2.4 Analysis of the regression coefficient

Regression Analysis Coefficient is a statistical analysis method used to measure a linear relationship between one independent variable and one dependent variable (Shrestha, 2020). In the linear regression line, the equation is given by:

$$Y = b_0 + b_1X \tag{2.2}$$

Here b_0 is a constant and b_1 is the regression coefficient. The formula for the regression coefficient is given below. The observed data sets are given by x_i and y_i.

$$b_1 = \sum\left[\frac{(x_i - x)(y_i - y)}{(x_i - x)^2}\right] \tag{2.3}$$

Using available data and information, these equations can be applied and updated to explain the relationship between solar flux and the number of sunspots (Geryl and Alvestad, 2020). The equation obtained from the Regression Analysis Coefficient analysis will provide an overview of the linear relationship between the two variables and assist in making predictions about solar flux based on the number of sunspots (Nandy, 2021). In this study, the Regression Analysis Coefficient will be used to measure the linear relationship between the number of sunspots and solar flux from 2005. Data taken from open access sources such as SILSO and the Canadian government will be analyzed using the Regression Analysis Coefficient method to obtain the intercept and coefficient regression (McGranaghan et al., 2021). The results of the Regression Analysis Coefficient analysis will be used to find out how strong the influence of the number of sunspots is on the solar flux and whether there is a linear relationship between the two variables (Aparicio et al., 2020).

2.5 Time series analysis

Time Series Analysis is a statistical method used to study and understand the time relationship in a data. In Time Series Analysis, data is taken sequentially over a certain period of time and is used to predict future behavior. In this case, Time Series Analysis will analyze sunspot and solar flux data taken from 2005. This method will help to understand trends and time patterns present in the data, and predict future sunspot and solar flux values. To perform Time Series Analysis, data can be divided into certain parts, such as monthly data, annual data, or decadal data, depending on the time scale to be analyzed. Then, tools such as moving averages, exponential smoothing, or ARIMA (AutoRegressive Integrated Moving Average) can be used to create time series models that describe trends and patterns in the data. This model can then be used to predict future values by considering historical information and time trends

2.6 Multivariable Regression Models

The Multivariable Regression Model is a statistical research method used to examine the relationship between more than one independent variable and one dependent variable. In terms of sunspot and solar flux research, the Multivariable Regression Model can be used to determine how these two independent variables together affect the dependent variable to be studied. The Multivariable Regression Model will provide

April 2023. Vol. 7, No. 1
e-ISSN: 2549-2950
pp. 114-127

information about how the two independent variables affect the dependent variable and will also estimate the value of the dependent variable based on the value of the independent variable

2.7 ARIMA (*AutoRegressive Integrated Moving Average*) Model

The ARIMA *(AutoRegressive Integrated Moving Average)* method is a time series analysis method that aims to predict the value of a variable based on the previous value and to overcome the problem of non-stationarity in time series data. In this case, the ARIMA method can be applied to sunspot and solar flux research to predict the relationship between the two variables and see how these two variables change simultaneously in time. The ARIMA equation can be written as ARIMA (p, d, q) where p denotes the amount of autoregression, d denotes the amount of integration, and q denotes the amount of moving average. In this case, the ARIMA model will include information about the previous values of the two variables, namely sunspot and solar flux, as input for making predictions. Thus, the ARIMA method can be used to examine the relationship between sunspots and solar flux and make predictions about the development of these two variables over time. However, it is important to remember that determining the parameters p, d, and q in the ARIMA model requires careful analysis and considers the characteristics of the time series data to be analyzed.

2.8 Bayesian models

Bayesian Model Method is a statistical research method that uses Bayesian probability principles to model the relationship between variables (Hinne et al., 2020). Bayesian models use prior knowledge and data information to update the probabilities of a hypothesis, make parameter estimates and determine conditional probabilities (Rohmer, 2020). In sunspot and solar flux research, the Bayesian Model method can be used to model the relationship between the two variables and estimate the parameters of the relationship (Penza et al., 2022). The Bayesian model equation in this case will focus on determining the conditional probabilities of solar flux and sunspots based on existing data (Leka et al., 2019). Thus, the Bayesian Model method can be used as an alternative advanced research method to examine the relationship between solar flux and sunspots in 2005.

