



논문

# 강릉 지역의 주야간 에어로솔 광학 특성과 머신러닝을 이용한 예측 Day and Night Aerosol Optical Properties in Gangneung and their Prediction using Machine Learning

이권호\*

국립강릉원주대학교 복사위성연구소, 대기환경과학과

Kwon-Ho Lee\*

Research Institute for Radiation-Satellite, Department of Atmospheric & Environmental Sciences,  
Gangneung-Wonju National University, Gangneung, Republic of Korea

접수일 2023년 8월 17일  
수정일 2023년 8월 25일  
채택일 2023년 8월 25일

Received 17 August 2023  
Revised 25 August 2023  
Accepted 25 August 2023

\*Corresponding author  
Tel : +82-(0)33-640-2319  
E-mail : kwonho.lee@gmail.com

**Abstract** This study reports the results of a statistical analysis of day and night aerosol optical properties measured from the Cimel CE318-T sun-lunar-photometer operated at Gangneung-Wonju National University (37.771°N, 128.867°E) from January 2015 to December 2022. Aerosol observations during the day and night are an important means of monitoring aerosols during the diurnal cycle, and we aimed to use these observations to the study of regional aerosol properties. In addition, since long-term observations can be used to predict current and future variations through machine learning-based time series analysis technique, the optimal modeling technique was determined and future predictions were presented. Statistical analysis of the day and night observations showed that daily and seasonal aerosol optical depth at 500 nm wavelength ( $AOD_{500}$ ) and angstrom exponent at 440 nm and 675 nm wavelengths ( $AE_{440,675}$ ) are characterized by similar patterns, respectively. More specifically, the  $AOD_{500}$  and  $AE_{440,675}$  datasets observed at night range approximately 5.7% and 1.9% higher than the observed during day. Additionally, for modeling time series analysis with long-term aerosol observations, the computation of auto regressive moving average models among machine learning techniques was applied and evaluated by analyzing the differences in the results. Pairwise comparisons, correlation coefficients, root mean square errors, and mean bias also supported the reliability of the data. The results also showed that the Seasonal AutoRegressive Integrated Moving Average with Exogenous Variables (SARIMAX) model was the most accurate, and future predictions were compared.

**Key words:** Machine learning, Lunarphotometer, Aerosol optical depth, Angstrom exponent, Sunphotometer

## 1. 서론

대기 에어로솔은 대기 중에 떠다니는 작은 미립자로서 각종 먼지, 황산염, 질소산화물 등이 포함된다. 이러한 에어로솔은 기후, 대기 화학, 인간 건강 등에 영향을 미치는 것으로 알려져 있다 (Lelieveld *et al.*, 2019; IPCC, 2014). 에어로솔의 관측 방법 중 지상에서 복사 관측 장비를 이용한 관측 방법은 광원(light source)으로부터 방사(emit)되는 광학 정보를 측정하여 에어로

솔 입자의 광학 특성을 획득할 수 있다. 대표적인 복사 관측 장비인 썬포토미터(Sun photometer)는 태양 빛의 투과도를 측정하여 대기 에어로솔의 광학적 농도와 크기 분포 정보를 측정하는 데 사용되고 있다. 특히, 전 세계적인 썬포토미터 관측 네트워크인 Aerosol Robotic Network (AERONET)는 장기간에 걸친 대기 관측을 통하여 지역이나 전 지구에 대한 대기오염 평가와 기후변화 연구에 기여하는 중요한 데이터를 제공해왔다(Holben *et al.*, 1998).

썬포토크미터는 기본적으로 태양을 광원으로 이용하여 대기 에어로솔의 광학적 특성을 측정하는 데 사용되지만, 야간에도 에어로솔을 관측할 수 있는 방법이 개발되었다(Leiterer *et al.*, 1995). 야간 동안에는 태양 빛을 대체하여 천체에서 측정가능한 광원(예: 달, 별, 인공 광원 등)의 빛을 활용하여 대기 에어로솔을 측정하는 방법이 사용된다. 이 중 달빛을 이용한 썬포토크미터는 루나포토크미터(Lunarphotometer)라고도 하며 달 빛(태양광이 달표면에서 반사된 빛)이 지구 대기를 통과하면서 감쇄되는 투과도를 측정하여 야간 에어로솔 광학 두께(Aerosol Optical Depth (AOD))를 결정한다(Barreto *et al.*, 2016, 2013). 야간 에어로솔 관측값은 대기오염 평가, 기후변화 연구 등에 필요한 일변화 특성에 관한 정보를 제공할 수 있으므로, 대기 중 에어로솔의 시간적 특성 변화 파악, 미세 먼지 발생 원인 등에 대한 중요한 정보를 얻을 수 있다. 예를 들면, 주야간 변화는 태양 복사 프로세스와 대기 역학의 변화에 따른 대기 입자 형성 과정을 설명할 수 있는 중요한 수단이 된다(Huang *et al.*, 2013; Pérez-Ramírez *et al.*, 2012). 또한 일 주기 변화를 따르는 배출원(예: 교통, 난방, 농업 활동 등)은 야간 관측 정보가 포함되지 않으면 대기질이나 기후학 관련 연구에서 편향이 생길 수 있다(Roman *et al.*, 2020).

썬포토크미터 관측은 연속적인 에어로솔 광학 특성 정보를 생산할 수 있는 수단으로서 주간 관측 정보만을 생산하였기에 에어로솔의 전체 일변화를 분석하기 어려웠으나, 야간 관측 정보의 추가로 인하여 보다 관측 범위가 확장되었다. 또한, 주야간 에어로솔 연속관측 정보는 시계열 분석 및 예측을 통하여 현재와 미래 환경변화에 대한 분석을 가능하게 함으로써 대기질 현황과 기후변화 관련 요인 분석에 있어 중요한 역할을 담당할 것이다. 이렇게 시간 단위로 정렬된 일련의 관측값은 시계열 자료로 정의되며, 시계열 자료의 분석에서 사용되는 기본 가정은 미래의 값은 과거의 변화 패턴에 따라 달라진다는 것이다. 이러한 시계열 분석을 위한 통계적 접근 방식이 사용되고 있으며, 확률적 변동 프로세스는 그중 하나로서 변수값의 불확실

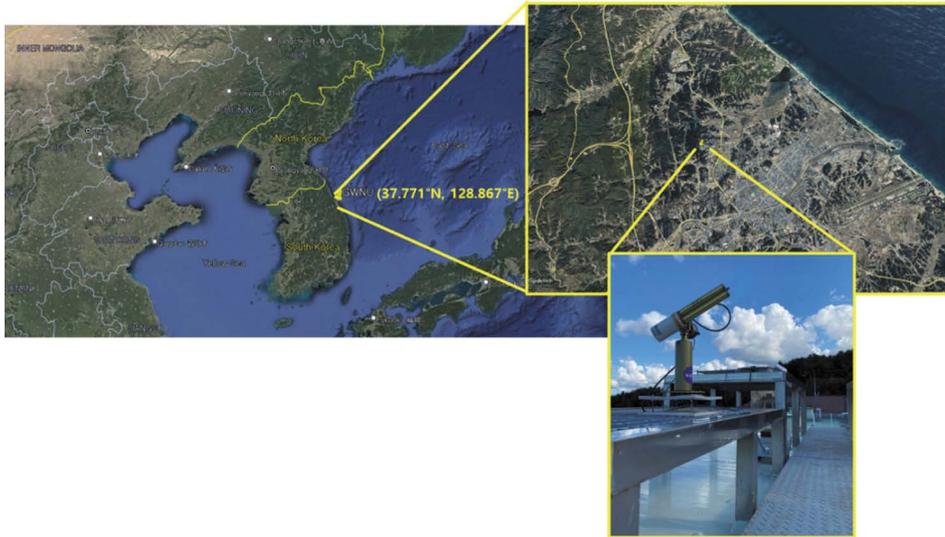
성을 예측에 반영하기 위한 일련의 프레임워크를 제공한다. 그리고 가장 널리 사용되는 모델은 Box-Jenkins 자기회귀 이동 평균(Auto Regression Moving Average (ARMA)) 모델이다(Box *et al.*, 1994). 이 모델은 관찰된 시계열과 동일한 지속성 구조를 가진 합성 시계열을 생성하여 과거값으로부터 미래를 예측하는 통계 확률적 방법을 사용한다. 이 모델은 풍속(Cadenas and Rivera, 2010), 강수량(Kripalani and Kulkarni, 2001), 대기 및 수온(Tularam and Ilahee, 2010), 대기오염 기준물질의 예측(Chaudhuri and Dutta, 2014; Chattopadhyay and Chattopadhyay, 2009; Kim *et al.*, 1988)에 사용되었다. 그러나 국내에서는 확률 모델을 사용한 에어로솔 광학 특성의 미래 추세를 예측하는 연구가 거의 수행되지 않았다.

