

Prediction of Sunspot using Time-series model

20232589 도경근

20232863 배기웅

목차

1. 서론	3
1.1. 흑점의 정의	3
2. 이론적 배경	4
2.1. 시계열 분해	4
2.2. 정상성	4
2.3. ADF 검정, KPSS 검정	4
2.4. ACF, PACF	5
2.5. ARIMA 모형, SARIMA 모형	5
3. 데이터 탐색	6
3.1. 데이터 수집 및 전처리	6
3.2. 데이터 분석	6
3.2.1. 일별 / 주별 / 월별 / 연도별 데이터 탐색	6
3.2.2. 시계열 분해	6
3.2.3. 상관관계 분석	8
3.2.4. ACF 와 PACF 계산	9
3.2.5. 분포추정	10
3.3. 정상성 검정	11
3.3.1. 흑점의 개수 정상성 검정	11
3.3.2. 흑점의 월별 평균크기 정상성 검정	11
4. 시계열 모델 추정	12
4.1. 시계열 모델 추정 과정	12
5. 성능 비교	14
6. 결론 및 한계점	15
7. 참고 문헌	16

1. 서론

1.1. 흑점의 정의

흑점은 태양의 광구(photosphere) 상에 나타나는 상대적으로 어두운 영역으로, 태양의 나머지 부분보다 낮은 온도(약 3,500~4,500K)를 가지고 있다. 이러한 흑점은 태양의 강한 자기장 활동에 의해 발생하며, 자기장이 강할수록 플라즈마의 대류가 억제되어 표면 온도가 낮아진다. 흑점의 크기는 지구보다 큰 것들도 많으며, 때로는 수주 동안 지속되기도 한다

흑점의 수와 활동은 대략 11년 주기의 태양 활동 주기, 즉 태양 주기와 밀접한 관련이 있는 것으로 알려져 있다. 태양 주기의 최소기에는 흑점이 거의 관찰되지 않을 수 있으며, 최대기에는 매우 빈번하고 큰 규모로 나타난다. 이러한 주기성은 태양의 자기장이 뒤집히는 과정과 관련이 있으며, 각 주기의 시작과 끝에서는 자기극의 위치가 반대로 바뀐다. 흑점의 활동이 증가하면 태양 플레어 및 코로나 질량 방출과 같은 다른 태양 활동 현상의 빈도도 증가하는 경향이 있다. 이러한 현상들은 우주 날씨에 큰 영향을 미치며, 지구에도 다양한 영향을 줄 수 있다.

1.2. 흑점 예측의 중요성

흑점은 태양 플레어와 코로나 질량 방출(CME)의 주요 원인으로, 이러한 현상은 지구 주변의 우주 날씨에 큰 영향을 미치며 위성 운영, GPS 시스템, 지상 기술 및 전력망에 지장을 줄 수 있다. 흑점 활동의 정확한 예측은 이러한 우주 날씨 현상에 대한 사전 대비와 완화 전략을 수립하는 데 필수적이다.

다음으로, 흑점 활동은 태양의 에너지 출력 변화와 관련이 있으며, 이는 지구의 기후 패턴에 영향을 줄 수 있다. 흑점 수의 장기적 변화는 지구 온도 변화와 관련이 있어, 예를 들어 '작은 빙하기'와 같은 기후 변동과 연관된다. 따라서 흑점을 이해하고 예측하는 것은 기후 과학에서 중요한 역할을 한다.

또한, 흑점 패턴을 연구하는 것은 태양 동력기구와 흑점 주기를 이끄는 메커니즘을 이해하는 데 도움을 준다. 이러한 이해는 별의 자기 활동 모델을 구축하는 데 기초적인 역할을 하며, 이는 천체물리학의 넓은 맥락에서 근본적인 중요성을 가진다.

1.3. 연구의 목적

위와 같은 다양한 이유로, 흑점 예측은 과학적 연구뿐만 아니라 기술적 응용에 있어서도 매우 중요하다. 이 연구에서는 SARIMA 모델을 포함한 다양한 시계열 및 머신러닝 기법을 활용하여 시간에 따른 흑점의 빈도와 크기를 예측하고, 이러한 방법들의 예측 성능을 비교 분석하였다. 이러한 분석을 통해 보다 정확한 흑점 예측 모델을 개발하고, 관련된 기상 문제 또는 기후 과학 문제에 대응할 수 있는 토대를 마련하였다.

1.4. 관련 연구

시간에 따른 흑점 빈도와 크기를 예측하기 위한 기존 연구는 다수 존재한다. Abdel-Rahman[1]은 ARIMA(Autoregressive Integrated Moving Average)를 통하여 흑점의 빈도를 예측하였다. Lee[2]에서는 LSTM(long short-term memory)을 활용하여 흑점의 빈도를 예측하였다. 또한 Wilson[3]은 회귀분석과 통계 기법을 사용하여 흑점의 크기를 예측하였다. Peng[4]은 LSTM을 활용하여 흑점의 크기를 예측하였다. 이외에도 다양한 시계열 분석과 머신러닝 기법을 활용하여 흑점 빈도와 크기를 예측하기 위한 연구들이 다양하게 존재한다.

2. 이론적 배경

2.1. 시계열 분해

시계열 분해는 시계열 데이터를 추세(Trend), 계절성(Seasonality), 그리고 잔차(Residual)로 분해하여 분석하는 방법이다. 이 방법을 통해 각 구성 요소가 데이터의 특성을 어떻게 나타내는지 파악할 수 있다.

먼저, **추세(Trend)**는 시계열 데이터의 장기적인 변동을 나타낸다. 이는 데이터가 시간이 지남에 따라 전반적으로 증가하거나 감소하는 경향을 보여주며, 데이터의 장기적인 방향성을 이해하는 데 중요한 지표이다. 추세는 주로 장기적인 변동 요인에 의해 영향을 받는다.

