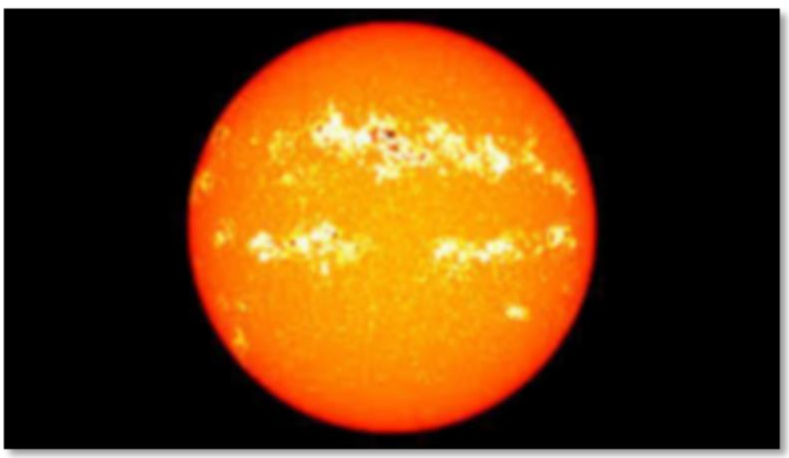
Sunspots



TIME SERIES PROJECT 2024

DR.RASHA ELSODA

________Table of content _________

Introduction ……………………………………………………….2

Data analysis ………………………………………………......4

- Data description ……………………………………………4
- TS graphs ………………………………………………..5
- Stationarity …………………………………………………6
- ACF & PACF ………………………………………………7
- Box & Jenkins Analysis……………………………………10
- Final model & diagnosis …………………………………..11
- Forecasting ………………………………………………..12

Conclusion ………………………………………………………..13

Reference ……………………………………………….………..13

Introduction

Sunspots are temporary phenomena on the Sun's photosphere that appear as dark spots compared to surrounding areas. They are caused by magnetic activity and typically occur in pairs or groups. Sunspots are cooler regions on the Sun's surface, but they are still incredibly hot compared to terrestrial temperatures. We are interested in this topic because it is an interesting phenomenon to study which is a natural phenomenon with a specific records and variables.

- **<u>Why to study sunspots??</u>**

1- **Understanding Solar Dynamics:** By studying sunspots and their behavior, scientists can gain insights into the complex dynamics of the Sun, including its magnetic field structure, energy output, and internal processes. This understanding is important for advancing our knowledge of solar physics and how it impacts the Earth and the solar system.

2- **Space Weather Prediction:** Sunspots are closely related to solar flares, coronal mass ejections (CMEs), and other solar phenomena that can affect space weather. These events can release large amounts of energy and particles into space, potentially impacting satellite operations, communication systems, power grids, and even astronaut safety. By forecasting sunspot activity, scientists and space agencies can better predict and prepare for space weather events, mitigating their potential impacts on Earth and space-based infrastructure.

3- **Climate Studies:** Some research suggests that solar activity, including sunspot cycles, may influence Earth's climate patterns over long periods. By understanding sunspot behavior and its correlation with climate variability, scientists can improve climate models and predictions, helping us better understand and adapt to changes in Earth's climate.

4- **Technological Implications:** Sunspot activity can induce geomagnetic storms on Earth, which, in turn, can disrupt power grids, GPS systems, and communication networks. By forecasting sunspot activity, engineers and operators of critical infrastructure can take preventive measures to minimize the impact of geomagnetic disturbances, ensuring the reliability and resilience of technological systems.

5- **Space Exploration and Satellite Operations:** For space missions and satellite operations, accurate forecasts of space weather, driven by sunspot activity, are essential for ensuring the safety of spacecraft and astronauts and optimizing mission planning. By predicting periods of heightened solar activity, space agencies can schedule critical operations and take precautions to safeguard spacecraft and equipment.

- **Literature review:**

Historically, sunspot observations date back to ancient civilizations, with notable records from Chinese astronomers as early as 800 BCE. However, systematic sunspot observations began in the early 17th century with the advent of the telescope, notably by astronomers like Galileo Galilei.

