Prediction of Sunspot using Time-series models

Time Series Analysis Term Project

2024.5.29

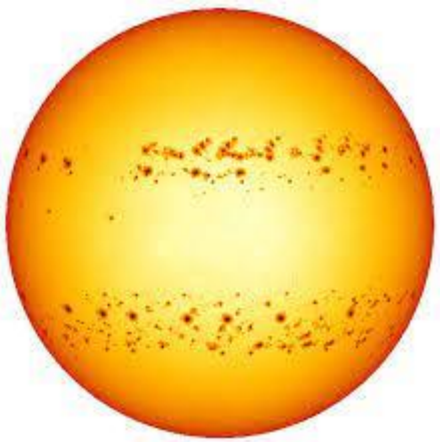
배기웅, 도경근

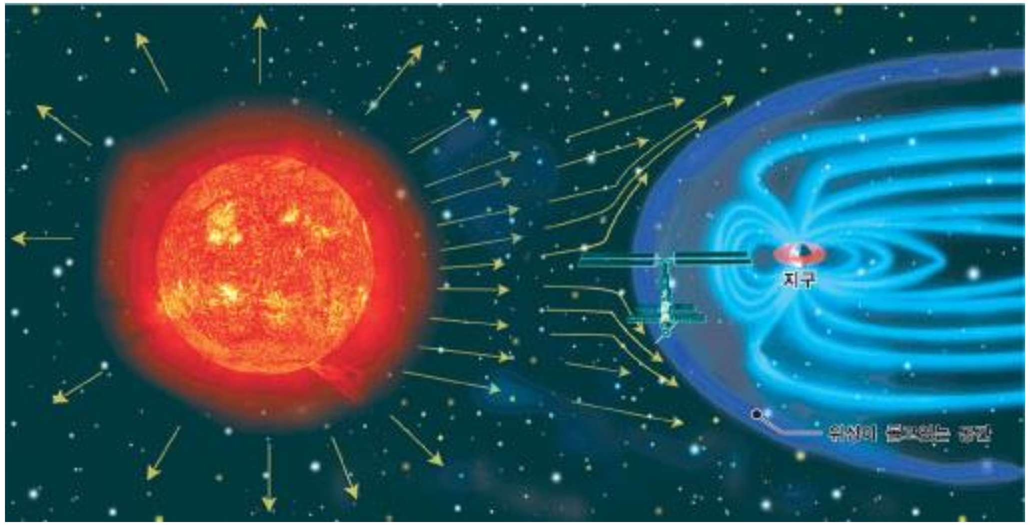
Contents

1. Introduction

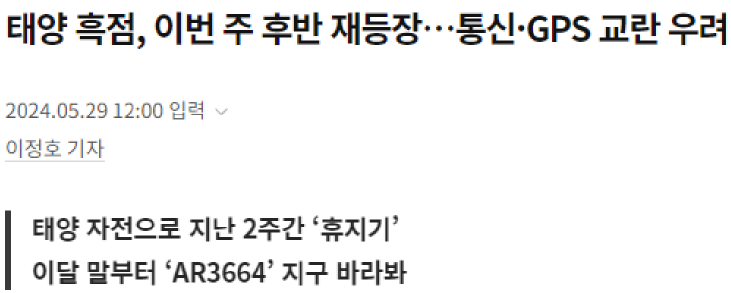
- **Objective of the project**
  - Develop predictive models using historical sunspot data to forecast future numbers and sizes of sunspots.
  - Enhance the accuracy of solar activity predictions, contributing to the fields of space weather, climate science, and astrophysical research.

- **Definition of Sunspot**
  - Temporary phenomena on the Sun's photosphere that appear as spots darker than the surrounding areas
  - Caused by intense magnetic activity, associated with solar magnetic storm activities





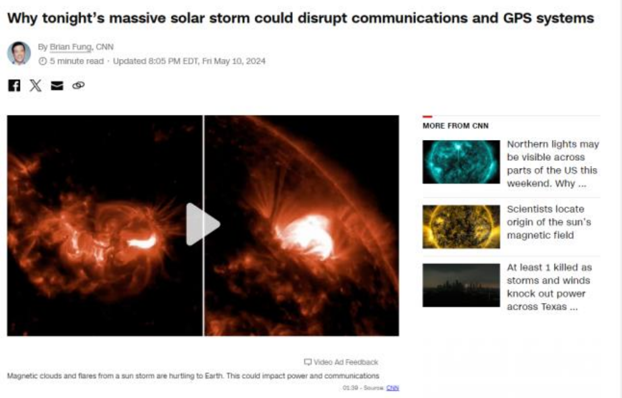
1. Introduction


태양 흑점, 이번 주 후반 재등장…통신·GPS 교란 우려

2024.05.29 12:00 입력

이정호 기자

태양 자전으로 지난 2주간 '휴지기'
이달 말부터 'AR3664' 지구 바라봐


Why tonight's massive solar storm could disrupt communications and GPS systems

By Brian Fung, CNN
5 minute read · Updated 8:05 PM EDT, Fri May 10, 2024

MORE FROM CNN

Northern lights may be visible across parts of the US this weekend. Why ...

Scientists locate origin of the sun's magnetic field

At least 1 killed as storms and winds knock out power across Texas ...

Video Ad Feedback

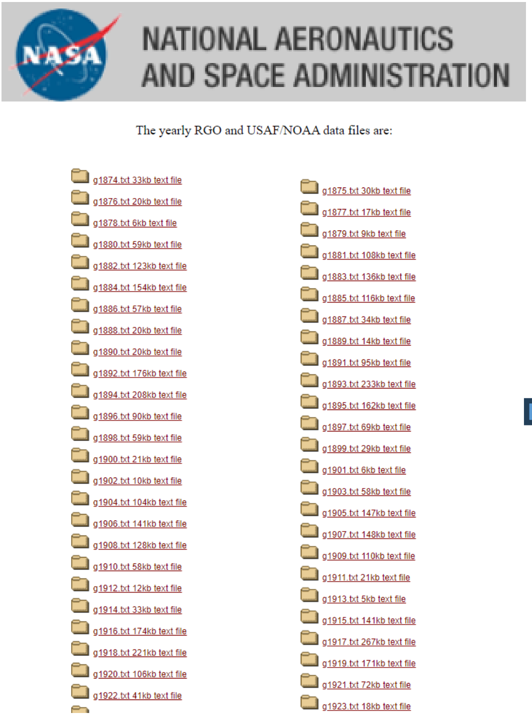
Magnetic clouds and flares from a sun storm are hurtling to Earth. This could impact power and communications

01:39 - Source: CNN

- **Importance of sunspot prediction**
  - Space Weather Forecasting
    - Sunspots are the origin of the significant solar flares and coronal mass ejections (CME), which can lead to geomagnetic storms that often affect satellite operations, GPS systems, and even ground-based technologies and power grids. Accurate predictions of sunspot activities enable better preparation and mitigation strategies against these disturbances.
  - Climate Research
    - Understanding and predicting sunspots contribute to climate science as variations in solar output associated with sunspot activity can influence Earth's climate patterns. Long-term changes in sunspot numbers have been correlated with Earth's temperature variations, such as the Little Ice Age.
  - Solar Physics Insights
    - Studying sunspot patterns helps scientists understand the solar dynamo mechanism that drives the entire sunspot cycle. This understanding is crucial for building models of stellar magnetic activity, which is fundamental in the broader context of astrophysics.