다음으로, **계절성(Seasonality)**은 시계열 데이터가 일정한 주기로 반복되는 패턴을 가지는 성질을 의미한다. 계절성은 연도, 분기, 월, 주 단위 등 특정 주기에 따라 데이터가 반복되는 패턴을 보여주며, 이는 주기적인 변동 요인에 의해 발생한다. 계절성을 분석함으로써 주기적인 변동 패턴을 파악하고 향후 데이터의 변동을 예측하는 데 도움을 줄 수 있다.

마지막으로, **잔차(Residual)**는 추세와 계절성을 제외한 나머지 부분으로, 추세와 계절성으로 설명되지 않는 변동 요인을 나타낸다. 잔차는 불규칙한 요소로서, 노이즈나 예측할 수 없는 변동을 포함하고 있으며, 데이터에서 발생하는 임의의 변동을 나타내며, 모델의 오차로 해석될 수 있다.

2.2. 정상성

정상성(Stationary)는 시계열은 뚜렷한 추세나 계절성이 관측되지 않고 시간의 흐름에 따라 일정한 변동을 가지는 성질을 의미한다. 정상적이지 않은 시계열은 차분(differencing)을 통해 추세, 계절성을 제거한 후 정상성을 가지는 시계열로 변환해주어야 한다.

2.3. ADF 검정, KPSS 검정

ADF (Augmented Dickey-Fuller) 검정은 단위근(unit root) 검정 방법 중 하나로 시계열 모델이 정상성을 가지는지 알아보는 검정이다. 단위근은 시계열 데이터가 비정상적인 경향을 가지며, 평균이 일정하지 않고 시간에 따라 변동하는 경우를 의미한다. 단위근 문제가 있다면 시계열 데

이터는 정상성을 가지지 않으며, 이는 신뢰할 수 없는 결과를 가져올 수 있다.

ADF 검정에서는 주어진 시계열 데이터에 대한 회귀 모델을 설정하고 회귀 계수에 대한 검정을 수행한다. 사용하는 가설은 다음과 같다. 귀무가설(H_0)은 자료에 단위근이 존재하여 정상성을 만족하지 않음을, 대립가설(H_1)은 정상성을 만족함을 뜻한다. 검정 결과에는 ADF 검정 통계량과 p-value가 포함되는데, p-value가 유의수준보다 작으면 단위근 문제가 없다는 귀무가설을 기각하고 데이터가 정상성을 가진다고 결론 낼 수 있다. 그렇지 않으면 데이터가 정상성을 가지지 않는다고 판단된다. 검정 통계량이 작을수록 단위근이 존재하는 귀무가설을 기각할 수 있는 근거가 더 강력하다고 해석할 수 있다.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests도 ADF test처럼 시계열 데이터의 정상성을 검정하는 방법으로, 단위근 검정방법 중 하나이다. KPSS test의 가장 큰 특징은 ADF 테스트와 달리 귀무가설(H_0)이 "시계열이 정상 시계열이다"로 설정하며, 반대로 대립가설(H_1)을 "시계열이 비정상 시계열이다"라고 설정하여 검정한다.

2.4.ACF, PACF

ACF(자기상관함수, Autocorrelation Function)는 시차에 따른 자기상관을 의미한다. 현재 데이터와의 시점 차이인 lag에 따른 ACF를 관찰해 정상성을 결정한다. 정상성을 가지는 시계열은 빠르게 0에 수렴한다.

ACF는 모든 시계열 데이터의 특성을 분석하는 것에 한계가 있다. 시계열 모델 중에 AR(q) 모델이나 MA(p)모델 중에 어떤 것을 활용하고, 어떤 시차(lag)를 적용할지에 대해서 ACF를 활용하여 결정하는 것은 어렵다. 따라서 이러한 경우, PACF를 도입하여 추가적인 분석을 진행할 수 있다. PACF (편자기상관함수, Partial Autocorrelation Function)은 시차에 따른 편자기상관으로 y_t 와 y_{t-k} 간의 순수한 상관관계로 두 시점 사이에 포함된 모든 시점의 영향은 고려하지 않는다.

2.5.ARIMA모형, SARIMA모형

ARIMA(Autoregressive Integrated Moving Average)은 AR 모델과 MA 모델을 결합한 모델이다. ARIMA 모델은 정상성을 가정하기 때문에 비정상 시계열을 차분을 통해 정상 시계열로 변형해줘야 한다.

Seasonal ARIMA 모형은 기존의 ARIMA모형이 계절성을 반영을 못한다는 점을 보완한 시계열 예측 모형이다. SARIMA 모형은 $(p, d, q)(P, D, Q)_m$ 의 형태를 가진다. 모형의 계수에 대한 설명은 다음과 같다.

- p: 비계절성 AR 모형의 lag
- q: 비계절성 MA 모형의 lag
- d: 비계절성 모형의 차분

- P: 계절성 AR 모형의 lag
- Q: 계절성 MA 모형의 lag
- D: 계절성 모형의 차분
- m: 계절성 주기

모형은 다음 식 (1) 을 따른다.

$$\phi_p(B)\Phi(B^m)(1-B)^d(1-B^m)^D X_t = \theta_q(B)\Theta(B^m)a_t \quad (1)$$

3. 데이터 탐색

3.1. 데이터 수집 및 전처리

NASA, Greenwich Royal observatory에서 제공하는 USAF/NOAA Sunspot Dataset에서 1874년 5월부터 2016년 10월까지의 sunspot count와 sunspot Average size 데이터를 획득하였다. 그 후 결측값 처리, 데이터 병합과 같은 전처리를 통하여 시간에 따른 흑점 빈도와 크기를 예측하기 위해 이한 데이터 형태로 전처리를 진행하였다. 최종적으로 날짜, 해당 날짜의 흑점 빈도, 흑점 평균크기에 대한 데이터를 만들었다.