The study of sunspots gained momentum in the 19th century with the pioneering work of astronomers like Richard Carrington and Johann Rudolf Wolf, who developed systematic methods for recording and cataloging sunspot data. Their efforts led to the creation of the sunspot number, a quantitative measure of sunspot activity that remains a key metric in solar physics.

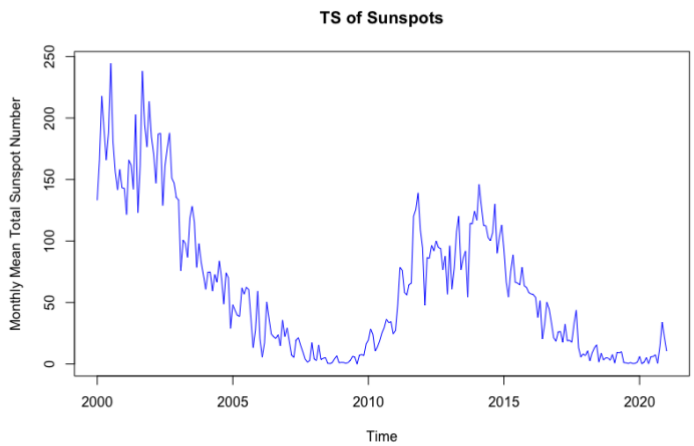
Modern advancements in observational techniques, such as space-based telescopes and solar satellites, have revolutionized our understanding of sunspots and solar activity. Data-driven approaches, including machine learning and statistical modeling, have become increasingly prevalent in analyzing sunspot datasets to uncover patterns, predict solar activity, and mitigate potential impacts on Earth's technology infrastructure.

Data analysis:

- ***<u>Data description:</u>***

1. Source of the Data: The sunspot data in this dataset originates from the Royal Observatory of Belgium's Sunspot Index and Long-term Solar Observations (SILSO) project. SILSO is a long-term research initiative dedicated to monitoring and recording sunspot activity using historical observations and modern telescopic data.
2. Variable Measured: The primary variable measured in the dataset is the monthly mean sunspot number, which represents the mean count of sunspots observed on the Sun's surface during each month.
3. Scale Unit: The scale unit for the sunspot numbers is simply the count of sunspots observed within a particular month which measurement scale is ratio.
4. Sample Size: The dataset contains a total of 3235 monthly observations, covering the period from January 1749 to January 2021. Each observation includes the month and year of the observation, along with the corresponding sunspot number recorded for that month (our variable). However, we only cover the recent period from January 2000 to January 2021 for better visualizations.

➢ ***T.S Plot*:**



This graph is showing the monthly mean total sunspots number from year 2000 till 2020,

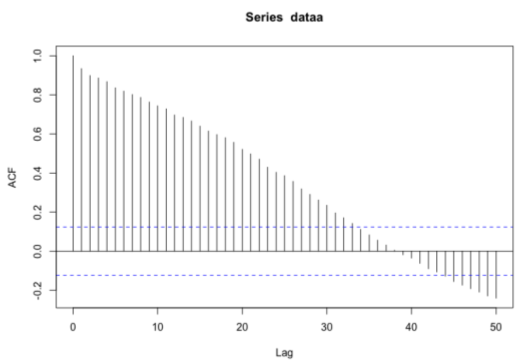
(2000-2010): The graph commences with a notable peak in sunspot activity around the year 2000, with the monthly mean total reaching approximately 250 sunspots. This peak is indicative of heightened solar activity at the onset of the millennium. However, following this peak, there is a discernible downward trend in sunspot numbers over the next decade, indicating a gradual decline in solar activity during this period.

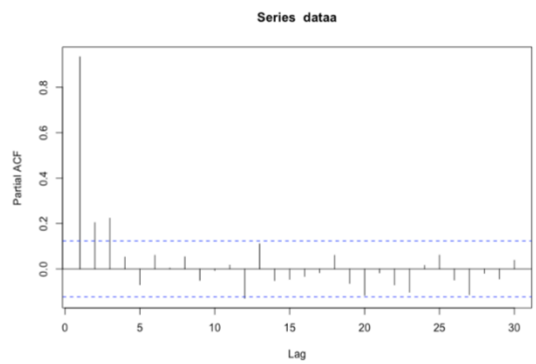
(2010-2015): Around the midpoint of the observed period there is a gradual recovery in sunspot activity. Although the increase is evident, it appears to be modest compared to the preceding decline.