RESEARCH RESULTS

The graphical plot of the 2005 flux data is as follows:

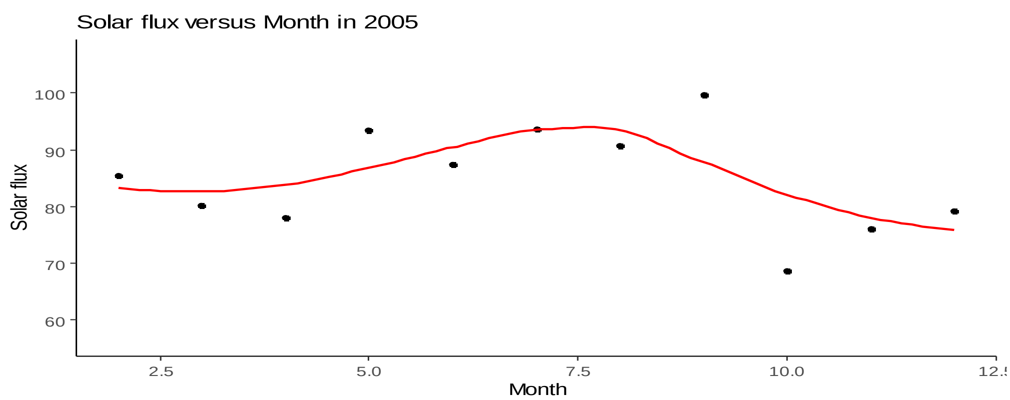

Figure 1. Smooth Graphical plot of the 2005 flux data
Source: Ruben Siagian in r program plot (2023)

Based on the descriptive statistical data of Solar Flux in 2005, there are several values that can describe the characteristics of the data. The Mean *(average)* value of this data is 84.73545, the Median value *(middle value)* is 85.45, and there is no Mode *(the most frequently occurring value)* in this data. The Standard Deviation value *(data spread)* of this data is 9.239713 and the Variance Value *(a measure of data spread)* of this data is 85.37229. The range *(the difference between the highest and lowest values)* of this data is 68.6 to 99.69, and the Quantile *(division of the data into equal parts)* of this data is 68.6, 78.555, 85.45, 92.05 and 99.69. The IQR value *(Interquartile Range or the difference between the 75% and 25% quantile values)* of this data is 13,495. Based on the calculation of the Pearson Correlation Coefficient, it can be seen that the correlation between the moon and the solar flux in 2005 was -0.1149093. This figure shows that there is a very weak relationship between the two variables and we cannot accurately predict solar flux based on the month.

Table 2. ARIMA (*Auto Regressive Integrated Moving Average*) model with parameters (0,0,0) and mean non-zero

Model ARIMA (*Auto Regressive Integrated Moving Average*) dengan parameter (0,0,0) dan mean non-zero	
Mean	85.288
Sigma^2	91.13
log likelihood	-36.22
AIC (*Akaike Information Criterion*)	76.45
AICc (*Corrected Akaike Information Criterion*)	78.16
BIC (*Bayesian Information Criterion*)	77.05

Source: Analysis data from author (2023)

The ARIMA *(Auto Regressive Integrated Moving Average)* model with parameters (0,0,0) and mean non-zero has been applied to the 2005 Solar Flux data. The mean value of this data is 85.2880 with a standard error of 2.8638. The value 2 *(variance)* of this data is 91.13. The log likelihood value of this model is -36.22 and the AIC *(Akaike Information Criterion)* value is 76.45, the AICc *(Corrected Akaike Information Criterion)* value is 78.16, and the BIC *(Bayesian Information Criterion)* value is 77.05.

Table 3. ARIMA model evaluation results

ARIMA model evaluation results	
Metric	**Value**
ME	-1.13688E-14
RMSE	9.056153
MAE	7.6824
MPE	-1.184356
MAPE	9.279572
MASE	0.8038786
ACF1	0.05453511

Source: Analysis data from author (2023)

After the ARIMA model is applied to the data, the ME *(Mean Error)* value is -1.136877e-14, the RMSE *(Root Mean Square Error)* value is 9.056153, the MAE *(Mean Absolute Error)* value is 7.6824, the

MPE *(Mean Percentage Error)* value is -1.184356, the MAPE *(Mean Absolute Percentage Error)* value is 9.279572, and the MASE *(Mean Absolute Scaled Error)* value is 0.8038786. The ACF1 *(Autocorrelation Function)* value of the training set is 0.05453511.