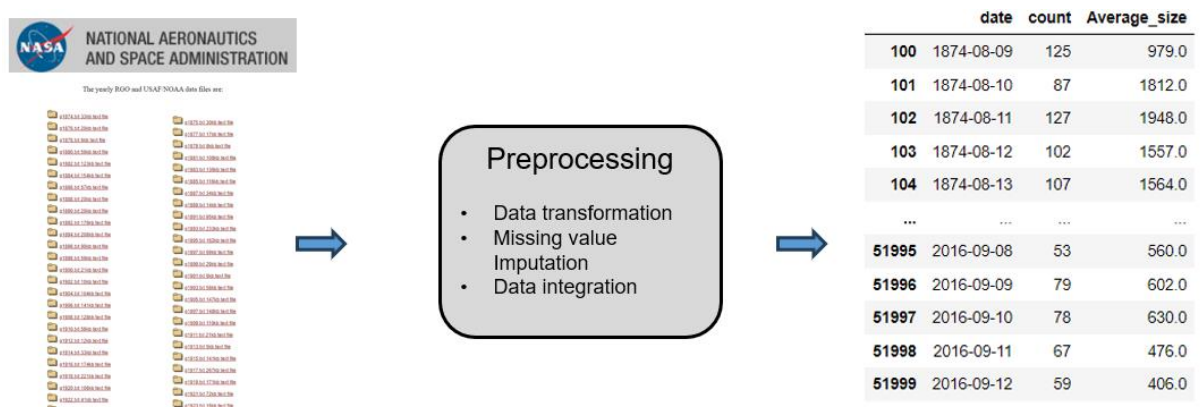


Figure 1 데이터 수집 및 전처리 overview

3.2. 데이터 분석

3.2.1. 일별/주별/월별/연도별 데이터 탐색

수집한 데이터를 바탕으로 일별/주별/월별/연도별 흑점 빈도와 흑점 평균크기를 시각화 해보았다.

흑점 개수와 같은 경우에는 11년 주기의 태양 흑점 활동 주기가 모든 그래프에서 명확히 관찰된다. 특히, 1950년까지 흑점의 빈도는 점차 증가하는 추세를 보이다가 그 이후로는 점차 감소하는 경향이 있다. 이러한 추세는 태양 활동의 장기적 변동성을 반영하는 것으로, 태양의 자기장 활동과 관련된 다양한 외부 요인에 의해 영향을 받을 수 있다. 일별 데이터에서는 매우 높은 변

동성을 보이며, 연도별 데이터에서는 이 변동성이 상대적으로 매끄러운 형태로 나타난다.

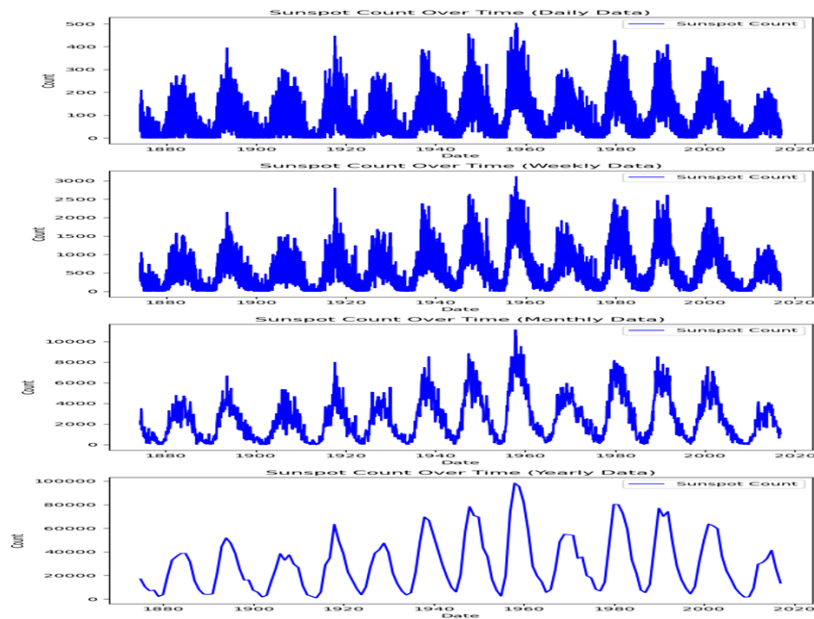


Figure 2 일별/주별/월별/연도별 흑점 개수 시각화

흑점 평균크기와 같은 경우에서도 11년의 주기성이 확인되며, 1950년을 기점으로 크기는 점진적으로 감소하는 추세를 보인다. 이는 흑점의 빈도와 유사한 경향을 보이는데, 이는 태양의 자기장 강도 변화와 직접적인 관련이 있을 수 있다. 연도별 데이터에서는 흑점 크기의 변동 폭이 가장 크게 나타나며, 일별 데이터에서는 높은 변동성을 보여준다.