1. Introduction (Data Description)

- **USAF/NOAA Sunspot Data - Sunspot Average size, Sunspot count**
  - By. NASA, Greenwich Royal observatory
  - Data period: May 1874 ~ October 2016
  - https://solarscience.msfc.nasa.gov/greenwch.shtml



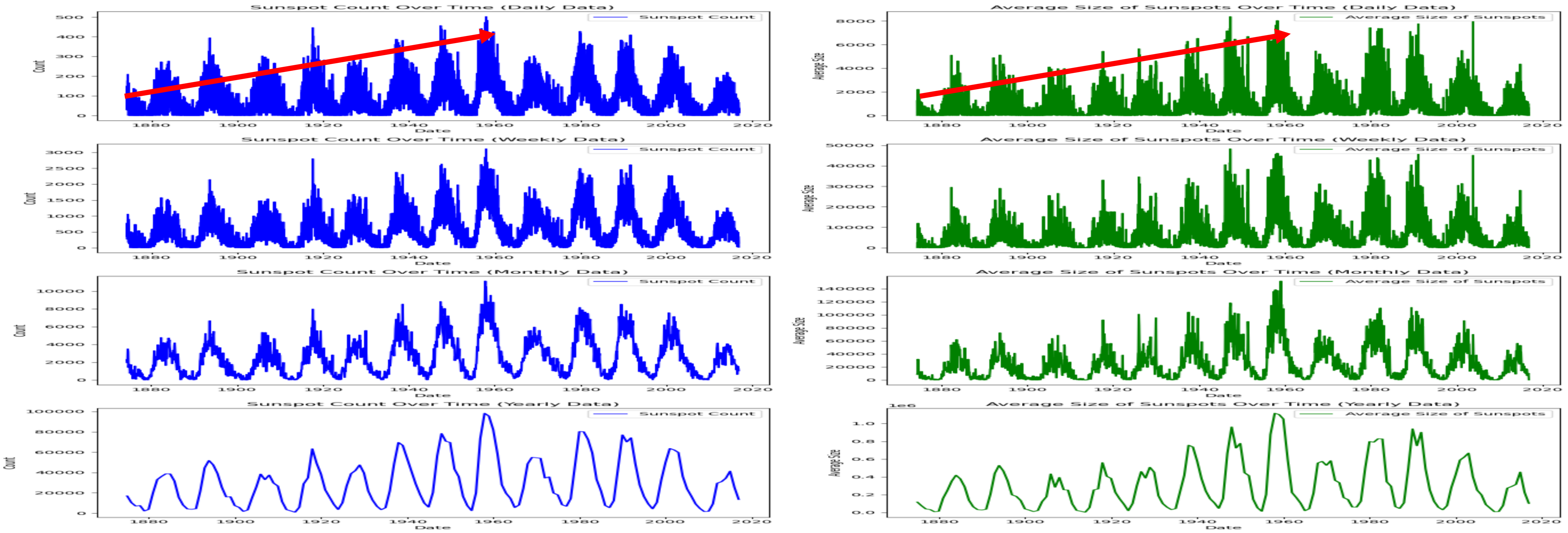
Preprocessing

- Data transformation
- Missing value Imputation
- Data integration

	date	count	Average_size
100	1874-08-09	125	979.0
101	1874-08-10	87	1812.0
102	1874-08-11	127	1948.0
103	1874-08-12	102	1557.0
104	1874-08-13	107	1564.0
...	...	...	...
51995	2016-09-08	53	560.0
51996	2016-09-09	79	602.0
51997	2016-09-10	78	630.0
51998	2016-09-11	67	476.0
51999	2016-09-12	59	406.0

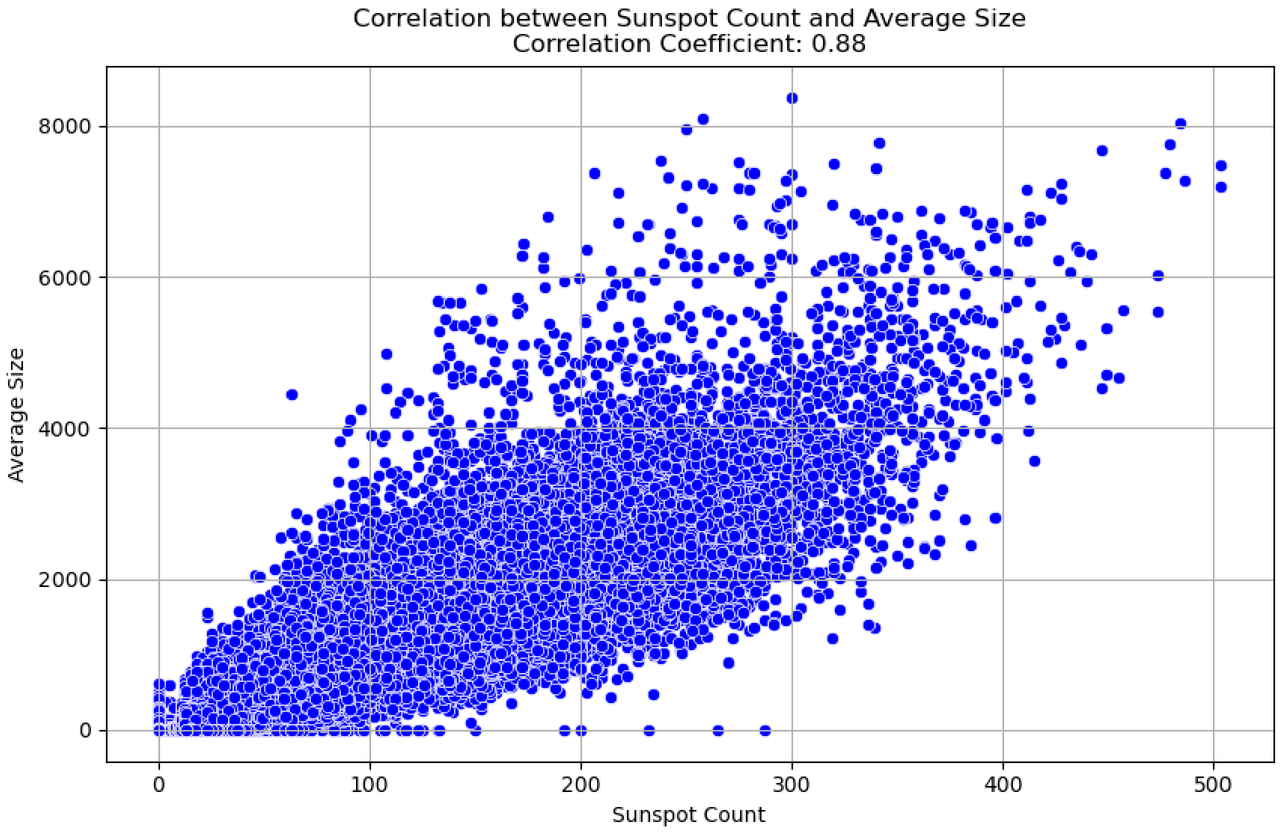
2. Data analytics

- **EDA**
  - Shows a seasonal pattern of period of 11 years
  - Shows a slightly upward trend until 1950s and after that turns around to a downward trend