따라서 본 연구에서는 강릉 지역에서 장기간 썬포토크미터 관측으로부터 획득한 주야간 대기 에어로솔의 광학 특성을 비교 분석하였다. 장기간 관측된 주야간 대기 에어로솔 자료를 기반으로 한 시계열 변화 유형, 그리고 인공지능 기법 중 시계열 예측이 가능한 머신러닝 모델을 이용한 예측 정확도를 분석하였다. 마지막으로 최적의 예측 정확도를 가지는 예측 모델을 이용한 미래의 예측값의 특성범위에 대해 논의하였다.

## 2. 자료 및 방법

### 2.1 에어로솔 관측자료

본 연구에서 사용된 에어로솔 관측자료는 강원도 강릉시 강릉원주대 캠퍼스 내에 위치한 AERONET 관측 지점(북위 37.771°, 동경 128.867°, 해발 60.0 m)에서 2005년부터 2022년까지의 Cimel CE318-T sun-sky-lunar multiband photometer 관측자료이다. 관측 지점은 동해안에서 5 km, 최 근접한 대관령까지 약 12 km 정도 떨어져 있다(그림 1). 강릉의 대기 에어로솔은 비교적 낮은 수준의 교외 특성을 나타내고 있으며, 때때로 장거리 이동성 에어로솔 및 해양성 오염원으로부터 영향을 받는 것으로 간주될 수 있다(Lee and



**Fig. 1.** Geographical location of the CIMEL-318T observation site at the Gangneung-Wonju National University in Gangneung, Korea (Pointed as a yellow pin in the map, latitude: 37.771°N, longitude: 128.867°E, altitude: 60.0 m).

Lee, 2020). 따라서 다양한 출처의 에어로솔 특성에 관한 모니터링에 이상적이다.

CE318-T sun-sky-lunar multiband photometer는 프랑스의 Cimel Electronique에서 개발하였으며, 기존의 CE318 모델의 확장형으로서 주간 및 야간 측정이 가능하다. CE318-T은 대략 1.29°의 관측각과 8개의 중심 파장(1020, 940, 870, 675, 500, 440, 380 및 340 nm)을 가지는 실리콘 광 다이오드 검출기와 2개의 파장(1020 및 1640 nm)을 측정하는 InGaAs 검출기를 사용한다. 이 중 자외선 영역의 340 nm와 380 nm는 낮은 신호값이 측정되므로 야간에는 측정할 수 없다.

CE318-T 모델은 특히 야간의 달빛을 측정하기 위하여 기존 모델에 비하여 더 높은 신호 대 잡음비(60 dB 이상)를 가지도록 제작되었으므로 야간에 에어로솔과 수증기에 대한 정보를 획득할 수 있다(Barreto *et al.*, 2013). 이 장비는 50% 이상의 달 조도에서 야간 측정을 수행할 수 있으며 달 주기의 50%를 할 수 있어 기존의 주간 관측 대비 야간 관측의 연속성을 크게 확장할 수 있다(Barreto *et al.*, 2016). CE318-T를 이용하여 스페인에서 다른 관측 기기와 비교한 결과는 주간 동안의 AOD 관측값에 대한 불확도가 약 0.002~0.009

범위로 다른 관측 기기의 오차범위값과 유사한 것으로 나타났으며, 야간 동안의 관측 결과는 약 0.011~0.018 범위 내에 있는 것으로 보고되었다(Barreto *et al.*, 2016). 또 다른 집중관측 캠페인의 결과에서는 야간 AOD의 불확실성은 보정 기술에 따라(Langley 보정: ~0.014, moon-langley 보정: 0.012~0.022, 태양-달의 gain 값 보정: 0.1) 최대 0.14으로 보고되었다(Barreto *et al.*, 2019). 이를 통하여 야간 AOD 계산에 사용된 달 조도 모델의 불확실성이 예상 범위인 5~10% 이내임을 증명했다.

본 연구에서는 2015년부터 2022년까지 강릉원주대 관측 지점에서 관측된 자료 중 Version 3 Direct Sun Algorithm의 Level 2 주간자료와 Version 3 Direct Lunar Algorithm Level 1.5 야간자료를 사용하였다. Version 3 알고리즘은 주간의 품질 관리를 높이기 위하여 사용되는 알고리즘이지만, 야간 데이터를 분석할 수 있도록 수정되었다. Level 1.5 산출물은 구름에 의한 영향이 최소화된 산출물로서 광학적으로 얇은 권운에 의한 오차를 줄이기 위한 태양광 산란 측정 자료를 분석하지만, 야간에는 달에 의한 측정값이 부족하므로 권운 구름 오염의 영향을 받을 수 있는 것으로

알려져 있다(Perrone *et al.*, 2022).

### 2.2 머신러닝 기법을 이용한 시계열 모델링

최근까지, 인공지능 기술의 도약으로 인하여 머신러닝 및 딥러닝 기반의 분석 기법을 이용한 대기오염 물질의 농도 및 분포변화를 예측하는 연구 결과가 보고되고 있다(Kim *et al.*, 2023, 2022; Pyo and Lee, 2022; Pyo *et al.*, 2021). 대기 에어로솔의 시계열 분석을 위하여 머신러닝 기법 중 Box-Jenkins 기법을 사용하여 에어로솔 특성에 대한 시계열 분석을 수행하였다. 시계열 분석은 시간의 흐름에 따른 데이터의 패턴과 동향을 이해하고 예측하는 데 사용되는 통계적 기법이며, 본 연구에서는 시계열 데이터를 모델링하고 예측하는 데 널리 사용되는 인공지능 기법을 이용한 모델 중 AutoRegressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average (SARIMA), Seasonal AutoRegressive Integrated Moving Average with Exogenous Variables (SARIMAX)를 사용하였다.

ARIMA는 시계열 데이터를 자기회귀(Auto Regression (AR))와 이동평균(Moving Average (MA))의 두 가지 기본 요소로 설명하는 모델이다(Box *et al.*, 1994). AR은 이전 시점의 데이터가 현재 시점의 데이터에 영향을 미치는 모델이며, MA는 이전의 예측 오차가 현재 시점의 데이터에 영향을 미치는 모델이다. 그리고 ARIMA는 과거 시점에서의 데이터가 가지고 있는 추세성과 오차가 현재에 영향을 미치는 것으로 가정하며, 시계열 데이터의 정상성(Continuity)을 보장하기 위해 차분(Differencing) 과정을 통해 비정상적인 데이터를 정상성을 갖는 데이터로 변환한다. ARIMA의 계산식은 자기회귀 분석에 대해 고려해야 할 과거 관측치의 수(p), 원시 관측치의 차이 횟수(q), 그리고 이동 평균 기간의 크기(d)와 함께 식 (1)과 같이 정의된다(Box *et al.*, 1994).

$$\phi_p(B) \cdot (1 - B)^d Y_t = \theta_q(B) \varepsilon_t \quad (1)$$

위 식에서  $Y_t$ 와  $\varepsilon_t$ 는 특정 시점  $t$ 에서의 관측값과 잡음

(noise)이다. 그리고  $B$ 와  $\phi_p$ ,  $\theta_q$ 는 각각 후행연산자(Backshift operator), AR 연산자, MA 연산자를 의미하며 다음의 식 (2)~(4)와 같이 표현된다.