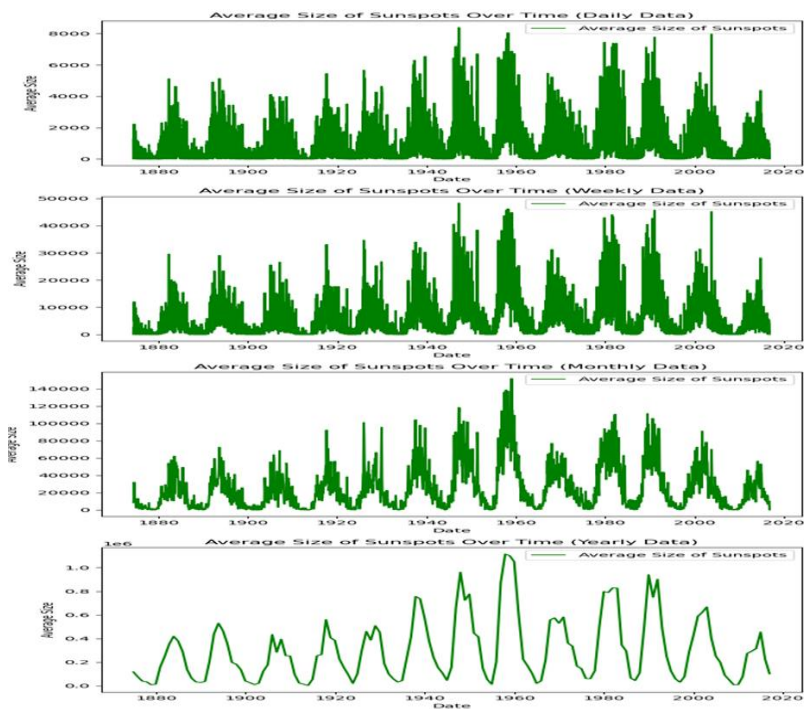


Figure 3 일별/주별/월별/연도별 흑점 평균크기 시각화

일별/주별/월별/연도별 데이터의 시각화에서 나타나는 차이점은 주로 데이터의 집계 단위에 따른 변동성의 차이에서 기인한다. 일별 데이터는 매우 높은 변동성을 보이며, 특히 단기적인 이상치가 더 자주 나타난다. 주별과 월별 데이터에서는 이 변동성이 다소 줄어들지만 여전히 주기적 패턴이 명확히 관찰된다. 연도별 데이터는 가장 안정적인 트렌드를 보여주며, 장기적인 주기성과 경향성을 분석하기에 적합하다. 이러한 차이는 데이터의 집계 단위가 클수록 단기적 변동성이 평탄화 되고 장기적인 경향이 더 명확하게 드러나기 때문이다.

3.2.2. 시계열 분해

시계열 분해를 통하여 태양의 흑점 빈도와 흑점 평균크기의 추세, 계절성, 잔차에 대하여 분석해보았다.

태양의 흑점 개수와 흑점 평균크기의 시계열 분해를 통한 분석은 다음과 같다. 첫째, 트렌드 성분에서는 1950년대까지 약간의 상승 추세를 보이다가 그 이후로 하강 추세로 전환되는 경향을 보인다. 이는 태양 활동의 장기적인 변화를 반영하는 것으로 해석될 수 있다. 둘째, 계절성 성분에서는 약 11년의 주기성을 나타내며 이는 잘 알려진 태양의 활동 주기와 일치한다. 마지막으로, 잔차 성분을 보면 잔차 그래프에서 흑점 수에 대한 점들이 무작위로 분포하고 있는 것을 확인할 수 있는데, 이는 모델이 데이터의 주요 패턴을 잘 포착하고 있음을 시사한다

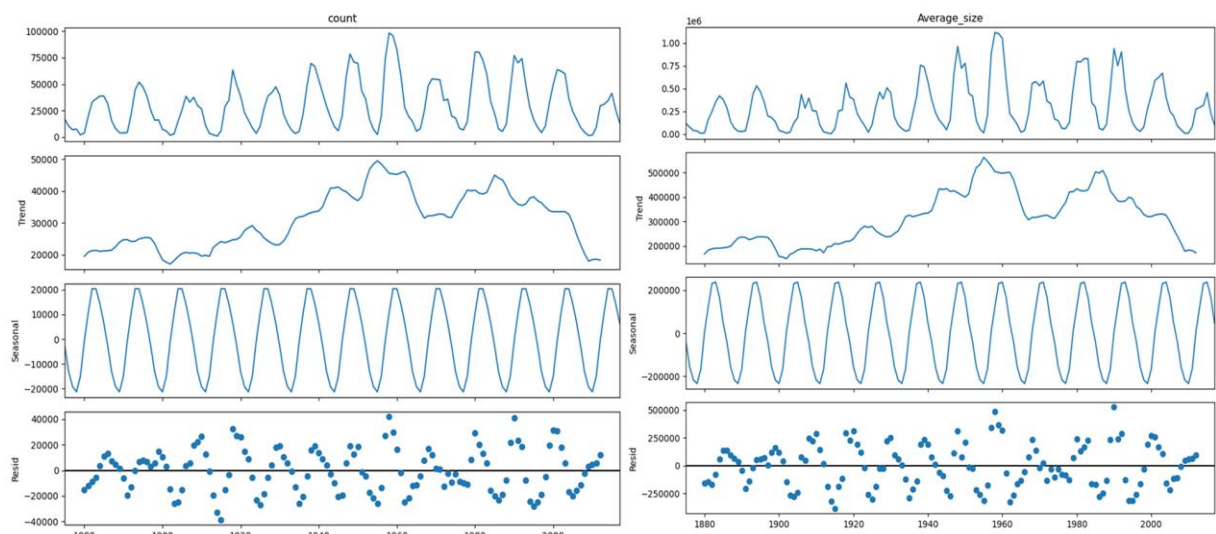


Figure 4 태양의 흑점 개수(좌)/흑점 평균크기(우)의 시계열 분해 분석 결과

3.2.3. 상관관계 분석

태양의 흑점 개수와 흑점 평균크기 사이의 상관관계를 파악하기 위하여 상관관계 분석을 시행하였다. 상관계수는 0.88로 상당히 강한 양의 상관관계를 가진다고 할 수 있다. 이를 통하여 흑점의 개수가 늘어나고 태양의 활동이 활발해지는 태양의 극대기에는 흑점의 크기 또한 같이 커진다는 점을 알 수 있다.