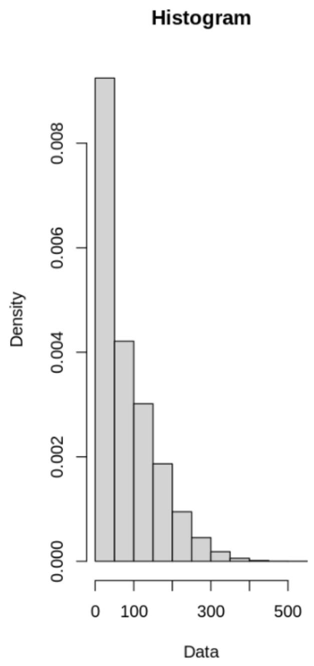
2. Data analytics

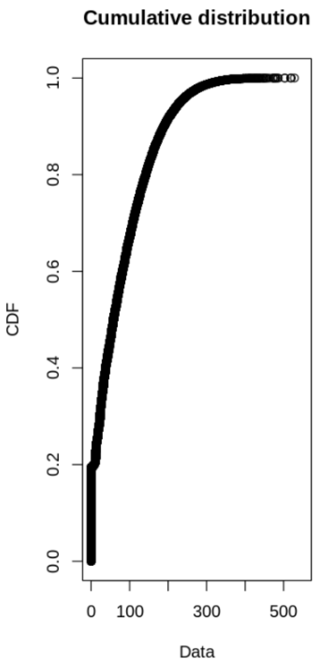
- **Correlation between sunspot count and average size**
  - Correlation coefficient : 0.88
  - Strong correlation between sunspot count and average size exists



2. Data analytics

- **Estimation of Distribution (Frequency of sunspot)**
  - With Poisson distribution and non-negative binomial distribution, we estimated the frequency distribution.



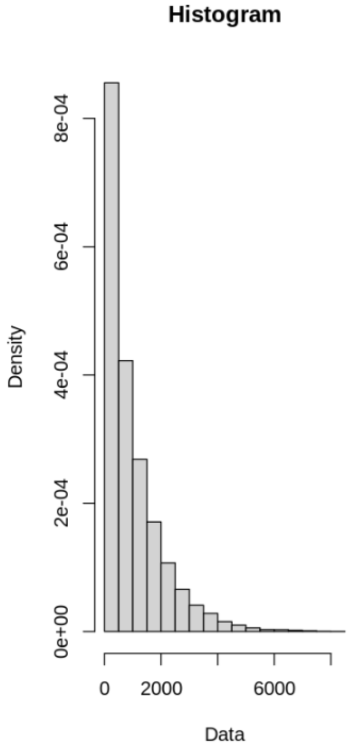


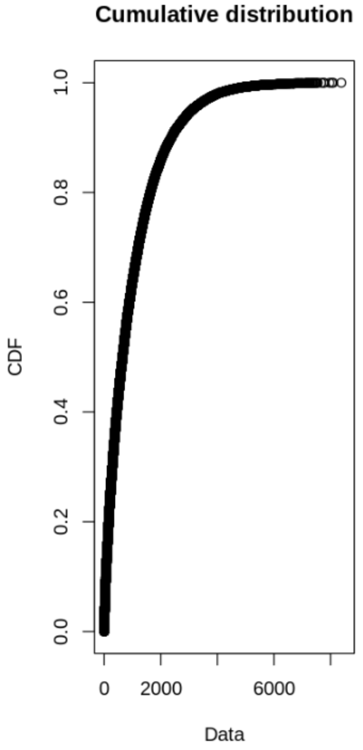

```
Chi-squared statistic:  Inf 38208.36
Degree of freedom of the Chi-squared distribution:  93 92
Chi-squared p-value:  0 0

 Goodness-of-fit criteria
                                  pois   nbinom
Akaike's Information Criterion 6199372 784827.6
Bayesian Information Criterion 6199381 784846.0
```

  - The frequency of sunspot follows a non-negative binomial distribution

2. Data analytics

- **Estimation of Distribution (Size of sunspot)**
  - With the continuous probability distributions, we estimated the size (severity) distribution.





'Exponential 로그우도: ' '-349028.09990711'

'Gamma 로그우도: ' '-348441.215993977'

'Log-normal 로그우도: ' '-352690.555525963'

'Log-logistic 로그우도: ' '-352227.57541394'

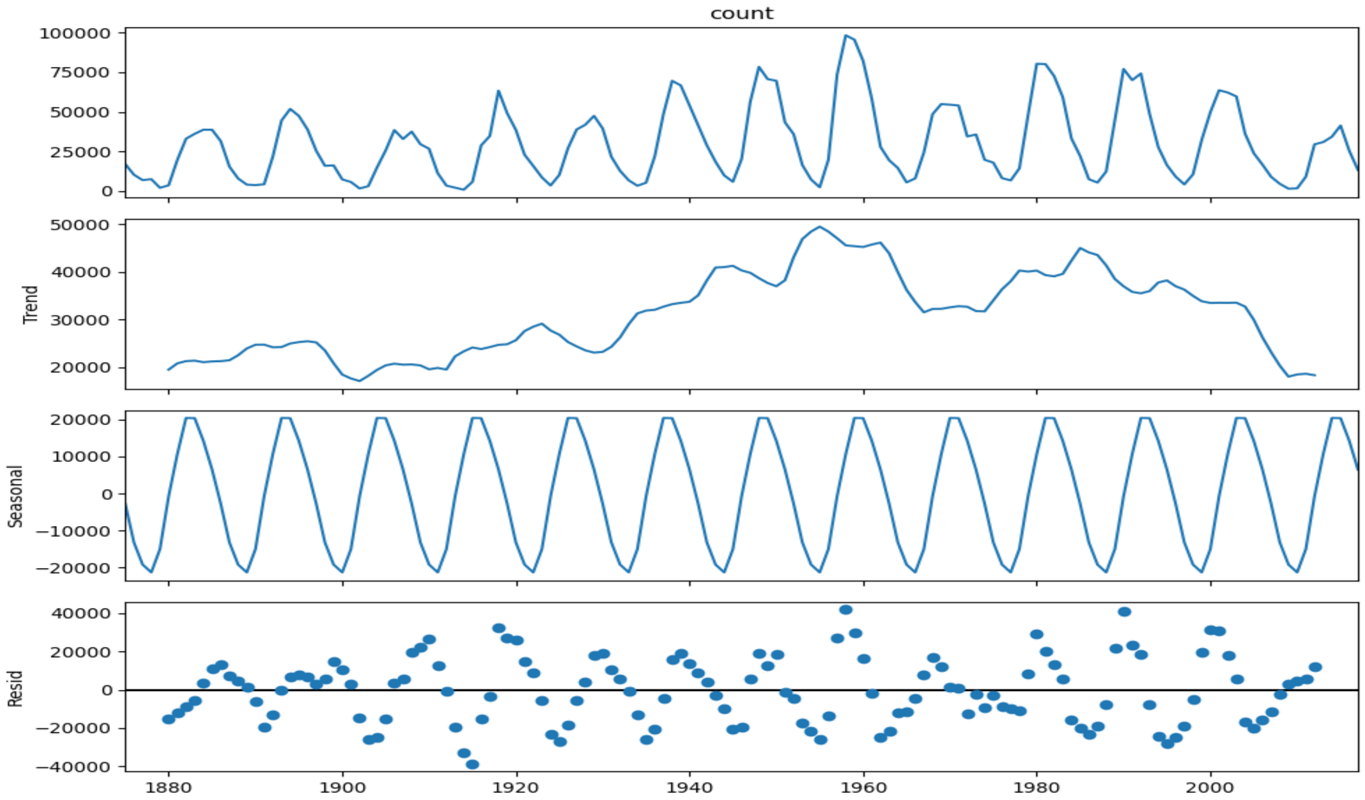
'Weibull 로그우도: ' '-348523.248616827'