(2015-2020): Following the brief resurgence in solar activity, the graph depicts another downturn in sunspot numbers from around 2015 onwards. This decline continues until the end of the observed period in 2020.

Throughout the entire time span, there is notable variability and fluctuations in sunspot numbers from month to month. Regarding seasonality, we conducted Kruskal Wallis test and we got p-value=0.564>0.1, which indicates no seasonality across the period.

- ***Stationarity***

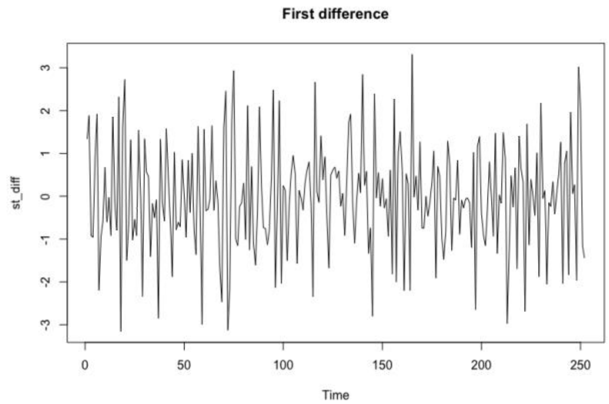




- Both ACF and PACF plots show slowly decreasing patterns ‘’PACF have sin wave pattern which is also decreasing slowly’’, it suggests that the data may not be stationary.
- Also, time series plot itself shows no stationarity in mean ‘’ it is decreasing and increasing which shows some negative and positive trends ‘’ and in variance ‘’it shows a fluctuation, and we have both decreasing and increasing variances.”

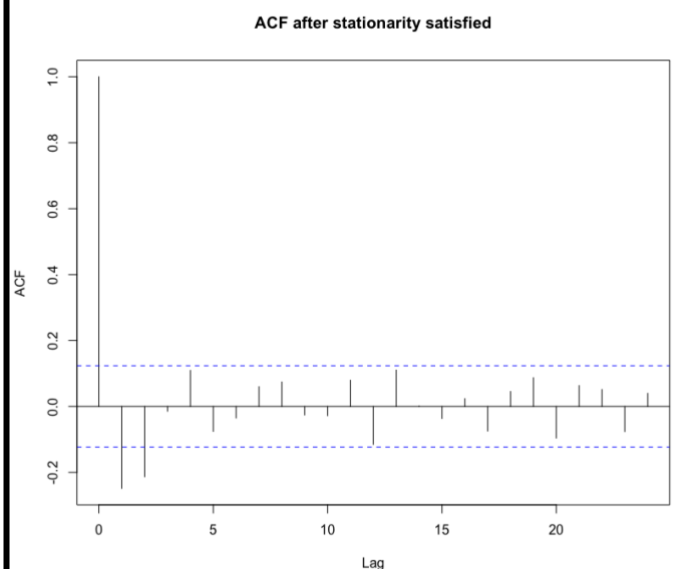
So, to turn data to stationary in mean, we transform the data by a square root we will also take the first difference ‘'Δ Y(T) ‘’ to achieve stationarity in mean.

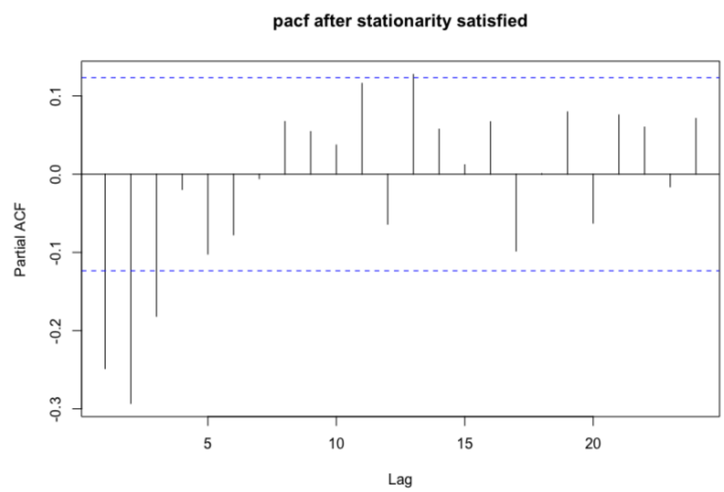
➢ The following graph shows the data after the first difference:



It seems that our data now is stationary, and it shows no trends or high fluctuations.