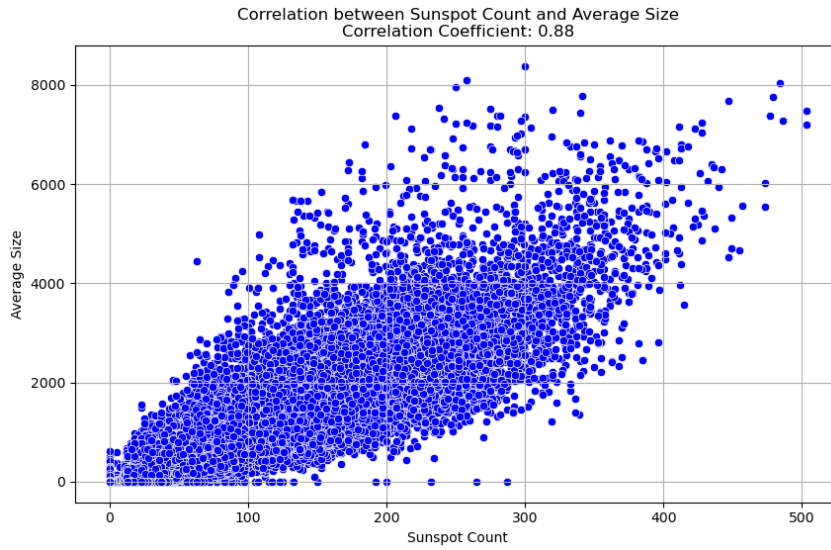


Figure 5 흑점의 개수와 평균크기의 상관관계 분석

3.2.4. ACF와 PACF 계산

태양의 흑점 개수와 흑점 평균크기의 ACF, PACF의 패턴을 확인했다.

태양의 흑점 개수의 경우에는 계절성을 제거하지 않은 경우 ACF는 11번째의 lag에서 피크를 보이며 점진적으로 줄어드는것으로 보인다. 계절성 차분을 한 후 ACF에서는 3번째 lag에서 cut-off가 보인다. PACF는 계절성을 제거하기 전과 후 모두 lag이 3에서 cut-off가 보인다.

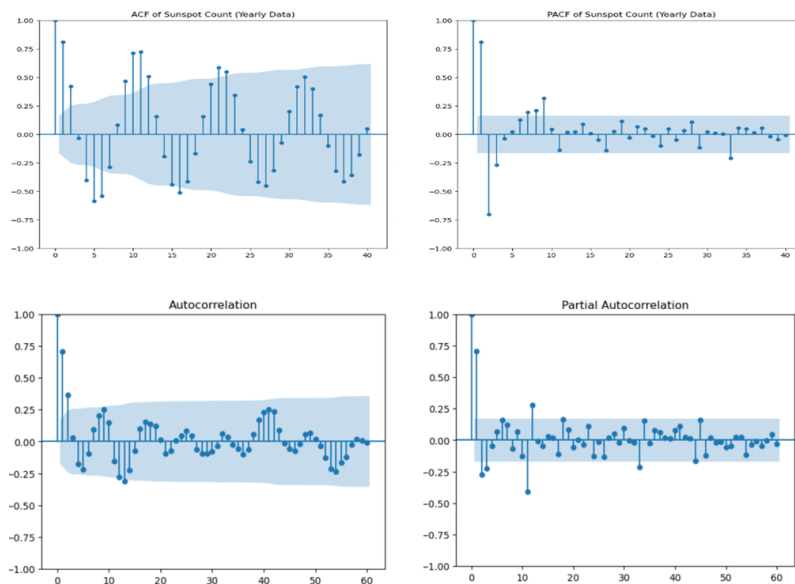


Figure 6 흑점 개수 계절성을 제거하지 않은 ACF, PACF (위), 계절성 차분 후 ACF, PACF (아래)

태양의 흑점 평균크기의 경우에는 계절성을 제거하지 않은 경우 ACF는 11번째의 lag에서 피크를 보이며 점진적으로 줄어드는 것으로 보인다. 계절성 차분을 한 후 ACF에서는 3번째 lag에서 cut-off가 보인다. PACF는 계절성을 제거하기 전과 후 모두 lag이 3에서 cut-off가 보인다.

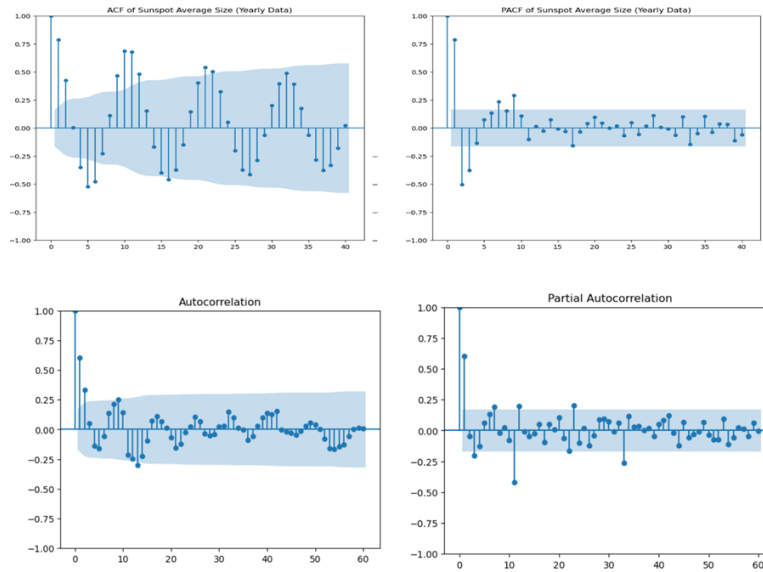


Figure 7 흑점 개수 계절성을 제거하지 않은 ACF, PACF (위), 계절성 차분 후 ACF, PACF (아래)

3.2.5. 분포추정

흑점의 개수와, 흑점의 크기에 대하여 분포를 추정하였다. 흑점의 개수는 정수형이며, 이산형 확률분포를 따르고, 흑점의 크기는 연속형 확률분포를 따른다. 이산형 확률분포는 포아송분포와 음이항 분포를 가정하고, 연속형 확률분포는 지수분포, 감마분포, 로그노말분포, 로그 로지스틱 분포, 웨이블 분포, 그리고 skew-normal을 가정하여 분포 추정을 진행하였다.