'Skew-normal 로그우도: ' '-353558.537108571'

  - The severity of sunspot follows a Gamma distribution

3. Sunspot count

- **Sunspot count**
  - Trend: shows a slightly upward trend until 1950s and after that turns around to a downward trend
  - Seasonal: shows periodicity of 11 years
  - Errors: residuals graph shows points distributed randomly in sunspot counts



POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

3.1 Normality test

- **Normality**
  - Plot a graph of 142 years of wildfires and see a repeat ev ery 11 years.
  - ADF-test and KPSS-test with 90% significance level

  - ADF test H_0 : Time series is non-stationary
  - KPSS test H_0 : Time series is stationary

  - Ordinary data : non-stationary
  - 11-years differential data : stationary

original

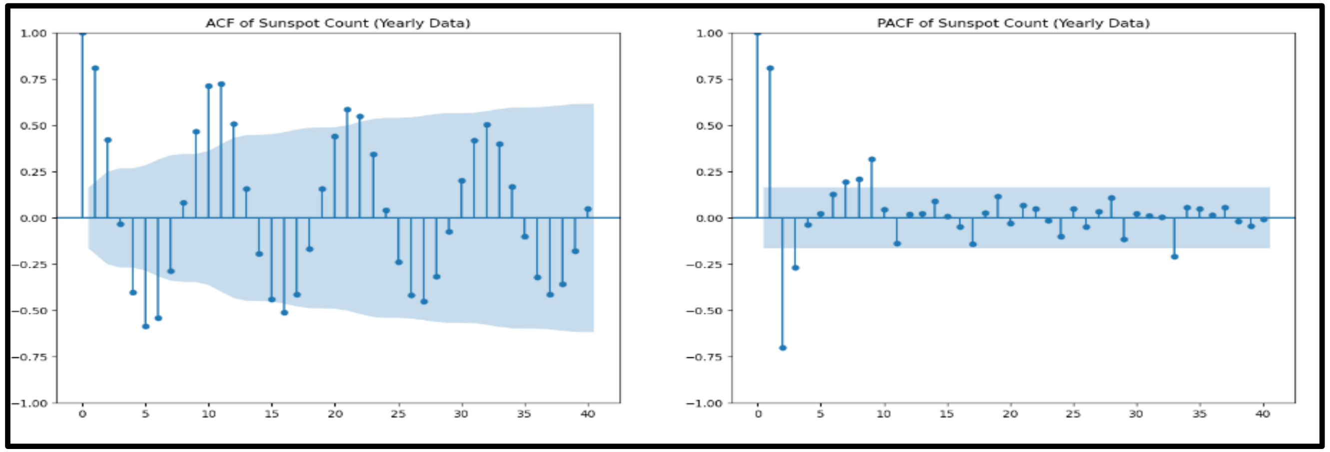
```
#### Results of Dickey-Fuller Test ####
Test-Statistic                   -1.722885
p-value                           0.419362
#Lags Used                        8.000000
Number of Observations used     134.000000
Critical Value (1%)              -3.480119
Critical Value (5%)              -2.883362
Critical Value (10%)             -2.578407
dtype: float64
------------------------------------------------------------
#### Results of KPSS Test ####
KPSS Statistic: 0.3843234434215026
p-value: 0.08391230887004199
num lags: 5
Critical values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
10%:0.347
5%:0.463
2.5%:0.574
1%:0.739
```

11-years differential data

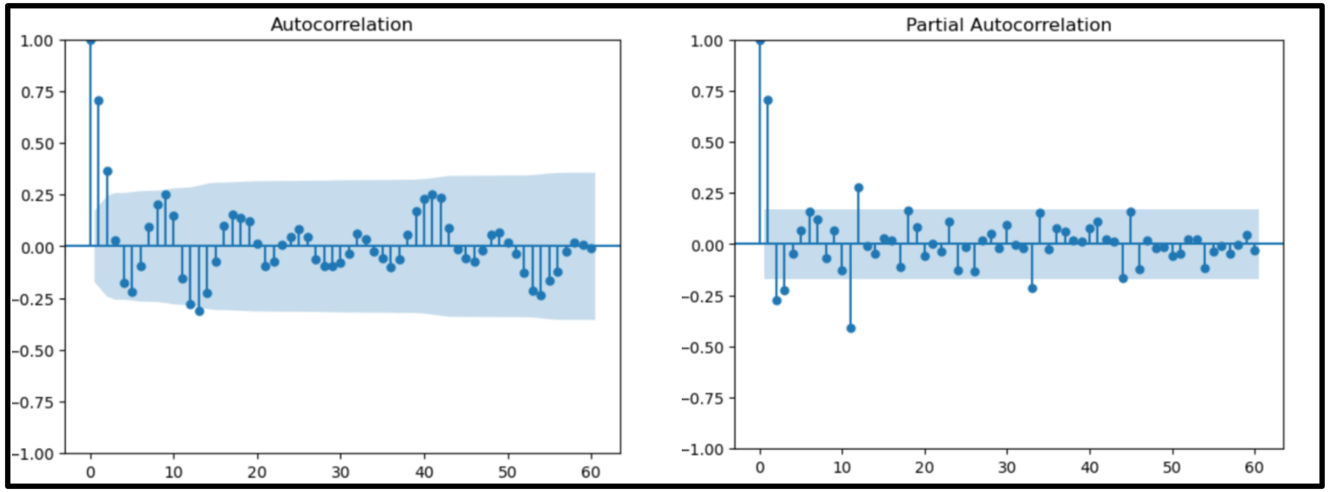
```
#### Results of Dickey-Fuller Test ####
Test-Statistic                   -2.738863
p-value                           0.067551
#Lags Used                       11.000000
Number of Observations used     119.000000
Critical Value (1%)              -3.486535
Critical Value (5%)              -2.886151
Critical Value (10%)             -2.579896
dtype: float64
------------------------------------------------------------
#### Results of KPSS Test ####
KPSS Statistic: 0.2939900891472019
p-value: 0.1
num lags: 5
Critical values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
10%:0.347
5%:0.463
2.5%:0.574
1%:0.739
```