➢ ***ACF & PACF:***





ACF:

- As all ACFs' it starts at lag (0) with r (0) = 1.
- It cuts off after the 2^{nd} lag which refers to a model of MA (2).
- This implies that there is a significant autocorrelation at lag 1 and lag 2, but no significant autocorrelation beyond lag 2.

PACF:

- Here the plot starts with the first lag.
- It cuts off after the 3rd lag which refers to a model of AR (3).
- The ACF plot shows significant autocorrelation at lags 1, 2, and 3, but no significant autocorrelation beyond lag 3. This suggests that there is a pattern in the data that repeats every 1, 2, and 3 time points.

➢ Based on the ACF & PACF and that we've taken the first difference we suggest to start with the model ARIMA (3,1,2).

Initial model

ARIMA (3,1,2)

$$\boxed{(1-\phi 1B-\phi 2B2-\phi 3B3)(1-B)Yt=(1+\theta 1B+\theta 2B2)\varepsilon t}$$

```
z test of coefficients:

     Estimate Std. Error z value Pr(>|z|)
ar1 -0.558578   0.331815 -1.6834  0.09230 .
ar2 -0.183871   0.183409 -1.0025  0.31609
ar3 -0.190351   0.095632 -1.9904  0.04654 *
ma1  0.165280   0.332583  0.4970  0.61922
ma2 -0.267394   0.214409 -1.2471  0.21235
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After checking coefficients of our model we found that some estimators are insignificant, therefore, this model is not appropriate for our data, so we tried model (3,1,0) because estimators Φ are insignificant.

ARIMA (3,1,0)

- All assumptions were satisfied in this model but for simplicity we will use ARIMA (2,1,0).

Final model

ARIMA (2,1,0)

$$\boxed{(1-\phi 1B-\phi 2B2)(1-B)Yt=\varepsilon t}$$

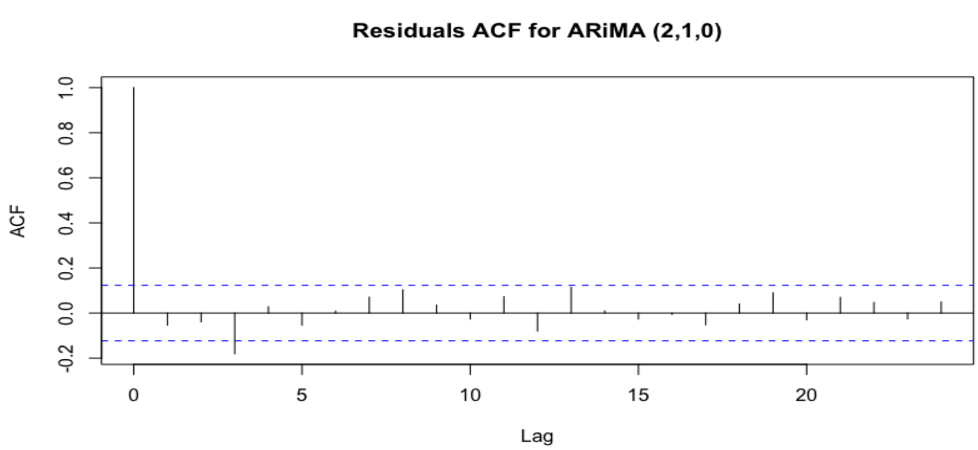
- **the significance of the coefficients**

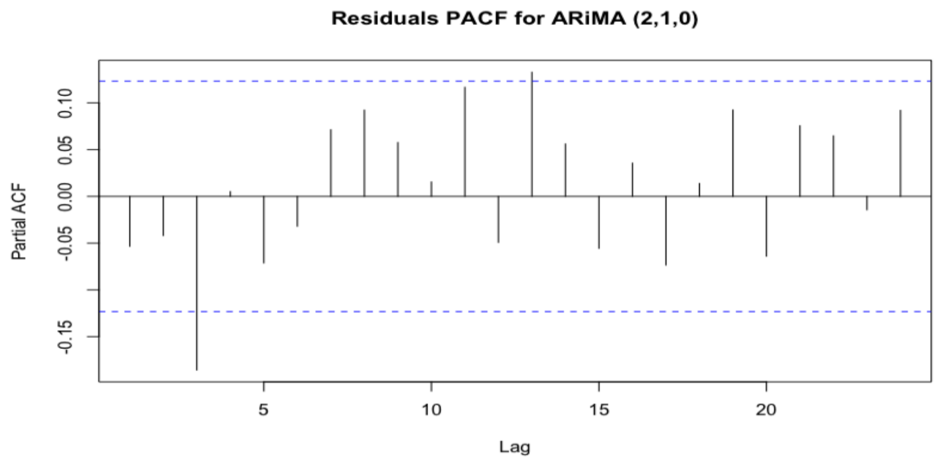
```
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1 -0.327701   0.060535 -5.4134 6.182e-08 ***
ar2 -0.301182   0.060686 -4.9630 6.941e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From z test of coefficient, all estimates are significant. So we will use this model .