After the model is applied to the data, several evaluation values are obtained such as ME, RMSE, MAE, MPE, MAPE, MASE, and ACF1. The ME value of -1.136877e-14 indicates that the model has a very small mean error, so it can be said that the model has good performance. However, the RMSE value of 9.056153 indicates that the model has a fairly large root mean square error. The MAE value of 7.6824 indicates that the average absolute error of this model is quite large. MPE value of -1.184356 indicates that there is a small percentage error in the model. However, the MAPE value of 9.279572 indicates that there is a fairly large absolute percentage error in the model. The MASE value of 0.8038786 indicates that the model has a relatively low mean absolute scaled error. The ACF1 value of 0.05453511 indicates that the model has a very low autocorrelation function, so the model can be said to have good performance. Overall, the ARIMA model with parameters (0,0,0) and mean non-zero has good performance, although there are still some error values that need to be improved.

Nonlinear and multivariate regression analyzes have been performed on the sunspot number data. In the nonlinear regression analysis the results show that only parameter c has a significant value with a t value of 2.095 and a p-value of 0.0695. The residual standard error is 16.85 with 8 degrees of freedom. Meanwhile, in the multivariate regression analysis the results showed that only the intercept parameter had a significant value with a t value of 4.171 and a p-value of 0.00241. The Multiple R-squared value is 0.02913 and the Adjusted R-squared is -0.07875.

Based on Bayesian analysis, the results show that the average intercept is 87.6 with a standard deviation of 8.7. This means that in each month, the average solar flux is 87.6, and the range of solar flux in each month is around 8.7. The mean slope of the Month is -0.4 with a standard deviation of 1.2. This means that for each rising of the month, the average solar flux will decrease by 0.4, and the range of reducing the solar flux for each rising of the month is about 1.2. Sigma is 10.9 with a standard deviation of 2.9, this shows that the solar flux variation range is about 10.9 and the variation range is about 2.9. The fit diagnostic shows that the average meanppd is 85.2 with a standard deviation of 5.0. This shows that the average solar flux per month is 85.2 and the range of solar flux per month is around 5.0. The MCMC diagnostic shows that for each parameter, the Monte Carlo standard error (mcse) is a measure of the standard error of the MCMC result, the roughly effective sample size (neff) is the number of samples used in the MCMC process, and the potential scale-reduction factor in the separate chains (Rhat) is the convergence factor of the MCMC chain. The closer the Rhat value is to 1, the better the convergence of the MCMC chains.

The conclusion from the Bayesian analysis is that every month, the average solar flux is 87.6 with a standard deviation of 8.7. Each month increases, the average solar flux will decrease by 0.4 with a standard deviation of 1.2. The variation range of solar flux is around 10.9 with a standard deviation of 2.9. The results of the fit diagnostic show that the average monthly solar flux is 85.2 with a standard deviation of 5.0. The results of the diagnostic MCMC show that the closer the Rhat value is to 1, the better the convergence of the MCMC chains. The standard error size of the MCMC results is called the Monte Carlo standard error (mcse) and the roughly effective sample size (neff) is the number of samples used in the MCMC process
The graphical plot of the 2005 flux data is as follows:

April 2023. Vol. 7, No. 1
e-ISSN: 2549-2950
pp. 114-127

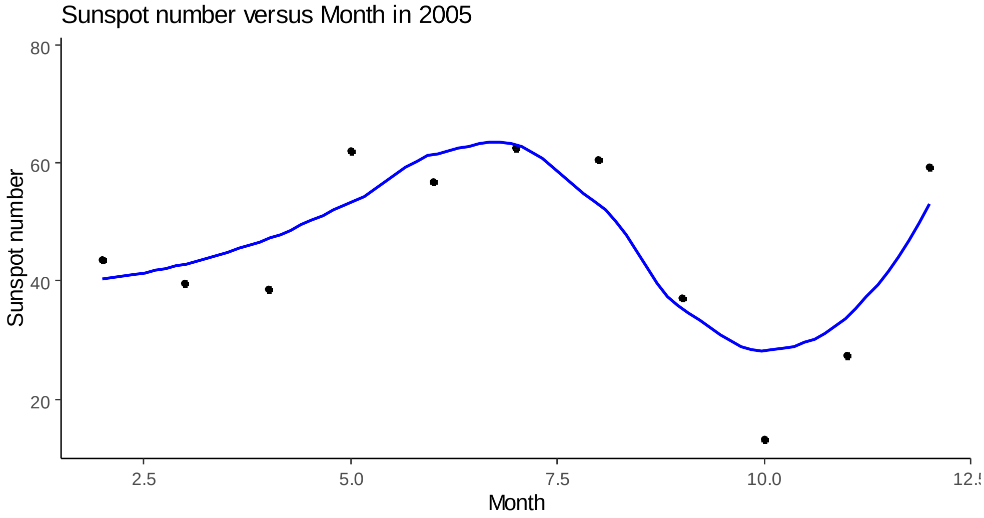

Figure 2. solarspots data in smooth Graphical plot of the 2005
Source: Ruben Siagian in r program plot (2023)

Based on existing sunspot data, statistical results show that the average number of sunspots is 45,509. The median number of sunspots is 43.5, which means that half of the sunspots have sunspots below this number and the other half have sunspots above this number. The standard deviation of the number of sunspots is 16,183, which shows how the number of sunspots differs from the average. The variance in the number of sunspots is 261,893, which is a measure of the variance of the sunspot data. The range of the number of sunspots is 13.2 to 62.4, which shows the difference between the maximum and minimum of the number of sunspots. The quantiles of the sunspot data are 27.5, 43.5, and 61.9, which show the distribution of the sunspot data. The IQR (Interquartile Range) of the number of sunspots is 21.95, which shows the difference between the quantiles at the top and bottom of the sunspot data.

For nonlinear regression analysis, the results of the analysis show that only parameter c has a significant value with a t value of 2.095 and a p-value of 0.0695. The residual standard error is 16.85 with 8 degrees of freedom. For multivariate regression analysis, the analysis shows that only the intercept parameter has a significant value with a t value of 4.171 and a p-value of 0.00241. The Multiple R-squared value is 0.02913 and the Adjusted R-squared is -0.07875.

In this sunspot analysis, we use ARIMA *(Auto Regressive Integrated Moving Average)* and Bayesian methods. The ARIMA model used is ARIMA (0,0,1) with a non-zero mean. The coefficient obtained is ma1 with a mean of 0.7351 and a standard error of 0.2590. The sigma^2 value is 181. The log likelihood obtained is -43.49, the AIC is 92.97, the AICc is 96.4, and the BIC is 94.16. After training, the error measure obtained was ME of 0.2615561, RMSE of 12.16969, MAE of 9.03306, MPE of -15.14689, and MAPE of 30.42013. Then, this model is also analyzed using Bayesian techniques using a sampling algorithm and 4000 observations in the sample. Priors used can be seen in help('prior_summary'). There are 2 predictors in this model and 11 observations. The results of the estimates show a mean (Intercept) of 51.3 with a standard deviation of 13.1, a mean Month of -0.8 with a standard deviation of 1.7, and a mean sigma of 18.1 with a standard deviation of 4.5. Diagnostic fit obtained is the mean mean_PPD of 45.5 with a standard deviation

of 7.9. MCMC Diagnostic shows the value of mcse, Rhat, and n_eff for each parameter. The value of mcse is the Monte Carlo standard error, n_eff is the effective size of the sample coarsely, and Rhat is the scaling factor on the split chain (at convergence Rhat=1).
As for the correlation graphic plot of flux and sunspot data in 2005 are as follows:

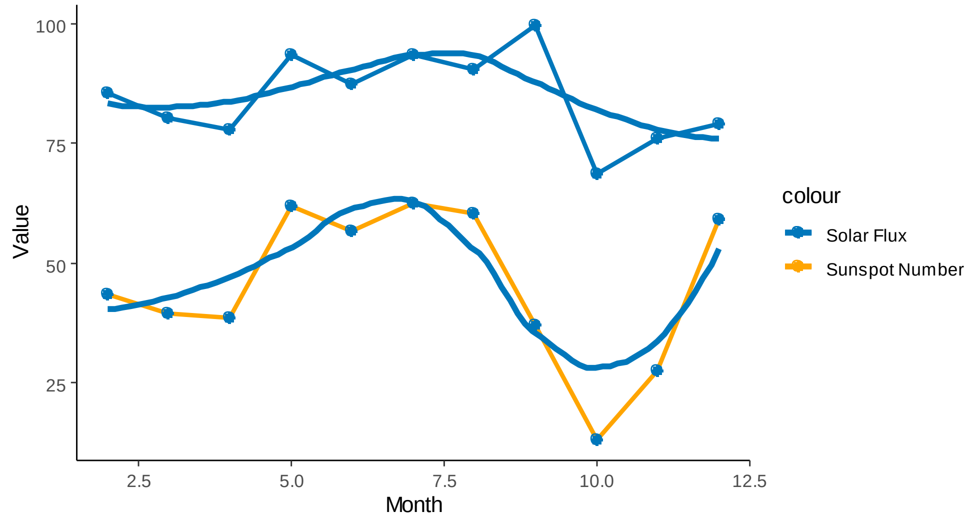

Figure 3. Graphical plot of the 2005 solarspot and solar flux correlation data
Source: Ruben Siagian in r program plot (2023)

This is an figure showing the solar flux number and sunspot number for 2005 in certain months. The solar flux number measures the intensity of radiation emitted from the Sun, while the sunspot number measures the number of sunspots on the Sun's surface. In this data, there are 11 observations in 2005 in months 2 to 12. For each observation, there are 2 line graphs showing the year and month of observation, the solar flux number, and the sunspot number. In addition, there is an additional graph showing the error (error) between the year and month of observation from the solar flux and sunspot data.

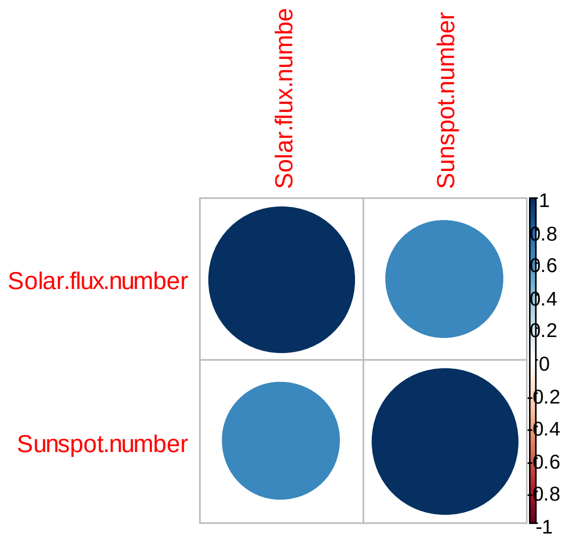

Figure 4. Corrplot of sunspot and sunpot correlations in 2005

Source: Ruben Siagian in r program plot (2023)

In nonlinear regression analysis, there are 4 weights used, namely b->h_1, i_1->h_1, b->o, and h_1->o. The initial value of this model is 25091.912932 and after the convergence process, the final value becomes 24410.580000. In the summary of the analysis results, it is stated that this is a 1-1-1 network with 4 weights and has the specified options, with a weight value of b->h_1 of 0.25 and i_1->h_1 of 0.37, and a weight value of b->o and h_1 ->o respectively 212.08 and 212.17.

The analysis shows the results of multivariate linear regression analysis between the parameters Solar Flux Number and Sunspot Number. From the results of the analysis, the coefficient *(Estimate)* of the Solar Flux Number is 1.1239, with a standard error of 0.4478. The t value of 2.510 indicates that the coefficient of the Solar Flux Number is significant at the level of 0.0333 *(meaning there is a 3.33% probability that changes in the Solar Flux Number do not affect the Sunspot Number)*. The F-statistic value is 6.301 with a p-value of 0.0333 indicating that the linear regression model has a good fit *(p-value less than 0.05)*. The R-squared value of 0.4118 indicates that the variation in the Sunspot Number is around 41.18% affected by the variation in the Solar Flux Number. However, because the Adjusted R-squared is only 0.3464, this shows that there are still variations in the Sunspot Number that cannot be explained by variations in the Solar Flux Number. While the residuals *(the difference between the observed value and the predicted value)* range from -25.1171 to 20.0012, with a median of 0.8736. The residual standard error is 13.08 at 9 degrees of freedom indicating that there are variations in the residuals that are not predicted by this linear regression model.