3.2 ACF, PACF of original data(yearly)

- **Sunspot count**
  - Original data
    - ACF decaying slowly
    - PACF peaks at 4 lags



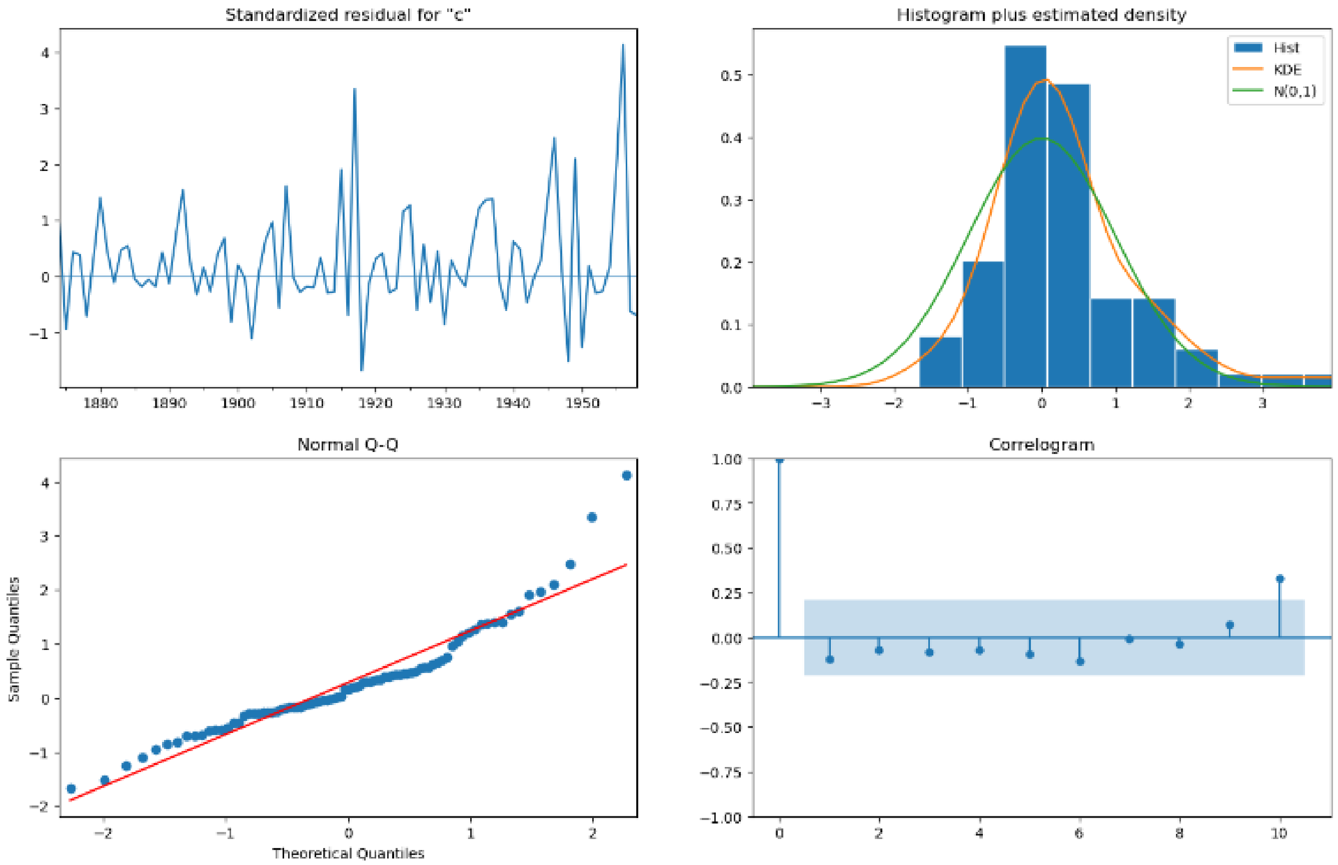
  - Differencing to remove seasonality
    - Cut off after 3 lag & Cutoff after 2 lag



3.3 Time series prediction on count of sunspot

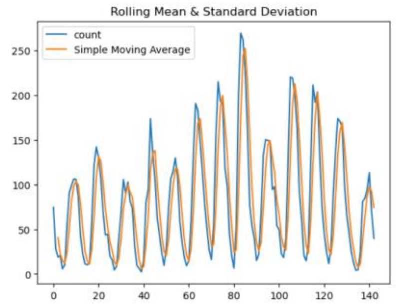
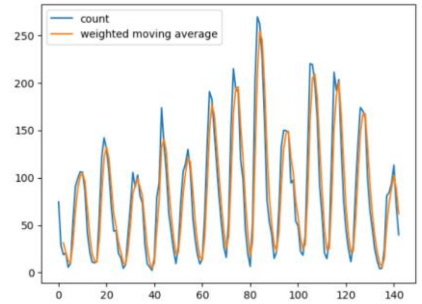
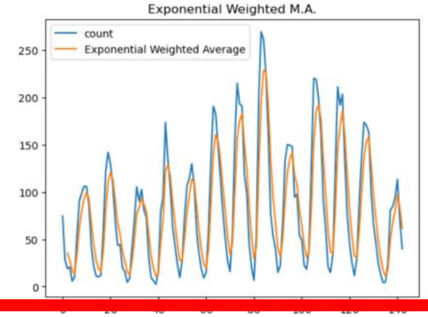
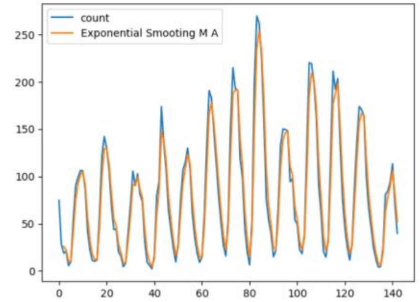
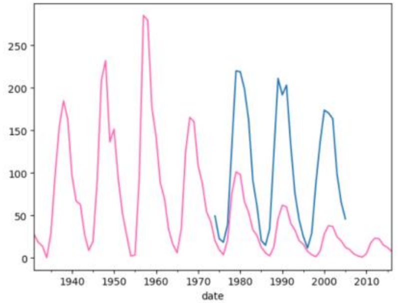
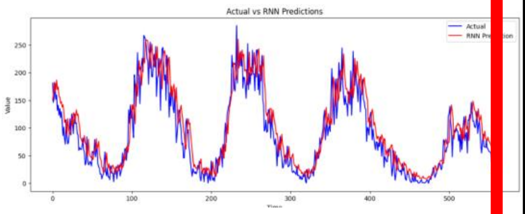
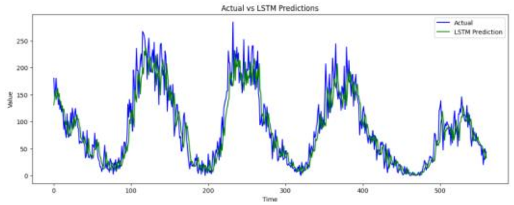
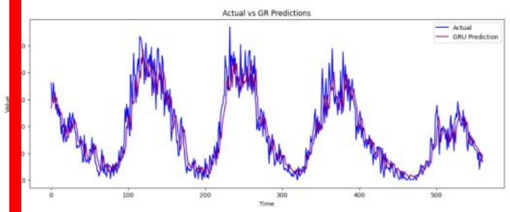
- **ARIMA (1, 0, 2)x(1, 0, 0, 11)**
  - $(1 - 0.5979B)(1 - 0.6120B^{11})Z_t = (1 - 0.6131B - 0.5012B^2)a_t$

```
best_model SARIMA
nonseasonal (1, 0, 2), seasonal (1, 0, 0, 11)
                                     SARIMAX Results
==================================================================================
=======
Dep. Variable:                             count   No. Observations:
85
Model:              SARIMAX(1, 0, 2)x(1, 0, [], 11)   Log Likelihood              -
400.750
Date:                           Wed, 29 May 2024   AIC
811.501
Time:                                   12:34:58   BIC
823.714
Sample:                               01-01-1874   HQIC
816.413
                                    - 01-01-1958
Covariance Type:                             opg
============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------
ar.L1          0.5978      0.160      3.740      0.000       0.284       0.911
ma.L1          0.6131      0.116      5.266      0.000       0.385       0.841
ma.L2          0.5012      0.139      3.614      0.000       0.229       0.773
ar.S.L11       0.6120      0.174      3.517      0.000       0.271       0.953
sigma2       668.9715     83.141      8.046      0.000     506.017     831.926
=================================================================================
Ljung-Box (L1) (Q):                   1.35   Jarque-Bera (JB):                48.94
Prob(Q):                              0.25   Prob(JB):                         0.00
Heteroskedasticity (H):               4.34   Skew:                             1.22
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.80
=================================================================================
```