먼저, 흑점의 개수에 대한 분포를 추정한 결과, 음이항 분포에 대한 적합도가 가장 높았음을 알 수 있었다. 음이항 분포에 대한 추정결과가 AIC 값 784727.6을 가지며 최적의 분포로 추정이 되었다.

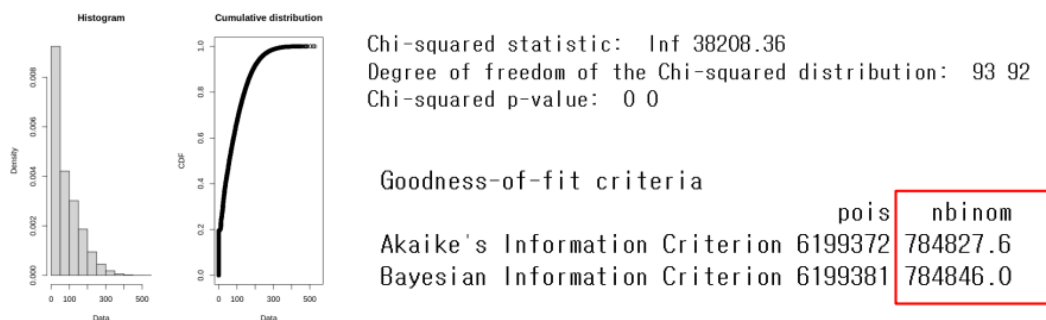
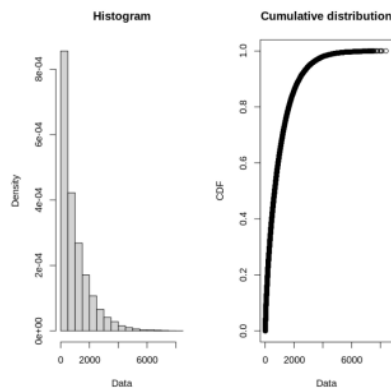


Figure 8 흑점의 개수에 대한 분포 추정결과

흑점의 월별 평균크기에 대한 분포를 추정한 결과, 총 6개의 분포 중, 감마분포에 대한 적합도가 가장 높았음을 알 수 있었다. 로그우도값을 기준으로 분포를 선정하였을 때, 감마분포가 -348441.2를 가지며 최적의 분포로 선정이 되었다.



'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'

Figure 9 흑점의 평균크기에 대한 분포 추정결과

3.3. 정상성 검정

총 1874년부터 2016년까지 142년동안의 흑점 시계열 데이터를 기반으로 정상성 여부에 대한 테스트를 진행하였으며, ADF 테스트와 KPSS 테스트를 이용하였다.

3.3.1. 흑점의 개수 정상성 검정

10%를 유의수준으로 검정한 결과, 차분하기 전 원래 데이터셋에 대해 p-value값은 ADF 테스트 0.41, KPSS 테스트 0.08의 값을 가지며 기각하지 못하였다. 11년을 기준으로 차분한 결과 p-value값은 ADF테스트는 0.06, KPSS 테스트는 0.1을 가지며 귀무가설을 기각하여 정상성을 가짐을 확인하였다.

#### 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	

#### 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	

Figure 10 흑점의 개수에 대한 정상성 검정 결과 (좌: 차분하기 전, 후 : 차분한 후)

3.3.2. 흑점의 월별 평균크기 정상성 검정

흑점의 평균크기 시계열 데이터셋도 동일한 방식으로 진행하였다. 10%를 유의수준으로 검정한 결과, 차분하기 전 원래 데이터셋에 대해 p-value값은 ADF 테스트 0.50, KPSS 테스트 0.05의 값을 가지며 기각하지 못하였다. 11년을 기준으로 차분한 결과 p-value값은 ADF테스트는 0.03, KPSS 테스트는 0.1을 가지며 귀무가설을 기각하여 정상성을 가짐을 확인하였다.

#### Results of Dickey-Fuller Test ####		#### Results of Dickey-Fuller Test ####	
Test-Statistic	-1.558619	Test-Statistic	-2.982868
p-value	0.504340	p-value	0.036526
#Lags Used	9.000000	#Lags Used	11.000000
Number of Observations used	133.000000	Number of Observations used	119.000000
Critical Value (1%)	-3.480500	Critical Value (1%)	-3.486535
Critical Value (5%)	-2.883528	Critical Value (5%)	-2.886151
Critical Value (10%)	-2.578496	Critical Value (10%)	-2.579896
dtype: float64		dtype: float64	
#### Results of KPSS Test ####		#### Results of KPSS Test ####	
KPSS Statistic	0.44436027438104553	KPSS Statistic	0.28006798301584906
p-value	0.05803436449092866	p-value	0.1
num lags: 5		num lags: 5	
Critical values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}		Critical values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}	
10%:0.347		10%:0.347	
5%:0.463		5%:0.463	
2.5%:0.574		2.5%:0.574	
1%:0.739		1%:0.739	

Figure 11 흑점의 평균크기에 대한 정상성 검정 결과 (좌: 차분하기 전, 후 : 차분한 후)

4. 시계열 모델 추정

4.1. 시계열 모델 추정 과정

본 프로젝트에서는 총 8가지의 시계열 예측모델을 사용하였으며, 그 중에 ARIMA모델을 이용하여 흑점의 개수와 흑점의 월별 평균크기 시계열 데이터를 예측하였다.