$$B^k y_t = y_{t-k} \quad (2)$$

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1(B) - \phi_2(B^2) - \dots - \phi_p(B^p) \\ &= 1 - \sum_{i=1}^p \phi_i(B^i) \end{aligned} \quad (3)$$

$$\begin{aligned} \theta_q(B) &= 1 + \theta_1(B) + \theta_2(B^2) + \dots + \theta_q(B^q) \\ &= 1 + \sum_{j=1}^q \theta_j(B^j) \end{aligned} \quad (4)$$

SARIMA는 ARIMA 모델의 확장된 형태로서 ARIMA에서 다루고 있지 않은 계절성 요인(Seasonal parameter)을 포함한 시계열 데이터를 모델링한다(Box *et al.*, 2015). 계절성 요인은 시계열 데이터에서 일정 기간 동안 주기적으로 반복되는 패턴이며, SARIMA에서는 AR, MA, 차분과 계절적인 AR, MA, 차분을 포함하여 시계열 데이터를 모의한다. SARIMA의 계산식은 ARIMA에서 사용되는 p, q, d 외에 추가적으로 계절별 요인을 설명하는 (P, D, Q)s가 사용되어 식 (5)와 같이 정의된다.

$$\begin{aligned} \phi_p(B) \cdot \Phi_P(B^S)(1 - B)^d \cdot (1 - B^S)^P Y_t \\ = \theta_q(B) \cdot \Theta_Q(B^S) \varepsilon_t \end{aligned} \quad (5)$$

위 식에서  $P$ 는 계절별 AR 모델의 순서,  $D$ 는 계절별 차분의 수,  $Q$ 는 계절별 MA의 순서를 나타내고,  $s$ 는 계절의 길이(주기성)를 의미한다. 그리고  $\phi_p$ ,  $\theta_q$ 는 각각 후행연산자(Backshift operator), 계절별 AR 연산자, MA 연산자를 의미하며 다음의 식 (6), (7)과 같이 표현된다.

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1(B) - \phi_2(B^2) - \dots - \phi_p(B^p) \\ &= 1 - \sum_{i=1}^p \phi_i(B^i) \end{aligned} \quad (6)$$

$$\begin{aligned} \theta_q(B) &= 1 + \theta_1(B) + \theta_2(B^2) + \dots + \theta_q(B^q) \\ &= 1 + \sum_{j=1}^q \theta_j(B^j) \end{aligned} \quad (7)$$

SARIMAX는 SARIMA 모델의 확장형으로 단일 시계열 데이터에 미치는 요인이 계절성 요인 외에도 외

생변수(Exogenous Variables)라고 불리는 추가적인 독립변수를 포함하여 모델링한다. 외생변수는 시계열 데이터에 영향을 미치는 다른 요인들을 나타내며, 원 데이터와 동일한 상관관계를 가지는 동일한 차원의 시계열 데이터가 사용된다. SARIMAX 모델은 식 (8)과 같이 표현되며(Manigandan *et al.*, 2021), 여기서  $y_{k,t}$ 는 시간  $t$ 에서의 외생 변수의 수를 의미하고,  $\alpha_k$ 는 외생 변수와의 상관계수이다.

$$\begin{aligned} \Phi_p(B) \cdot \Phi_p(B^s)(1-B)^d \cdot (1-B^s)^D Y_t \\ = \alpha_k y_{k,t} \theta_q(B) \cdot \theta_q(B^s) \varepsilon_t \end{aligned} \quad (8)$$

위에서 언급된 시계열 데이터의 예측을 위한 모델 중 가장 적합한 모델을 찾기 위하여 각 모델별 계산 결과에 대한 정확도 분석이 수행되었으며, 최소 오차 범위를 가지는 모델이 미래 값의 예측에 적합하다는 근본적인 가정을 사용한다. 따라서, 각 모델이 시계열의 변화를 설명하는 데 적합인지 확인하기 위해 본 연구에서는 식 (9)~(13)과 같이 Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE)를 사용하여 시계열의 분산을 설명하는 모델의 성능을 분석하였다. 가장 잘 설계된 모델은 MAE, MASE, RMSE, MAPE 및 SMAPE가 가장 적은 경우로서 모델의 예측능력이 가장 우수한 것으로 판단할 수 있다.

$$MAE = \frac{1}{n} \sum_{i=0}^n |y_i - \bar{y}_i| \quad (9)$$

$$MSE = \frac{1}{n} \sum_{i=0}^n (y_i - \bar{y}_i)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n (y_i - \bar{y}_i)^2} \quad (11)$$

$$MAPE = \frac{100}{n} \sum_{i=0}^n \left| \frac{y_i - \bar{y}_i}{y_i} \right| \quad (12)$$

$$SMAPE = \frac{100}{n} \sum_{i=0}^n \frac{|y_i - \bar{y}_i|}{(|y_i| + |\bar{y}_i|)/2} \quad (13)$$

위 식에서  $y$ 는 관측값,  $\bar{y}$ 는 예측값,  $n$ 은 사용된 자료의

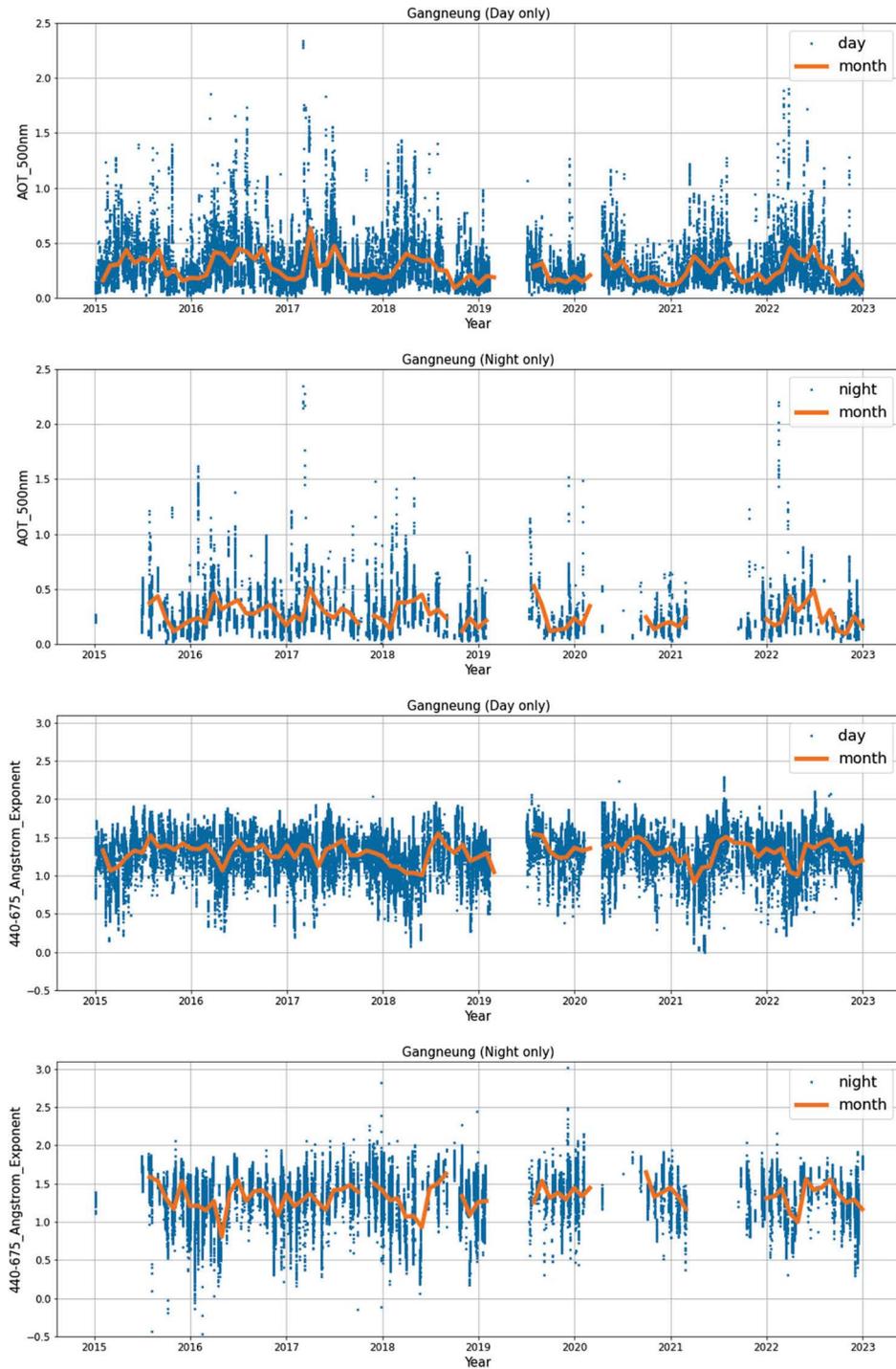
총 개수를 의미한다.