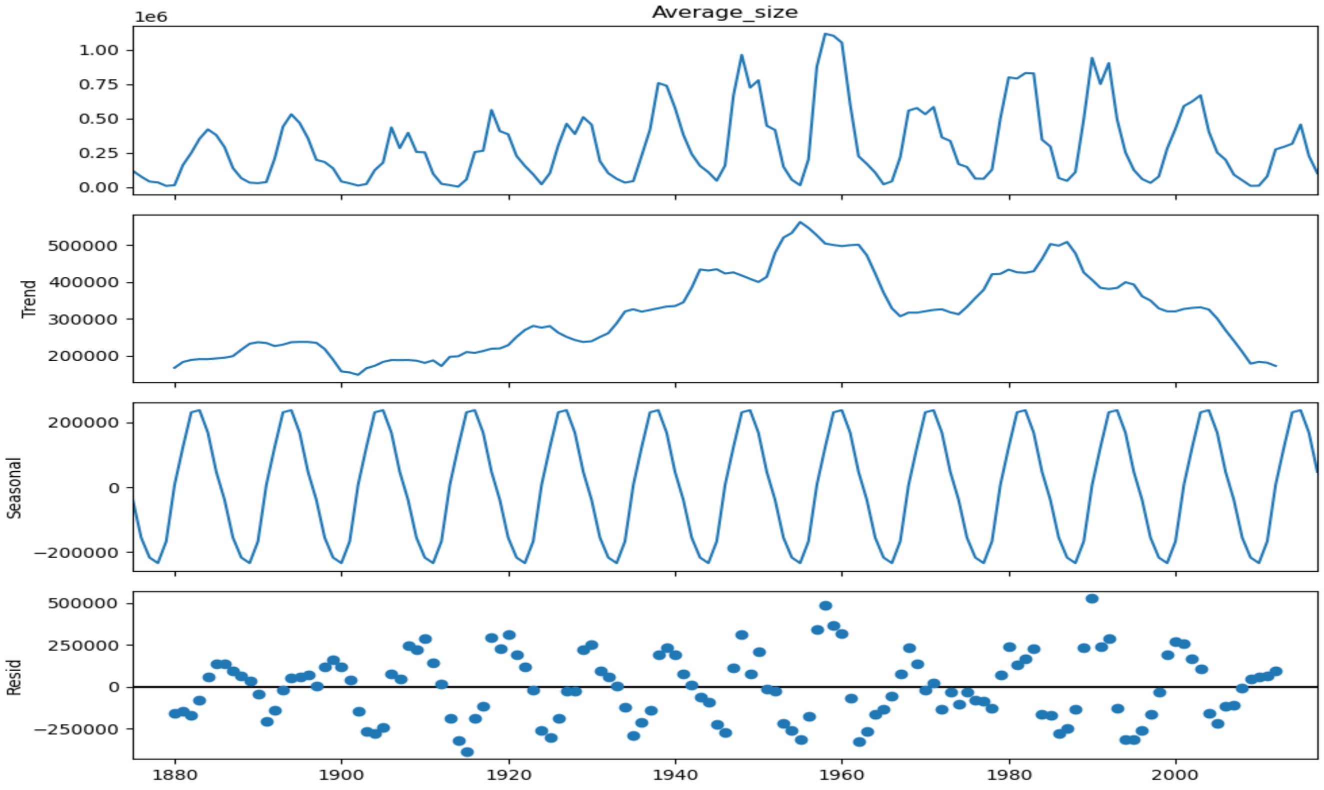
3.3 Time series prediction on count of sunspot

- **Forecast the count time-series dataset**

<table>
<tr><th>Time series</th><th>Simple MA</th><th>Weighted MA</th><th>Exponential MA</th><th>Exponential SMA</th></tr>
<tr><td>RMSE</td><td>415.15</td><td>281.99</td><td>323.85</td><td>168.74</td></tr>
<tr><td>Prediction result</td><td>
</td><td>
</td><td>
</td><td>
</td></tr>
<tr><th>Time series</th><th>ARIMA (1, 0, 2) x (1, 0, 0, 11)</th><th>RNN</th><th>LSTM</th><th>GRU</th></tr>
<tr><td>RMSE</td><td>60.11</td><td>26.45</td><td>24.33</td><td>24.48</td></tr>
<tr><td>Prediction result</td><td>
</td><td>
</td><td>
</td><td>
</td></tr>
</table>

4. Sunspot Size

- **Sunspot (average) size**
  - Trend: shows a slightly upward trend until 1950s and after that turns around to a downward trend
  - Seasonal: shows periodicity of 11 years
  - Errors: residuals graph shows points distributed randomly in sunspot size



POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

4.1 Normality test

- **Normality**
  - Plot a graph of 142 years of wildfires and see a repeat ev ery 11 years.
  - ADF-test and KPSS-test with 90% significance level

  - ADF test H_0 : Time series is non-stationary
  - KPSS test H_0 : Time series is stationary