먼저, 기존 시계열 데이터셋의 계절성 여부를 확인하였다.

흑점은 11년 주기성을 가진다. 이러한 주기성을 기반으로 연년 흑점 시계열 데이터를 차분하여 SARIMA 모형의 모수를 추정하였다.

먼저, 흑점의 개수 시계열 데이터셋에 대한 SARIMA모형을 추정하였다. 최종적으로 (1,0,2) (1,0,0)₁₁모형이 AIC값 811.501을 가지며 최적의 모형으로 선정이 되었다.

```
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			

Figure 12 흑점의 개수에 대한 SARIMA 모형 추정결과

test dataset에 대한 예측 결과, RMSE값이 60.11을 가짐을 확인하였다.

SARIMA 모형의 잔차를 확인한 결과, 자기 상관성은 없고, 정규성을 만족하지만 시간이 지날수록 잔차가 커지는 양상을 보였다.

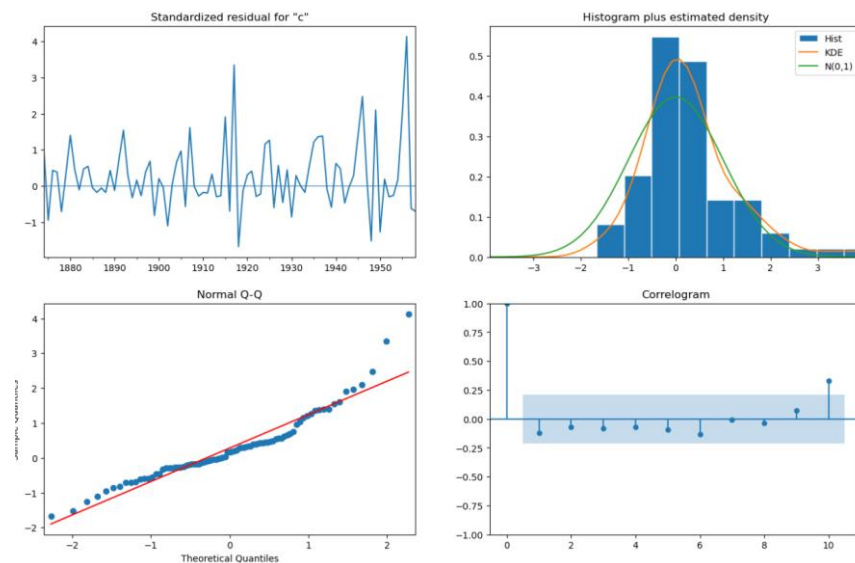


Figure 13 흑점의 개수 시계열 예측 잔차

다음은 흑점의 월별 평균크기 시계열 데이터에 대한 SARIMA모형을 추정하였다. 최종적으로 $(1,0,2)(1,0,0)_{11}$ 모형이 AIC값 1262.531을 가지며 최적의 모형으로 선정되었다.

```
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	1262.531
Time:		12:45:04	BIC	1274.745
Sample:		01-01-1874	HQIC	1267.444
		- 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

Figure 14 흑점의 평균크기에 대한 SARIMA 모형 추정결과

test dataset에 대한 예측 결과, RMSE 값이 845.28을 가짐을 확인하였다.

SARIMA 모형의 잔차를 확인한 결과, 흑점의 개수에 대한 잔차와 비슷한 결과를 내었음을 확인하였다.

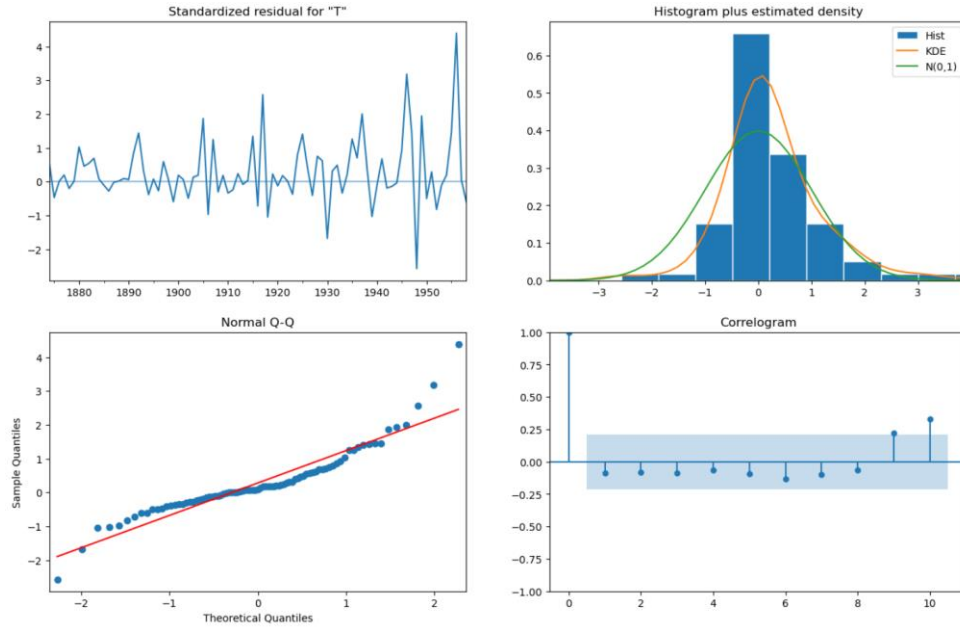


Figure 15 흑점의 평균크기 시계열 예측 잔차

5. 성능 비교

사용한 모델은 단순이동평균, 가중이동평균, 지수이동평균, 지수평활평균, SARIMA, RNN, LSTM, 그리고 GRU를 사용하였다. 각 모델을 이용하여 흑점의 개수 시계열과 흑점의 월별 평균크기 시계열을 예측하였고, RMSE를 평가지표로 설정하여 성능을 비교하였다.