The ARIMA analysis results show that there is one MA coefficient *(Moving Average)* with a value of 0.7351 and a mean of 45.9542. Value s.e. *(standard error)* for the MA is 0.2590 and for the mean is 6.1550. The sigma value squared is 181 and the log likelihood is -43.49. The value of AIC *(Akaike Information Criteria)* is 92.97, AICc *(Akaike Corrected Information Criteria)* is 96.4, and BIC *(Bayesian Information Criteria)* is 94.16. The error results on the training set show that the ME *(Mean Error)* value is 0.2615561, the RMSE *(Root Mean Squared Error)* value is 12.16969, the MAE (Mean Absolute Error) value is 9.03306, the MPE (Mean Percentage Error) value is -15.14689, the MAPE value *(Mean Absolute Percentage Error)* is 30.42013, and the MASE *(Mean Absolute Scaled Error)* value is 0.674109. Finally, the value of ACF1 *(Autocorrelation Function 1)* in the training set is 0.0808969.

CONCLUSSION

4.1 Analysis of solar flux analysis 2005 every month

a. Mean (average) value: 84.73545
b. Median value (middle value): 85.45
c. No Mode (the most frequently occurring value)
d. Standard Deviation value (data spread): 9.239713
e. Variance Value (a measure of data spread): 85.37229
f. Range (the difference between the highest and lowest values): 68.6 to 99.69
g. Quantile (division of the data into equal parts): 68.6, 78.555, 85.45, 92.05, 99.69
h. IQR value (Interquartile Range or the difference between the 75% and 25%
i. quantile values): 13.495 Pearson Correlation

April 2023. Vol. 7, No. 1
e-ISSN: 2549-2950
pp. 114-127

j. Coefficient between the month and the solar flux in 2005: -0.1149093 (very weak relationship)

Based on available data, there is no mode that occurs frequently in solar flux in every month in 2005. The average value of solar flux in every month is 84.73545, while the median value is 85.45 *(middle value of the data)*. The standard deviation value indicates how far the data is spread and in this data the default is 9.239713. The variance value is a measure of the spread of data and in this data the increase is 85.37229. The range shows the difference between the highest and lowest values, namely 68.6 to 99.69. Quantile divides data into equal parts and in this data it is 68.6, 78.555, 85.45, 92.05, 99.69. The IQR is the difference between the 75% and 25% quantiles, and in this data it is 13,495. The Pearson Correlation coefficient between the moon and solar flux in 2005 was -0.1149093, indicating a very weak relationship between the two variables.

4.2 Correlation sunspot analysis 2005 every month

a. The mean sunspots were 45,509 sunspots, the median were 43.5 sunspots, the standard deviation was 16,183, the variance was 261,893, and the range was 13.2 to 62.4.
b. Non-linear regression analysis shows that only parameter c has a significant value with a t value of 2.095 and a p value of 0.0695.
c. The residual standard error is 16.85 with 8 degrees of freedom.
d. Multivariate regression analysis shows that only the intercept parameter has a significant value with a t value of 4.171 and a p value of 0.00241.
e. The R-squared value is 0.02913 and the adjusted R-squared is -0.07875.
f. The ARIMA model used is ARIMA (0,0,1) with non-zero mean.
g. coefficient obtained is ma1 with a mean of 0.7351 and a standard error of 0.2590.
h. The sigma^2 value is 181.
i. diagnostic fit result was a mean PPD of 45.5 with a standard deviation of 7.9.
j. Bayesian analysis shows the values of mcse, Rhat, and neff for each parameter.
k. The value of mcse is the Monte Carlo standard error, neff is the effective size of the sample roughly, and Rhat is the scaling factor in the split chain (at convergence Rhat=1).