  - Ordinary data : non-stationary
  - 11-years differential data : stationary

original

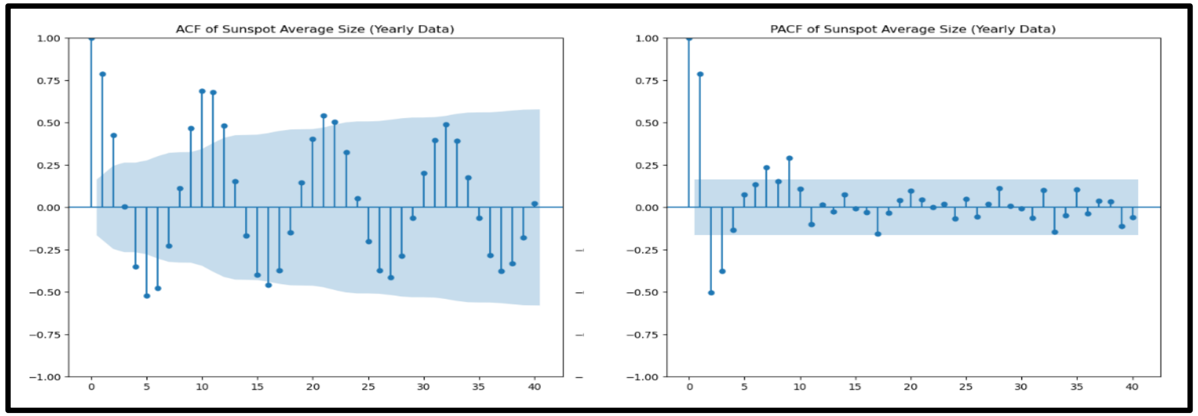
```
#### Results of Dickey-Fuller Test ####
Test-Statistic                  -1.558619
p-value                          0.504340
#Lags Used                       9.000000
Number of Observations used    133.000000
Critical Value (1%)             -3.480500
Critical Value (5%)             -2.883528
Critical Value (10%)            -2.578496
dtype: float64
------------------------------------------------------------
#### Results of KPSS Test ####
KPSS Statistic: 0.44436027438104553
p-value: 0.05803436449092866
num lags: 5
Critical values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
10%:0.347
5%:0.463
2.5%:0.574
1%:0.739
```

11-years differential data

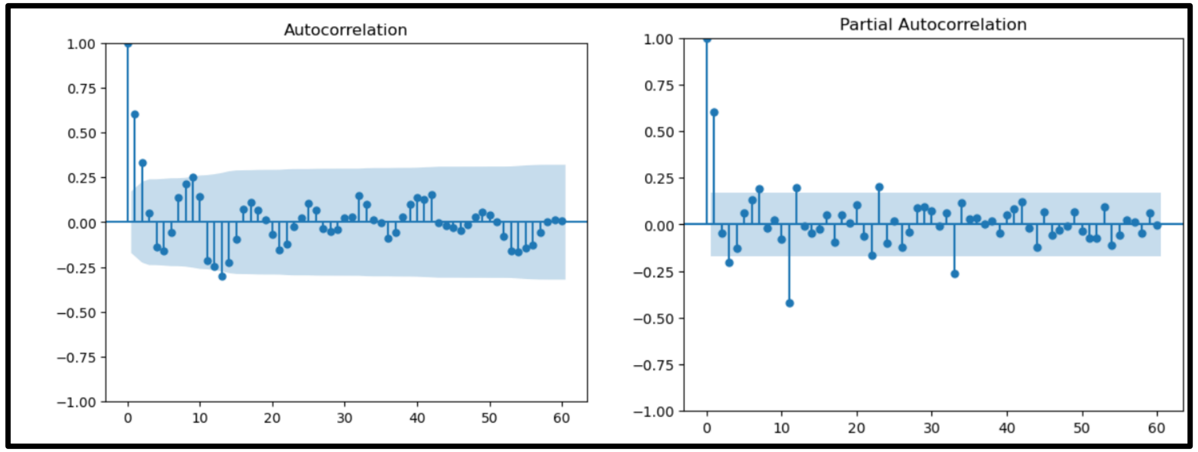
```
#### Results of Dickey-Fuller Test ####
Test-Statistic                  -2.982868
p-value                          0.036526
#Lags Used                      11.000000
Number of Observations used    119.000000
Critical Value (1%)             -3.486535
Critical Value (5%)             -2.886151
Critical Value (10%)            -2.579896
dtype: float64
------------------------------------------------------------
#### Results of KPSS Test ####
KPSS Statistic: 0.28006798301584906
p-value: 0.1
num lags: 5
Critical values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
10%:0.347
5%:0.463
2.5%:0.574
1%:0.739
```

4.2 ACF, PACF of original data(yearly)

- **Sunspot average size**
  - Original data
    - Peaks at 11 lags, with 1-2 lags ahead and be hind also showing correlation
    - Shows periodicity based on 11 lag and slowl y decays



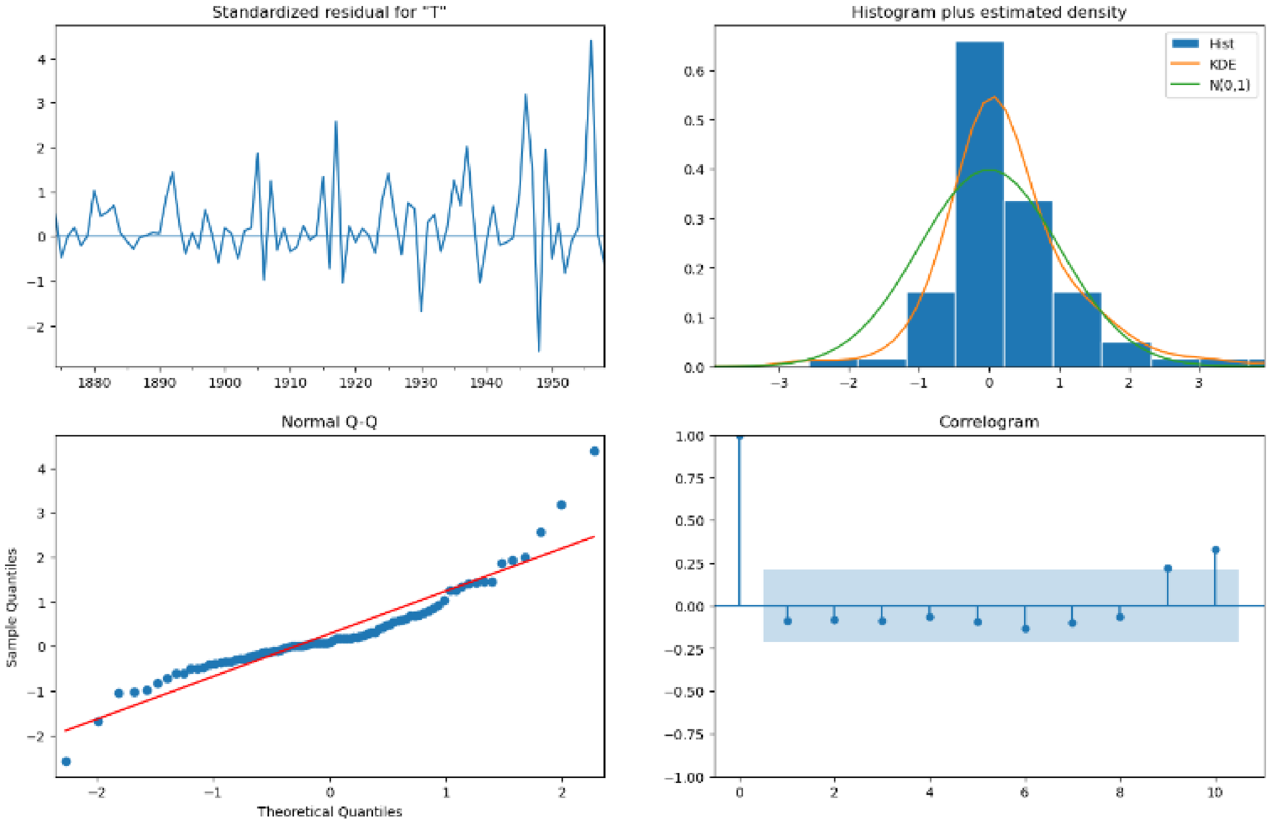
  - Differencing to remove seasonality
    - slowly decays