흑점의 개수에 대한 시계열 예측 결과, 총 8개의 모델들 중, LSTM의 RMSE값이 24.33으로 가장 좋은 성능을 내었음을 알 수 있었다.

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

Figure 16 흑점의 개수 시계열 예측 결과

반대로, 흑점의 월별 평균크기에 대한 시계열 예측에 대해서는 GRU의 성능이 RMSE 값 435.91을 가지며 가장 좋은 성능을 내었음을 알 수 있었다.

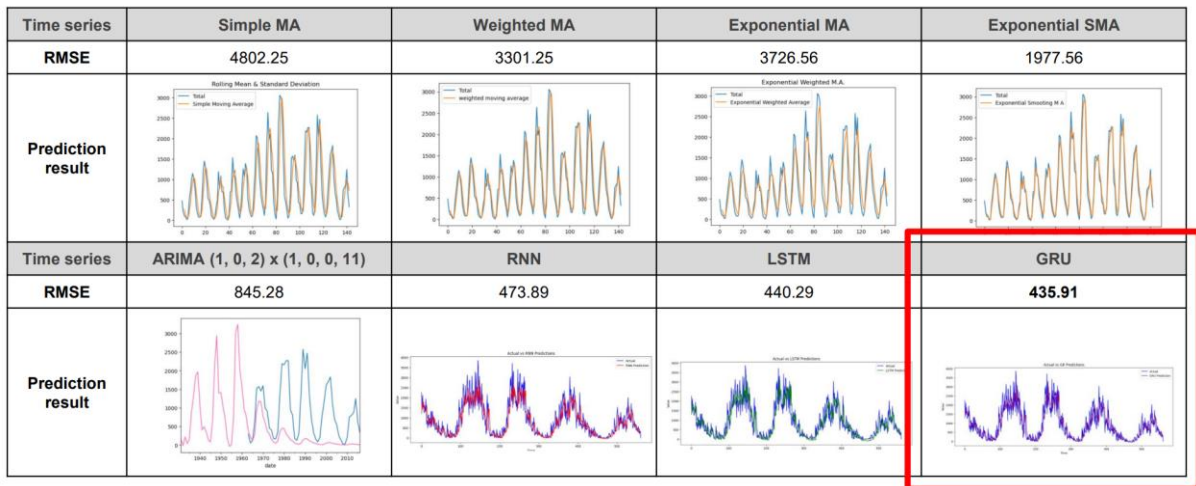


Figure 17 흑점의 평균크기 시계열 예측 결과

RMSE값의 전체 분포를 고려하였을 때, 딥러닝 기반의 시계열 예측 모형의 성능이 전반적으로 좋았음을 알 수 있다.

6. 결론 및 한계점

본 연구에서는 1874년부터 2016년까지의 태양 흑점 수와 평균크기 데이터를 분석하고는 SARIMA 모델을 포함한 다양한 시계열 및 머신러닝 기법을 활용하여 시간에 따른 흑점의 빈도와 크기를 예측 및 이러한 방법들의 예측 성능을 비교 분석하였다. 태양 흑점의 수와 평균크기 모두 약 11년의 주기성을 보이는 계절성이 확인되었으며, 미세하게나마 상승 및 하락 추세가 관찰되었다. 이러한 흑점 수와 크기 사이에는 강한 상관관계가 있음이 드러났다.

다양한 시계열 분석과 머신러닝 모델을 사용하여 태양 흑점 개수와 평균크기에 대하여 총 8개의 시계열 예측 모델을 구축하였고, 그 중 LSTM 딥러닝 모델은 흑점 수의 예측에, GRU 딥러닝 모델은 흑점 크기의 예측에 적합한 모델로 선정되었다. 이러한 결과는 태양 흑점의 특성을 이해하고, 흑점의 발생 및 크기를 예측하는 데 유용하게 사용될 수 있다. 더 나아가 흑점 활동의 정확한 예측은 우주 날씨 예측 및 기후 연구에 사용될 수 있을 것이다.

연구의 한계로는 월별 주기성, 다변량 예측과 같은 여러 관점에서의 시계열 예측을 수행하지 못했다는 점이 있다. 또한, 흑점 수와 크기 모두 장기적으로 계절 주기를 형성할 가능성이 있지만, 과거 데이터의 부족으로 이를 입증하기 어렵다.

추후 연구에서는 흑점의 위도와 경도에 따른 크기와 수의 분석이 필요하다. 태양의 적도로 이동하는 경향이 있는 흑점의 특성을 고려할 때, 이러한 분석은 태양 흑점 연구에 있어 중요한 시사점을 제공할 것이다. 이를 통해 보다 정밀한 태양 흑점의 행동 예측 모델을 개발할 수 있을 것으로 기대된다.

7. 참고 문헌

- [1] H.I. Abdel-Rahman, B.A. Marzouk, "Statistical method to predict the sunspots number", NRIAG Journal of Astronomy and Geophysics, Volume 7, Issue 2, Pages 175-179, (2018)
- [2] Lee, T. EMD, "LSTM Hybrid Deep Learning Model for Predicting Sunspot Number Time Series with a Cyclic Pattern", Sol Phys 295, Pages 82-88, (2020)
- [3] Wilson. R, "A prediction for the size of sunspot cycle 22", Geophysical Research Letters, Volume 15, Issue 2, Pages 125-128, (1988)
- [4] X. Peng, "Sunspot Activity and Cycle Prediction Study Based on Time Series Algorithm and CNN-LSTM Modeling", 2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, Chinapp. Pages 1211-1215, (2024)