- **The randomness of the noise:**





⇒ The residuals ACF & PACF we conclude that; all autocorrelation values are within the confidence bounds so all the residuals are insignificant.

⇒ Since the ACF and PACF plots of the residuals do not show any significant autocorrelation or partial autocorrelation at any lag, we can confidently conclude that the residuals behave like white noise.

➢ **Box-ljung & Box-Pierce test:**

- As shown from R results, the p-values of both Box-Ljung test and Box-Pierce test is greater than 0.05 so we reject the null hypothesis and conclude that the residuals are insignificant.

```
        Box-Ljung test

data:  Residuals
X-squared = 15.129, df = 8, p-value = 0.05669

> Box.test(Residuals,type="Box-Pierce",lag=10,fitdf=2)

        Box-Pierce test

data:  Residuals
X-squared = 14.744, df = 8, p-value = 0.06432
```

- **the accuracy measures :**

```
> accuracy(Model2)
                      ME     RMSE       MAE  MPE MAPE      MASE         ACF1
Training set -0.05953632 1.16426 0.9135558 -Inf  Inf 0.9111256 -0.001528532
> accuracy(Model3)
                      ME      RMSE      MAE  MPE MAPE      MASE       ACF1
Training set -0.04971135 1.184733 0.937492 -Inf  Inf 0.9349982 -0.0533968
```

We checked the accuracy measure for both models " ARIMA (3,1,0) & ARIMA (2,1,0) " but there is a slight difference between them so we will continue using the model ARIMA (2,1,0).

- **stationarity and invertibility:**

as the model we choose is ARIMA (2,1,0), there is no MA part to check stationarity and invertibility for, but for the AR part; it is always invertible with no conditions and stationary under some conditions .

```
Coefficients:
          ar1      ar2
      -0.3277  -0.3012
s.e.   0.0605   0.0607

sigma^2 estimated as 1.409:  log likelihood = -400.92,  aic = 807.84

Training set error measures:
                      ME     RMSE      MAE  MPE MAPE      MASE       ACF1
Training set -0.04971135 1.184733 0.937492 -Inf  Inf 0.9349982 -0.0533968
>
```