### 3. 결과 및 고찰

#### 3.1 주야간 에어로솔 광학 특성의 장기 변동 특성

CE318-T sun-sky-lunar multiband photometer는 다양한 파장대에서 에어로솔 광학 특성값을 측정할 수 있으나, 본 연구에서는 주간과 야간 관측에서 모두 사용이 가능하며 관측자료의 개수가 가장 많은 AOD<sub>500</sub> (파장 500 nm에서 관측된 에어로솔 광학 두께)와 AE<sub>440\_675</sub> (파장 440 nm와 675 nm에서 산출된 Angstrom exponent)를 사용하였다. 2015년 1월부터 2022년 12월까지 구간과 야간으로 구분된 AOD<sub>500</sub>와 AE<sub>440\_675</sub>의 시계열(시간별 관측, 청색 원, 월 평균: 주황색 실선) 변화량을 그림 2에 묘사하였다. 주간과 야간으로 구분된 AOD<sub>500</sub>의 시계열 변화 특성은 대기 중 에어로솔의 상대적 농도 변화를 이해하는 데 중요한 정보를 제공한다. 일반적으로 주간 동안은 인위적 활동 증가 및 대기의 이동 및 확산에 따라 에어로솔 농도가 높아지는 경향이 있다. 그러나 야간에는 인간 활동이 감소하지만 난방으로 인한 배출활동이 증가하거나 대기 혼합이 감소하여 농축이 될 수 있으므로 복잡한 양상을 띠는 것으로 알려져 있다(Perrone *et al.*, 2015).

전체 관측 기간 동안 주간과 야간의 평균값의 범위는 주간의 평균 AOD<sub>500</sub> = 0.261 ± 0.106, AE<sub>440\_675</sub> = 1.304 ± 0.136이었으며, 야간의 평균 AOD<sub>500</sub> = 0.264 ± 0.105, AE<sub>440\_675</sub> = 1.328 ± 0.163으로 관측되었다. 전체 관측 기간 동안 주간과 야간의 평균 범위는 서로 유사한 범위를 나타내었으나 야간의 에어로솔이 상대적으로 좀 더 큰 AOD<sub>500</sub>와 적은 AE<sub>440\_675</sub>를 가짐으로써, 야간 동안에 다소 적은 크기의 에어로솔 입자에 의하여 AOD<sub>500</sub>가 증가하는 것을 알 수 있다. 그리고 주간과 야간을 모두 고려한 평균값은 AOD<sub>500</sub> = 0.262 ± 0.105, AE<sub>440\_675</sub> = 1.314 ± 0.148으로 주간 관측값만으로 일변화를 대표하는 것과는 차이가 나타나는 것을



**Fig. 2.** Time series of AOD<sub>500</sub>, AE<sub>440-675</sub> for day, night, and combined data. The blue dots and orange line represent the all and monthly mean values.

**Table 1.** Seasonal mean of AOD and Angstrom exponent for day, night, and combined observations. Note that  $R_{AOD}$  and  $R_{AE}$  mean the ratios between observed values at daytime and night/combined, respectively.

Mode	Season	Aerosol optical depth (500 nm)	Angstrom exponent (440~675 nm)	$R_{AOD}$	$R_{AE}$
Day	MAM	0.362	1.184	-	-
	JJA	0.324	1.424	-	-
	SON	0.191	1.316	-	-
	DJF	0.187	1.248	-	-
Night	MAM	0.385	1.166	1.064	0.985
	JJA	0.331	1.466	1.022	1.029
	SON	0.192	1.340	1.005	1.018
	DJF	0.213	1.303	1.139	1.044
Combined	MAM	0.371	1.177	1.025	0.994
	JJA	0.327	1.442	1.009	1.013
	SON	0.192	1.327	1.005	1.008
	DJF	0.199	1.289	1.064	1.033

확인할 수 있다. 이러한 차이는 이전에 언급한 대로, 관측 지역의 에어로솔은 주야간 동안의 태양 복사 프로세스와 대기 역학의 변화에 따른 대기 입자 형성 과정이 다르게 영향을 받고 있는 것으로 설명할 수 있다.

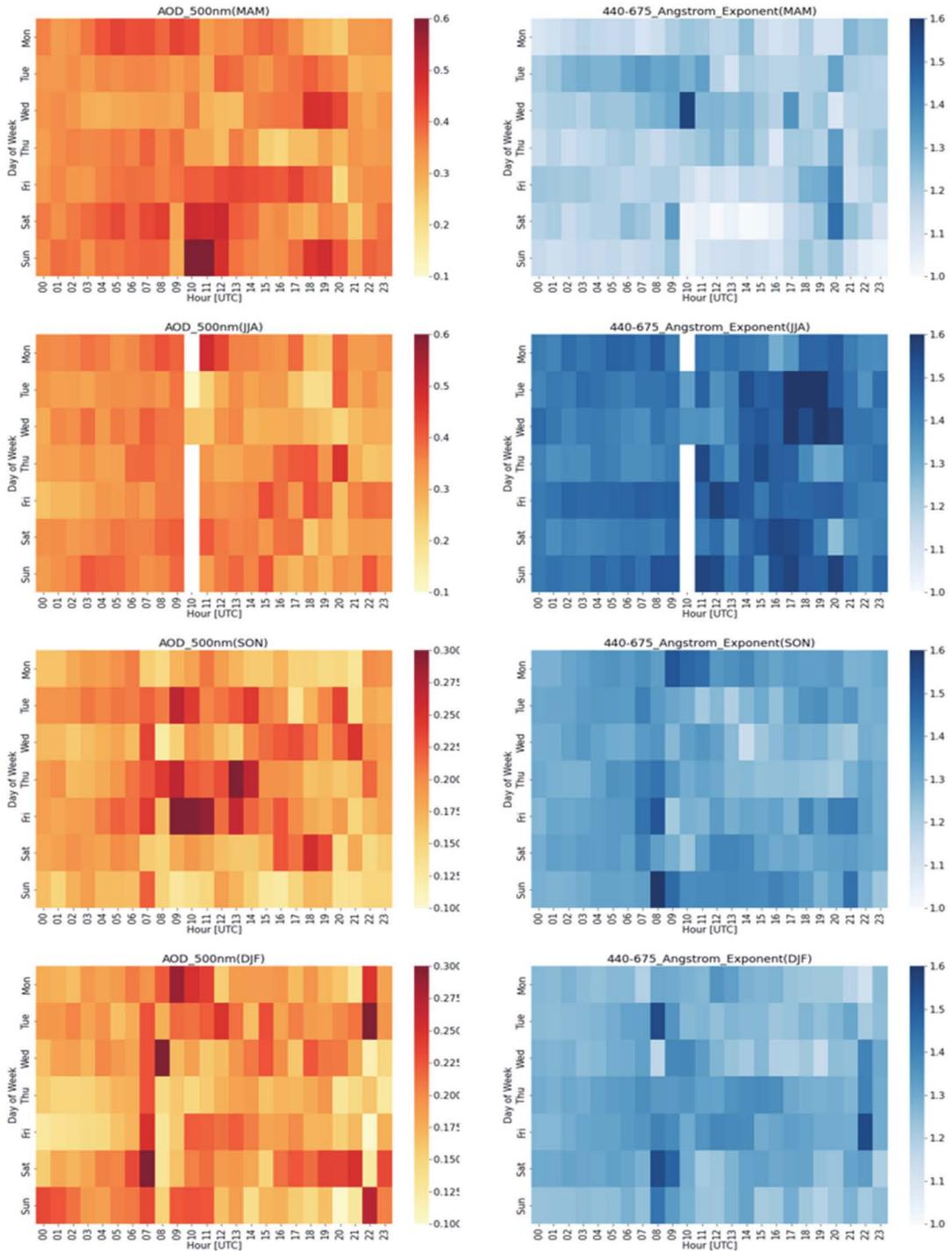
주야간의 에어로솔 특성의 일변화 차이를 비교하기 위하여 계절별 평균값을 표 1에 나열하였다. 각 연도별 에어로솔의 변화 추세는 주로 봄과 여름철에 증가하고 겨울에 감소하는 패턴을 보이며, 봄철의 증가된 AOD는 AE의 감소와 관련이 있다. 이러한 결과는 봄철의 황사와 같은 거대입자의 증가가 지역 내 에어로솔 입자가 증가하는 것과 관계가 있다. 주간 대비 야간의  $AOD_{500}$  관측값 비율 ( $R_{AOD}$ )은 1.005 (가을)~1.139 (겨울),  $AE_{440,675}$  관측값 비율 ( $R_{AE}$ )은 0.985 (봄)~1.044 (겨울)로서 주간 대비 다소 높은 값의 비율을 나타낸다. 그리고 주간 대비 주야간 결합된 관측값 비율 ( $R_{AOD}$ )은  $AOD_{500}$ : 1.005~1.064,  $AE_{440,675}$ : 0.994~1.033로서 주야간 결합된 관측자료의 일변화가 다소 높은 값의 범위를 가지므로 일 주기 대표성에 대해서 야간 관측자료의 기여도가 높은 것을 알 수 있다.