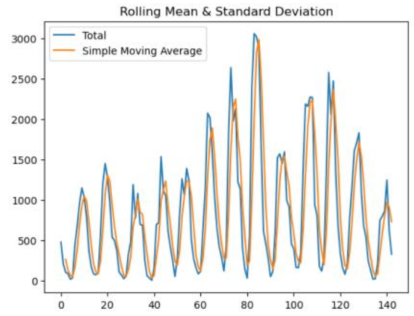
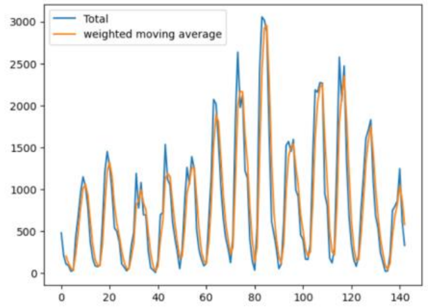
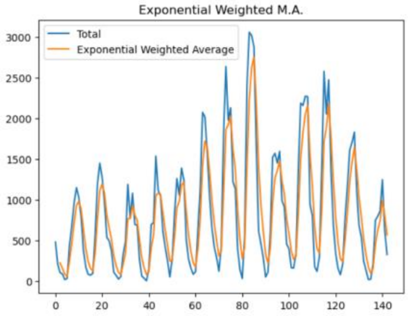
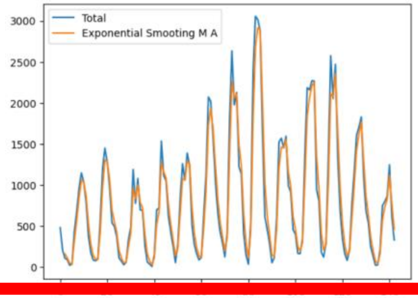
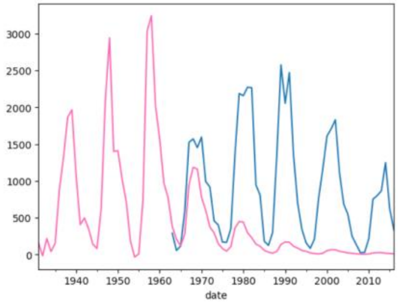
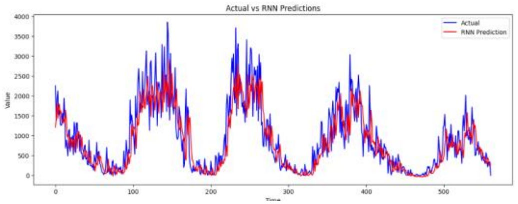
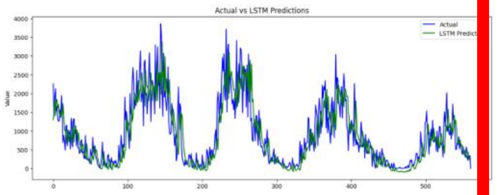
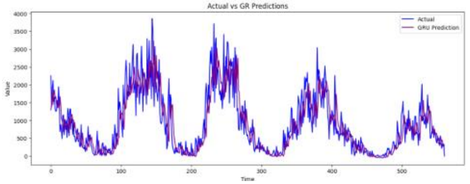
4.3 Time series prediction on size of sunspot

- **ARIMA (1, 0, 2)x(1, 0, 0, 11)**
    - $(1 - 0.6688B)(1 - 0.3778B^{11})Z_t = (1 - 0.3898B - 0.3972B^2)a_t$

```
best_model SARIMA
nonseasonal (1, 0, 2), seasonal (1, 0, 0, 11)
                                      SARIMAX Results
====================================================================================
=======
Dep. Variable:                                Total   No. Observations:
85
Model:               SARIMAX(1, 0, 2)x(1, 0, [], 11)   Log Likelihood                 -
626.266
Date:                              Wed, 29 May 2024   AIC                            1
262.531
Time:                                      12:45:04   BIC                            1
274.745
Sample:                                  01-01-1874   HQIC                           1
267.444
                                       - 01-01-1958
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6688      0.163      4.113      0.000       0.350       0.988
ma.L1          0.3898      0.119      3.287      0.001       0.157       0.622
ma.L2          0.3972      0.170      2.334      0.020       0.064       0.731
ar.S.L11       0.3778      0.179      2.114      0.035       0.027       0.728
sigma2      1.411e+05    1.3e+04     10.851      0.000    1.16e+05    1.67e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.64   Jarque-Bera (JB):                78.89
Prob(Q):                              0.43   Prob(JB):                         0.00
Heteroskedasticity (H):               8.23   Skew:                             1.13
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.14
===================================================================================
```



4.3 Time series prediction on average size of sunspot

<table>
<tr><th>Time series</th><th>Simple MA</th><th>Weighted MA</th><th>Exponential MA</th><th>Exponential SMA</th></tr>
<tr><td>RMSE</td><td>4802.25</td><td>3301.25</td><td>3726.56</td><td>1977.56</td></tr>
<tr><td>Prediction result</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><th>Time series</th><th>ARIMA (1, 0, 2) x (1, 0, 0, 11)</th><th>RNN</th><th>LSTM</th><th>GRU</th></tr>
<tr><td>RMSE</td><td>845.28</td><td>473.89</td><td>440.29</td><td>435.91</td></tr>
<tr><td>Prediction result</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td></td></tr>
</table>

6. Summary

- Sunspot count and average size from 1874 to 2016 was analyzed
- Sunspot count and average size shows a seasonality by period of 11 years
- There seems to be a slight upward and downward trend
- Strong correlation is seen between sunspot count and average size
- Time series prediction model was built by using various techniques and models
- LSTM Deep learning model was selected as a suitable prediction model for the sunspot count
- GRU Deep learning model was selected as a suitable prediction model for the sunspot size

5. Conclusion

- **Results can be used to understand the characteristics of sunspots, and to predict the occu rrence and size of sunspots**

- **Limitation**
  - Unable to forecast time series from multiple perspectives. (Monthly periodicity, multivariate forecasting)
  - There seems to be an upward and downward trend in both sunspot count and average size which could potentially become a seasonal cycle in long term, but can not be confirmed due to the lack of large past data

- **Further research**
  - Analysis of size and count according to the latitude and longitude of sunspot occurrence, since it is well known that sunspots tend to move toward the equator of the sun