$\boxed{\phi 1+=\phi 2<1}$ **-0.3277+-0.3012=-0.6289 <1**

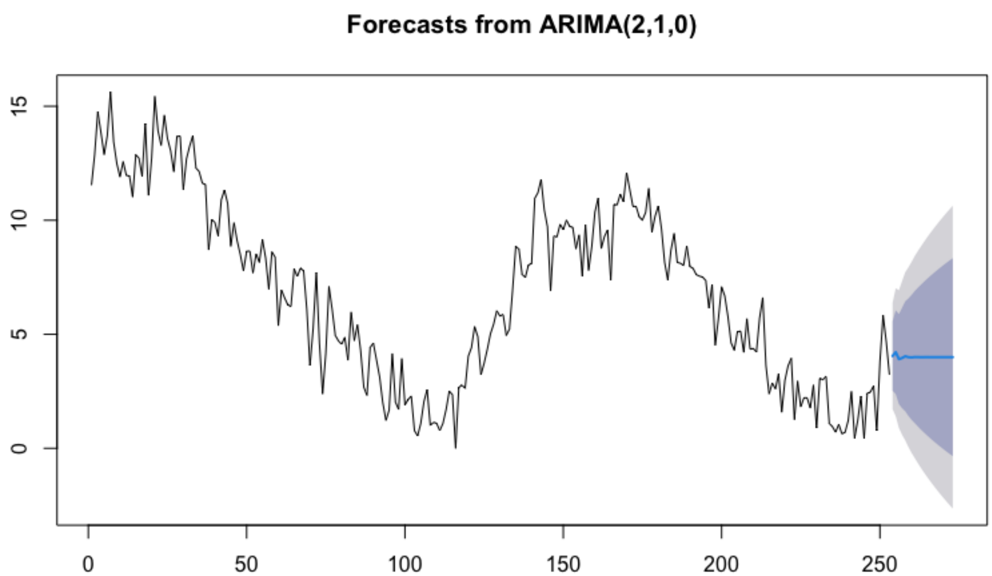
$\boxed{\phi 2 - \phi 1=<1}$ **-0.3012--0.3277=0.0265<1**

$\boxed{-1<=\phi 2<1}$ **-1<-0.3012 <1**

Therefore , it is stationary.

➢ Forecasting:

$$(1+0.3277B+0.3012B2)(1-B)Yt=\varepsilon t$$



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
254	4.048097	2.52679052	5.569403	1.72145975	6.374733
255	4.213286	2.38013692	6.046434	1.40972661	7.016845
256	3.911222	1.93882509	5.883618	0.89470150	6.927742
257	3.960456	1.76036186	6.160551	0.59570220	7.325210
258	4.035298	1.62573293	6.444864	0.35018601	7.720411
259	3.995944	1.42177531	6.570112	0.05909278	7.932795
260	3.986299	1.24978025	6.722818	-0.19884543	8.171444
261	4.001313	1.10647051	6.896155	-0.42596640	8.428592
262	3.999298	0.95805718	7.040538	-0.65187825	8.650473
263	3.995436	0.81475925	7.176113	-0.86898941	8.859862
264	3.997308	0.68196463	7.312652	-1.07307240	9.067689
265	3.997858	0.55338801	7.442328	-1.27000422	9.265720
266	3.997114	0.42840376	7.565824	-1.46075731	9.454985
267	3.997192	0.30827640	7.686108	-1.64451772	9.638902
268	3.997391	0.19204996	7.802731	-1.82237573	9.817157
269	3.997302	0.07904076	7.915563	-1.99516149	9.989766
270	3.997271	-0.03075691	8.025300	-2.16306627	10.157609
271	3.997308	-0.13758403	8.132200	-2.32646371	10.321080
272	3.997305	-0.24175138	8.236362	-2.48577244	10.480383
273	3.997295	-0.34342633	8.338017	-2.64126549	10.635856

The forecasted range falls within a relatively narrow band of 3.91 to 4.2, indicating an expected stability in the future behavior of the sunspot activity.

Conclusion:

The initial ARIMA(3,1,0) model was refined to a final ARIMA(2,1,0) model, indicating that the inclusion of two autoregressive terms better captures the underlying dynamics of the time series data. Specifically, the autoregressive coefficients (ϕ1$\phi 1$ and ϕ2$\phi 2$) were estimated to be -0.3277 and -0.3012, respectively. this narrow range signifies an anticipated stability in sunspot behavior, holding significant implications across various domains. From space weather prediction and climate analysis to agricultural planning, telecommunications infrastructure, and scientific research, these insights empower decision-makers to mitigate potential disruptions and leverage the predictability of solar variability for societal and technological resilience.

Reference :

1-The data source link: https://www.kaggle.com/datasets/robervalt/sunspots

2-Database from SIDC - Solar Influences Data Analysis Center - the solar physics research department of the Royal Observatory of Belgium. SIDC website