### 3.2 주야간 에어로솔 광학 특성의 일변화

주간과 야간으로 구분된 에어로솔 광학 특성 관측

정보는 에어로솔의 시간적 패턴에 대한 인사이트를 얻을 수 있다. 그림 3은 주간과 야간 관측값을 포함하여 각 요일별 하루 동안의  $AOD_{500}$ ,  $AE_{440,675}$  변화를 나타내는 히트맵(heat map)이다. 보통 교통이나 산업 시설 등의 대규모 배출원이 존재하지 않는 강릉 지역은 주간보다는 야간에  $AOD_{500}$ 가 상승하고, 주간에는  $AOD_{500}$ 가 하강하는 양상을 보이는데, 이러한 일일 패턴은 지역적 배출과 대기 순환 특성의 영향이 반영된 결과로 판단된다.

히트맵의 결과에서는  $AOD_{500}$ 가 가장 큰 요일과 시간대(히트맵의 붉은색으로 표시)가 계절별로 다르게 나타나며, 요일별로도 변동이 있음을 보인다. 전체적인 계절 변화의 양상은 여름을 제외한 나머지 계절의 야간  $AOD_{500}$ ,  $AE_{440,675}$  값의 분포가 주간보다 높은 것을 알 수 있다. 봄철의  $AOD_{500}$ 는 월, 목, 토요일 주간 시간대에 0.4 이상의 큰 값을 보이며, 다른 요일에는 주로 야간 시간대에 상대적으로 큰  $AOD_{500}$  값을 나타낸다. 일요일 저녁인 10~11 UTC 경의 최대값은 주로 큰 입자( $AE_{440,675} < 1.1$ )에 의한 영향을 고려하면, 주로 황사나 해염 또는 봄철에 흔히 발생하는 꽃가루가 원인일 것이다. 여름철에는  $AOD_{500}$ 의 시간대별 요일별 변화가 뚜렷하지는 않지만 일요일과 월요일이 다른 요일에 비하여 다소 높은  $AOD_{500}$ ,  $AE_{440,675}$ 가 관



**Fig. 3.** AOD<sub>500</sub> and AE<sub>440,675</sub> heatmaps by day of week and hour of day for each season. Each block of the heatmap represents a mean value of one day of the week over the eight years study period (MAM: March~April~May, JJA: June~July~August, SON: September~October~November, DJF: December~January~February).

측됨으로 인하여 인위적인 오염원이 영향을 미치고 있는 것과 연관지을 수 있다. 가을철에는 주말보다는 평일 동안의 AOD<sub>500</sub> 값이 오후와 야간에 여러 차례 높은 값과 낮은 AE<sub>440\_675</sub> 값(약 1.0~1.2의 범위)을 나타내고 있다. 겨울철에는 목요일을 제외한 대부분의 평일 오후 및 야간 시간대에 평균 AOD<sub>500</sub> 고점이 나타나며, 일요일에는 오전 시간대에 상대적으로 큰 값이 나타난다.

### 3.3 시계열 분석을 위한 머신러닝 적용

장기간 동안 관측된 AOD<sub>500</sub>, AE<sub>440\_675</sub>는 시계열 자료로서의 계절성과 추세성을 가지고 있음으로써, 시계열 자료의 매개변수 분석 결과를 기반으로 머신러닝 기법 중 시계열 분석 및 예측이 가능한 자기회귀모델을 사용하여 미래의 추세를 예측가능하다. 시계열 분석에 사용된 데이터는 2015년 1월 1일부터 2022년 12월 31일까지 관측된 자료 중 AOD<sub>500</sub>, AE<sub>440\_675</sub>이다. 각 데이터의 시계열 특성을 파악하기 위하여 오픈소스 프로그래밍 언어인 python (version 3.10.9)의 시계열 분리 함수인 `seasonal decompose()`를 이용하여 원본 데이터에서 추세(trend), 계절성(seasonal), 잔차(residual)를 분리하였다.

그림 4는 AOD<sub>500</sub>, AE<sub>440\_675</sub>에 대한 시계열 분리 결과를 나타내며, 위에서 아래 방향으로 각각 원시 자료, 추세, 계절성, 잔차 변화에 대한 시계열 그래프이다. AOD<sub>500</sub>의 추세 변화는 2019년 이전까지 감소하였으며, 이후 점차 증가하고 있는 것과 계절 변화를 통해 AOD<sub>500</sub>의 시계열 자료는 1년을 주기로 양적 변화를 가지고 있는 것을 확인할 수 있다. 잔차 변화는 평균이 약 1이고 분산이 일정하므로 AOD<sub>500</sub> 시계열 자료가 정상성(Stationarity)을 확보하지 못한 것을 시각적으로 확인할 수 있다. 여기서 시계열 자료의 정상성은 평균과 분산이 시간에 따라 일정하게 유지되는 성질을 의미하며, 정상성이 없는 시계열 자료는 추세나 계절성 등의 패턴으로 예측이 어렵게 된다.

AE<sub>440\_675</sub>에 대한 추세 변화에서도 AOD<sub>500</sub>의 경우와 마찬가지로 2019년 이전까지 감소하는 추세(입자

의 크기가 증가하는 추세)와 이후 점차 증가하는(입자의 크기가 감소하는) 패턴이 나타난다. 그리고 계절성에서도 AE<sub>440\_675</sub>는 1년 주기의 계절성 변화를 가지는 것으로 나타난다. 따라서 AOD<sub>500</sub>, AE<sub>440\_675</sub>에 대한 시계열 자료의 변화로부터 2020년 이후로 에어로솔의 양적인 변화가 증가하고 있는 것과 함께 주요 입자 크기는 점차 감소하고 있는 것을 해석할 수 있다.

시계열 예측 모델을 적용하기 위해서는 주어진 시계열 자료의 정상성 확보가 필요하다. 시계열 데이터의 정상성을 확인하기 위하여 AOD<sub>500</sub>와 AE<sub>440\_675</sub>에 대하여 Dickey Fuller Test (DFT)를 수행하였다. DFT는 시계열 데이터의 정상성을 확인하기 위한 통계적 방법으로 이전 시간과의 차이값을 이용하여 계산된 시간 차분값으로 선형회귀 분석 결과에 대한 t-통계량(t-statistic)과 임계값을 비교하여 유의수준(보통  $P < 0.05$ )에 따라 대립가설(정상성을 가짐)을 채택한다. AOD<sub>500</sub>와 AE<sub>440\_675</sub>에 대한 DFT 결과는 p-value가 각각 0.687,  $6.313 \times 10^{-7}$ 을 나타냄으로써 AOD<sub>500</sub> 자료는 정상성을 갖는다고 할 수 없다. 이러한 결과는 이전의 시계열 분해 결과에서 잔차의 범위에 따른 정상성을 설명한 것과 같다. AOD<sub>500</sub> 자료의 정상성 확보는 시계열 자료의 차분(differencing)을 통하여 가능하다. 본 연구에서는 최적의 차분 계수값을 찾기 위하여 1차 차분값에 대한 DFT 결과는 모든 입력값을 모델링한 사례에서 p-value 값이 0.05보다 작았으며, 차분된 데이터는 정상성을 확보하였음을 확인하였다.