The data provided describes various statistical analyses performed on the relationship between sunspot activity and the month of the year. It appears that the data was collected over a 12-month period in 2005, with the number of sunspots recorded each month. The data summary *(mean, median, standard deviation, variance, and range)* provides a general picture of the distribution of sunspot activity. The non-linear regression analysis indicates that there is a significant relationship between sunspots and one of the parameters (c), while the multivariate regression analysis suggests that only the intercept parameter is significant. The R-squared value indicates the proportion of variability in the response variable (sunspots) that can be explained by the predictor variable (month). The ARIMA model used provides an estimation of the future values of sunspots based on past values. The Bayesian analysis provides information on the accuracy of the model by showing the values of mcse, Rhat, and neff for each parameter.

April 2023. Vol. 7, No. 1
e-ISSN: 2549-2950
pp. 114-127

4.3 Solar flux and sunspot correlation analysis every month

a. The research examines the relationship between the solar flux number and sunspot number in 2005 using 11 observations in months 2 to 12.
b. The solar flux number measures the intensity of radiation emitted from the Sun, while the sunspot number measures the number of sunspots on the Sun's surface.
c. Nonlinear regression analysis was conducted using 4 weights (b->h_1, i_1->h_1, b->o, and h_1->o) with the final value of the model being 24410.580000.
d. Results of multivariate linear regression analysis between the Solar Flux Number and Sunspot Number showed that the coefficient of the Solar Flux Number was 1.1239 with a significant t value of 2.510 (3.33% probability of no effect on the Sunspot Number).
e. The linear regression model had a good fit with an F-statistic value of 6.301 and a p-value of 0.0333, with a R-squared value of 0.4118 indicating that 41.18% of the variation in the Sunspot Number was affected by the variation in the Solar Flux Number.
f. The adjusted R-squared value of 0.3464 showed that there were still variations in the Sunspot Number that could not be explained by variations in the Solar Flux Number.
g. ARIMA analysis results showed a MA coefficient of 0.7351 with a mean of 45.9542 and s.e. values of 0.2590 and 6.1550 respectively.
h. The AIC, AICc, and BIC values were 92.97, 96.4, and 94.16 respectively.
i. The error results on the training set showed a ME value of 0.2615561, RMSE value of 12.16969, MAE value of 9.03306, MPE value of -15.14689, MAPE value of 30.42013, and MASE value of 0.674109.
j. The ACF1 value in the training set was 0.0808969.

The results of this study show that there is a relationship between the amount of solar flux and the number of sunspots in 2005 which was analyzed using 11 observations in months 2 to 12. The solar flux number measures the intensity of radiation emitted from the Sun, while the sunspot number measures the number of sunspots on the surface of the Sun. This study uses multivariable linear regression analysis to analyze the relationship between Solar Flux Number and Sunspot Number. The results of the analysis show that the coefficient of the Solar Flux Number is 1.1239 with a significant t value of 2.510 (the probability that there is no effect on the Sunspot Number is 3.33%). The linear regression model has good results with an F-statistical value of 6.301 and a p-value of 0.0333, with an R-squared value of 0.4118 which indicates that 41.18% of the variation in the Sunspot Number is influenced by the variation in the Solar Flux Number. The adjusted R-squared value of 0.3464 indicates that there are variations in the Sunspot Number that cannot be explained by variations in the Solar Flux Number. The results of the ARIMA analysis show that there is an MA coefficient of 0.7351 with a mean of 45.9542 and a value of s.e. of 0.2590 and 6.1550. The AIC, AICc, and BIC values are 92.97, 96.4, and 94.16. The error results in the training set show that the ME value is 0.2615561, RMSE is 12.16969, MAE is 9.03306, MPE is -15.14689, MAPE is 30.42013, and MASE is 0.674109. The ACF1 value in the training set is 0.0808969.

REFFERENCE

Ansor, N.M., Hamidi, Z., Shariff, N., 2019. The impact on climate change due to the effect of global electromagnetic waves of solar flare and coronal mass ejections (CMEs) phenomena. Presented at the Journal of Physics: Conference Series, IOP Publishing, p. 012019.

Aparicio, A., Gallego, M., Antón, M., Vaquero, J., 2020. Relationship between solar activity and direct solar irradiance in Madrid (1910–1929). Atmospheric Res. 235, 104766.

Armstrong, R.A., 2019. Should Pearson's correlation coefficient be avoided? Ophthalmic Physiol. Opt. 39, 316–327.