식 (1)에서 제시된 ARIMA 모델에 필요한 (p, d, q) 조합을 결정하기 위하여 d는 1로 고정(1차 차분)하고, 개별 p, q 값에 대하여 그리드 탐색(grid searching) 과정을 수행하였다. (p, q)는 각각 0에서 3까지의 조합으로 총 16가지 조합을 사용하여 모델을 수행하여 Akaike Information Criterion (AIC)값을 계산하였다. 일반적으로, 시험 모델의 AIC가 최소인 경우의 조합에서 모델의 성능이 가장 좋은 것으로 알려져 있으며 (Taddy, 2019; Snipes and Taylor, 2014), AOD<sub>500</sub> 자료를 이용한 ARIMA 모델의 최적의 값을 찾기 위하여 탐색한 결과는  $p = 3, d = 1, q = 3$ 인 경우이다. 이때 각각의

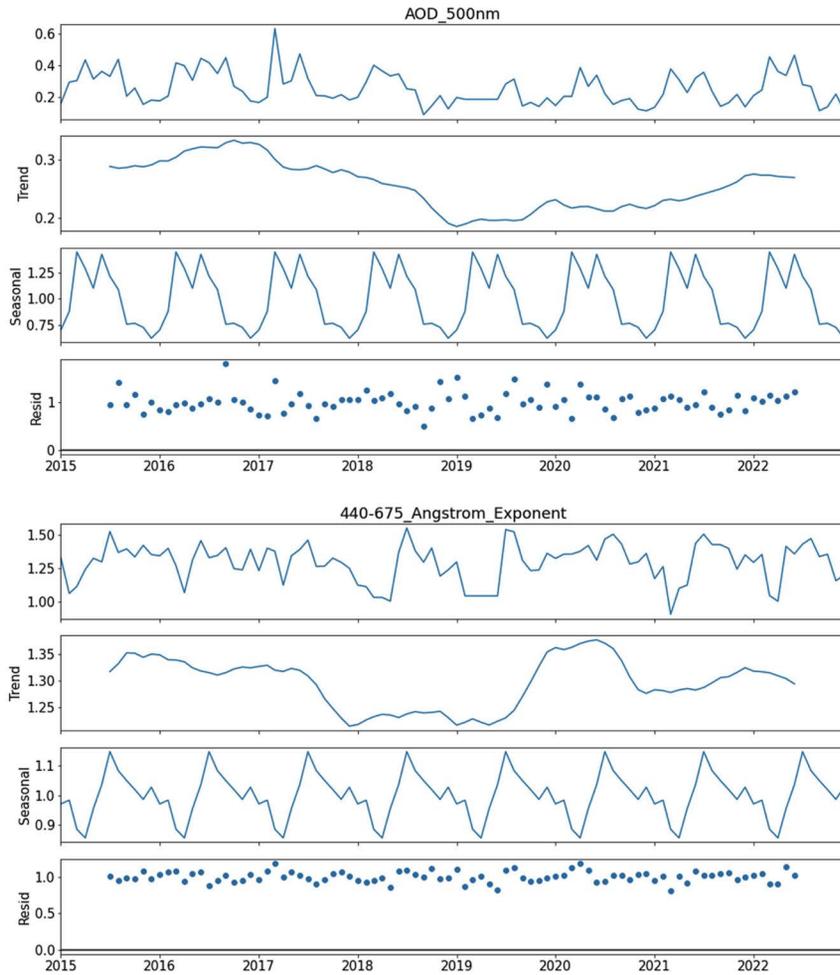
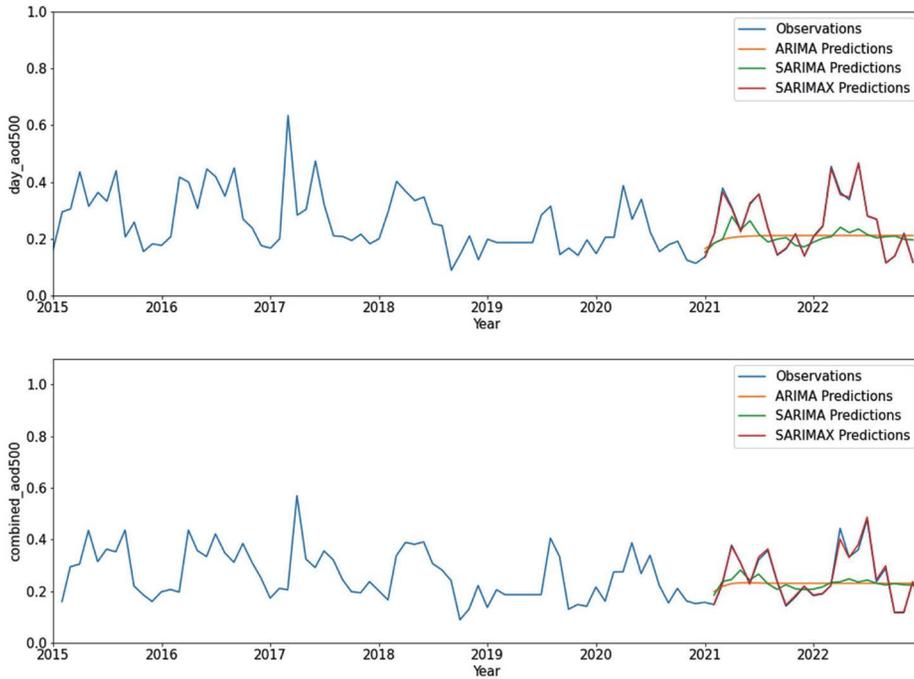


Fig. 4. Time series decomposition results of (top) AOD<sub>500</sub> and (bottom) AE<sub>440-675</sub>.

상수와 변수들의 p-value는 0.05보다 작았으므로 통계적으로 유의미한 값을 가지고 있는 것을 확인할 수 있다. 또한 모든 잔차값이 95% 신뢰 수준 내에 있으며, 잔차가 무작위로 발생하는 백색 잡음(white noise)임을 추론하여 모델의 예측이 잘 맞음을 나타낸다. 마찬가지로 방법으로 AE<sub>440-675</sub>에 대한 매개변수값을  $p=2, d=0, q=2$ 인 경우로 확정하였고, ARIMA 모델의 적용 가능성에 대해 확인하였다.

시계열 자료의 예측에 대한 평가를 수행하기 위해 전체 관측자료는 모델의 학습 및 hyper parameter tuning에 사용하기 위한 학습(Train) 자료와 모델 성

능을 평가하기 위한 시험(Test) 자료로 분리하였다. 그림 5는 2015년부터 2021년까지 기간의 주간 AOD<sub>500</sub> 자료로 학습한 ARIMA ( $p=3, d=1, q=3$ ), SARIMA ( $p=2, d=1, q=1$ ) ( $P=1, D=0, Q=0$ ) [ $m=12$ ], SARIMAX ( $p=2, d=1, q=1$ ) ( $P=1, D=0, Q=0$ ) [ $m=12$ ] 모델과 주야간이 결합된 AOD<sub>500</sub> 자료로 학습한 ARIMA ( $p=3, d=1, q=1$ ), SARIMA ( $p=0, d=1, q=3$ ) ( $P=1, D=0, Q=0$ ) [ $m=12$ ], SARIMAX ( $p=0, d=1, q=3$ ) ( $P=1, D=0, Q=0$ ) [ $m=12$ ] 모델을 통해 2021년부터 2022년까지의 기간의 AOD<sub>500</sub>를 예측한 결과를 동일 기간에 대한 시험용 관측자료와



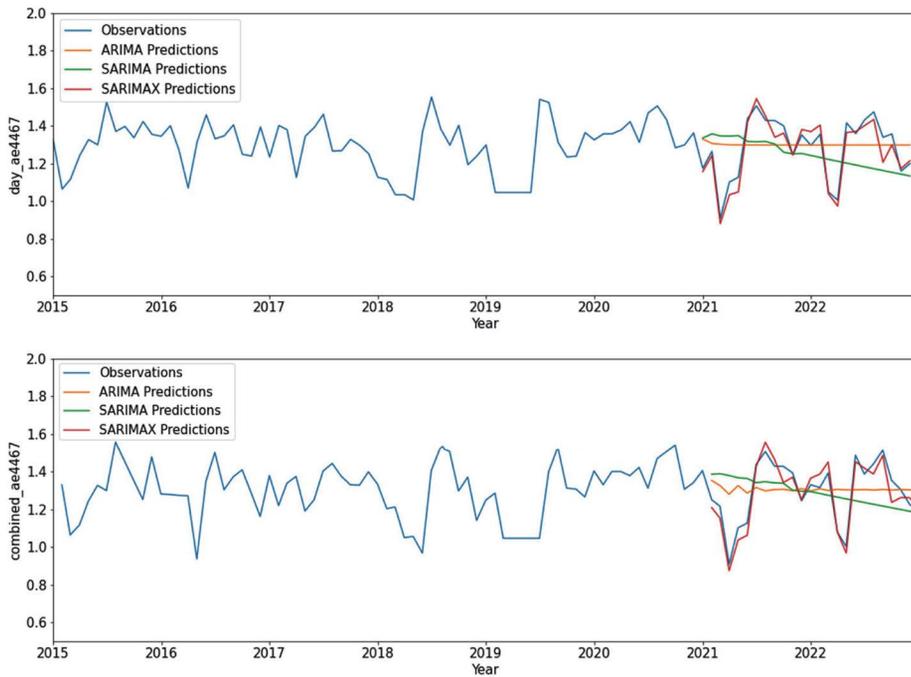
**Fig. 5.** Time series prediction results for (top) daytime and (bottom) combined AOD<sub>500</sub> by the ARIMA, SARIMA, and SARIMAX models.

**Table 2.** Accuracy results for the different time series forecast models. Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE).

Model (data)	MAE	MSE	RMSE	MAPE	SMAPE	BIAS	<Tolerance
ARIMA (day AOD <sub>500</sub> )	0.086	0.012	0.11	33.849	34.901	0.046	37.5
SARIMA (day AOD <sub>500</sub> )	0.077	0.01	0.101	30.238	31.499	0.045	45.833
SARIMAX (day AOD <sub>500</sub> )	4.25E-08	2.87E-15	5.35E-08	1.41E-05	1.41E-05	4.25E-08	100
ARIMA (day AE <sub>440_675</sub> )	0.134	0.026	0.16	11.216	10.641	-0.016	12.5
SARIMA (day AE <sub>440_675</sub> )	0.163	0.035	0.187	13.209	12.959	0.034	8.333
SARIMAX (day AE <sub>440_675</sub> )	4.98E-06	2.51E-11	5.01E-06	3.87E-04	3.87E-04	4.98E-06	100
ARIMA (com AOD <sub>500</sub> )	0.077	0.01	0.1	32.206	30.721	0.025	37.5
SARIMA (com AOD <sub>500</sub> )	0.07	0.009	0.093	29.107	27.739	0.024	50
SARIMAX (com AOD <sub>500</sub> )	1.48E-08	3.80E-16	1.95E-08	5.14E-06	5.14E-06	1.21E-08	100
ARIMA (com AE <sub>440_675</sub> )	0.135	0.026	0.16	11.118	10.622	-0.012	12.5
SARIMA (com AE <sub>440_675</sub> )	0.153	0.034	0.185	12.58	12.051	0.006	16.667
SARIMAX (com AE <sub>440_675</sub> )	4.91E-06	2.45E-11	4.95E-06	3.79E-04	3.79E-04	4.91E-06	100

비교하였다. ARIMA 모델결과는 단순히 학습용 자료의 선형적인 변화 경향을 반영하지만 모델 계산과정에

서 계절성이 반영되지 않기 때문에 계절 변화를 잘 예측해내지 못하는 것을 확인할 수 있다. 그림에도 불구



**Fig. 6.** Time series prediction results for (top) daytime and (bottom) combined  $AE_{440,675}$  by the ARIMA, SARIMA, and SARIMAX models.

하고 시험자료를 이용한 예측 결과는 95% 신뢰구간 안에 대부분의 관측값들이 위치하고 있다. SARIMA 모델결과는 추가적으로 계절성 변화가 고려되었기 때문에 경년 변화에 계절적 변동요인이 반영되어 좀 더 관측값에 근접한 결과를 나타내었다. SARIMAX의 결과는 가장 관측값에 근접한 모델링 결과값을 예측하였으며, 두 값의 편차도 세 가지 모델중에서 가장 최소 값을 나타낸다(표 2). 한편, 주간 기간의  $AOD_{500}$  자료만 학습하여 예측한 결과와 주야간 결합된  $AOD_{500}$ 를 학습하여 예측한 결과를 비교하면, 주야간을 결합한 자료를 사용하였을 때 모델의 예측 성능이 조금 더 나아지는 경향을 확인하였다.

그림 6은 2015년부터 2021년까지 기간의 주간  $AE_{440,675}$  자료로 학습한 ARIMA ( $p=2, d=0, q=2$ ), SARIMA ( $p=1, d=0, q=0$ ) ( $P=0, D=0, Q=1$ ) [ $m=12$ ], SARIMAX ( $p=1, d=0, q=0$ ) ( $P=0, D=0, Q=1$ ) [ $m=12$ ] 모델과 주야간 통합된  $AE_{440,675}$  자료

로 학습한 ARIMA ( $p=2, d=0, q=3$ ), SARIMA ( $p=1, d=0, q=0$ ) ( $P=0, D=0, Q=1$ ) [ $m=12$ ], SARIMAX ( $p=1, d=0, q=0$ ) ( $P=0, D=0, Q=1$ ) [ $m=12$ ] 모델을 통해 2021년부터 2022년까지의 기간의  $AE_{440,675}$ 를 예측한 결과를 나타낸다. ARIMA와 SARIMA 모델 결과는 선형적인 변화 경향만을 확인할 수 있다. 계절성이 추가된 SARIMA 모델에서는 오히려 관측값과의 차이가 커지게 되는데, 이러한 결과는 그림 4에서 나타나듯이 경년 변동이 계절적 변동보다 크게 작용을 한 것이 원인으로 판단된다.  $AOD_{500}$  모델 결과와 마찬가지로, SARIMAX의 결과는 가장 정확한 모델링 결과값을 예측하였다(표 2). 주간 기간의  $AE_{440,675}$  자료만 학습하여 예측한 결과와 주야간 결합된  $AE_{440,675}$ 를 학습하여 예측한 결과를 비교하면, 주야간을 결합한 자료를 사용하였을 때 모델의 예측 성능이 조금 더 개선되는 경향을 확인하였다.

표 2는 주간과 야간 그리고 주야간을 통합한 시계

열 AOD<sub>500</sub>와 AE<sub>440\_675</sub>의 시험자료를 이용하여 3가지 시계열 예측 모델 (ARIMA, SARIMA, SARIMAX)을 이용하여 예측한 결과에 대한 정확도 검증 수치 결과를 요약하였다. 여기서 사용된 검증지수는 RMSE, MAPE, MASE, MAE를 계산하였다. 모든 입력자료와 모델 예측 결과에 대하여 RMSE, MAPE, MASE 및 MAE의 최소값을 가진 모델은 SARIMAX로서 검증 지수의 값이 모두 10<sup>-4</sup>보다 낮은 값의 범위를 나타냄으로써, 모든 입력자료에 대한 예측을 생성하는 데 사용할 수 있는 가장 적합한 모델로 간주된다.