Avraam, D., Wilson, R., Butters, O., Burton, T., Nicolaides, C., Jones, E., Boyd, A., Burton, P., 2021. Privacy preserving data visualizations. EPJ Data Sci. 10, 2.

Cao, B., Feng, J., An, L., 2022. Long-term relationship between foF2 from China ionospheric stations and solar activity during the 24th solar activity cycle. Radio Sci. 57, 1–11.

Davies, S.J., Abiem, I., Salim, K.A., Aguilar, S., Allen, D., Alonso, A., Anderson-Teixeira, K., Andrade, A., Arellano, G., Ashton, P.S., 2021. ForestGEO: Understanding forest diversity and dynamics through a global observatory network. Biol. Conserv. 253, 108907.

Geryl, P., Alvestad, J., 2020. A formula for the start of a new sunspot cycle: Relating the adjusted 10.7 cm solar flux to the start of a new sunspot cycle. Astrophys. Space Sci. 365, 113.

Hinne, M., Gronau, Q.F., van den Bergh, D., Wagenmakers, E.-J., 2020. A conceptual introduction to Bayesian model averaging. Adv. Methods Pract. Psychol. Sci. 3, 200–215.

Kaliyadan, F., Kulkarni, V., 2019. Types of variables, descriptive statistics, and sample size. Indian Dermatol. Online J. 10, 82.

Kane, S.R., Arney, G.N., Byrne, P.K., Dalba, P.A., Desch, S.J., Horner, J., Izenberg, N.R., Mandt, K.E., Meadows, V.S., Quick, L.C., 2021. The fundamental connections between the Solar System and exoplanetary science.

Kjelseth, I.M.E., 2020. Intensity Contrast as a Function of Underlying Magnetic Field Through the Solar Atmosphere.

Leka, K., Park, S.-H., Kusano, K., Andries, J., Barnes, G., Bingham, S., Bloomfield, D.S., McCloskey, A.E., Delouille, V., Falconer, D., 2019. A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems. Astrophys. J. Suppl. Ser. 243, 36.

Maliniemi, V., Asikainen, T., Salminen, A., Mursula, K., 2019. Assessing North Atlantic winter climate response to geomagnetic activity and solar irradiance variability. Q. J. R. Meteorol. Soc. 145, 3780–3789.

McGranaghan, R.M., Ziegler, J., Bloch, T., Hatch, S., Camporeale, E., Lynch, K., Owens, M., Gjerloev, J., Zhang, B., Skone, S., 2021. Toward a next generation particle precipitation model: Mesoscale prediction through machine learning (a case study and framework for progress). Space Weather 19, e2020SW002684.

Mishra, H., Garofalo, G., Giordano, A.M., Ott, C., 2020. On the dynamics of floating-base robots: Linking the recursive formulation to the Reduced Euler-Lagrange Equations. Preprint 10.

Mukhtar, M., Jushua, B., Bonde, D., Kaoje, M., Asabe, B., Sulaiman, M., 2021. The relationship between sunspots number and some lower atmospheric parameters at solar minimum of solar cycle 24. Int. Res. J. Mod. Eng. Technol. Sci. 3, 228–233.

Nandy, D., 2021. Progress in solar cycle predictions: Sunspot cycles 24–25 in perspective: Invited review. Sol. Phys. 296, 54.

Penza, V., Berrilli, F., Bertello, L., Cantoresi, M., Criscuoli, S., Giobbi, P., 2022. Total Solar Irradiance during the Last Five Centuries. Astrophys. J. 937, 84.

Rohmer, J., 2020. Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: A comprehensive review. Eng. Appl. Artif. Intell. 88, 103384.

Shiokawa, K., Georgieva, K., 2021. A review of the SCOSTEP's 5-year scientific program VarSITI—Variability of the Sun and Its Terrestrial Impact. Prog. Earth Planet. Sci. 8, 1–43.

Shrestha, N., 2020. Detecting multicollinearity in regression analysis. Am. J. Appl. Math. Stat. 8, 39–42.

Tian, L., Oviatt, S., Muszynski, M., Chamberlain, B., Healey, J., Sano, A., 2022. Applied Affective Computing.

Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

Widén, J., Munkhammar, J., 2019. Solar radiation theory. Uppsala University.