### 3.4 최적 머신러닝 기법을 적용한 시계열 예측

시계열 자료의 분석을 위한 머신러닝 예측 모델의 적정성을 확인한 후, 가장 정확도가 높은 SARIMAX 모델을 이용하여 미래 시점에서의 AOD<sub>500</sub>와 AE<sub>440\_675</sub>에 대한 예측을 수행하였다. 그림 7과 8에는 2015년부터 2022년까지 관측값과 2023년부터 2025년까지 향후

3년 동안의 SARIMAX 예측값 및 예측한계범위를 포함한다. 각 모델의 예측 결과는 각 연도별 계절성과 경년 변화가 반영되어 예측된 결과로서, 모델의 학습에 사용된 입력자료의 특징이 식 (8)에 적용되어 계산된 결과이다.

표 3은 2015년부터 2022년까지 관측값과 2023년부터 2025년까지 향후 3년 동안의 SARIMAX 예측값 및 예측한계범위에 대하여 통계적으로 요약된 결과 (평균, 표준편차, 1차 선형 회귀식에 대한 기울기)에 대한 리스트이다. 2023년부터 2025년까지 예측 결과는 주간 AOD<sub>500</sub>를 제외한 모든 항목에서 음의 기울기 값인 감소추세를 보인다. 2023년부터 2025년까지 기간 동안 예상되는 값은 각각 주간 AOD<sub>500</sub> = 0.2072 ± 0.0835, 주야간 AOD<sub>500</sub> = 0.2156 ± 0.0829, 주간 AE<sub>440\_675</sub> = 1.3203 ± 0.1441, 주야간 AE<sub>440\_675</sub> = 1.3393 ± 0.1473이 될 것으로 예측된다. 그리고 주간과 주야간에 대한 예측 결과를 비교하면, AOD<sub>500</sub> (주간/

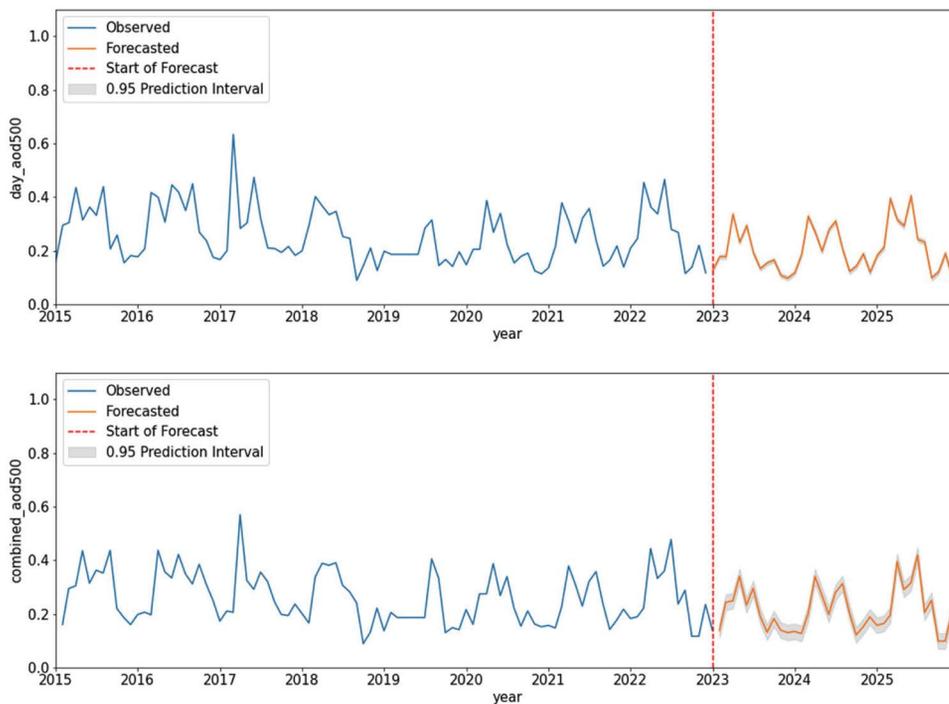


Fig. 7. Time series forecast results for (top) daytime and (bottom) combined AOD<sub>500</sub> by the SARIMAX model.

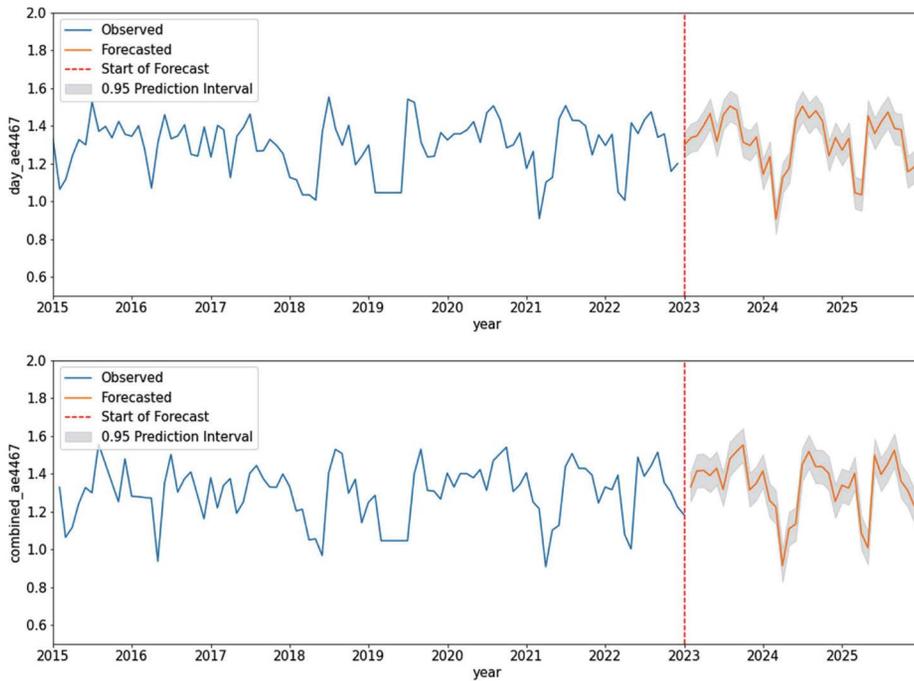


Fig. 8. Time series forecast results for (top) daytime and (bottom) combined  $AE_{440\_675}$  by the SARIMAX model.

Table 3. Summary of the SARIMAX forecasted  $AOD_{500}$  and  $AE_{440\_675}$ . For 2023 = 2025

Data	Mean	Standard deviation	Liner slope
$AOD_{500}$ (day)	0.2072	0.0835	6.934E-04
$AOD_{500}$ (combined)	0.2156	0.0829	-8.369E-05
$AE_{440\_675}$ (day)	1.3203	0.1441	-1.929E-03
$AE_{440\_675}$ (combined)	1.3393	0.1473	-2.413E-03

주야간)는 약 4.08%,  $AE_{440\_675}$  (주간/주야간)는 1.43%의 차이가 발생한다.

$AOD_{500}$ 와  $AE_{440\_675}$ 를 미리 예측할 수 있는 예측 모형을 개발함으로써 얻을 수 있는 시사점은 에어로솔에 대한 변화 추세를 예측하고 이에 대한 영향 평가를 효과적으로 계획할 수 있다는 것이다. 또한, 본 연구에서 사용된 시계열 예측 모델 중 가장 적합한 모델로 확인된 SARIMAX 모델은 기후학, 기후변화 등 다른 연구 분야에도 적용하여 미래 값을 예측할 수 있을 것으로 판단된다.

#### 4. 요약 및 결론

인공지능 기술 중 머신러닝이나 딥러닝 관련 기술의 발전은 매우 빠른 속도로 진행되고 있으나, 이러한 분석 기법이 대기질 예측에 적용된 역사는 비교적 짧은 편이다. 지난 수십 년 동안 축적된 대기질 관측자료는 다양한 관측 플랫폼과 분석기법의 발달로 자료의 질과 양이 크게 증가했다. 이러한 관측자료 기반의 빅데이터와 인공지능 기술 기반의 자료 분석을 통해 대기오염 물질의 거동에 대한 이해가 가능하게 되었다.

본 연구는 주야간 관측이 가능한 CIMEL 루나포토미터를 이용한 강릉 지역에서 장기간 대기 에어로솔의 주야간 관측자료를 이용하여  $AOD_{500}$ 와  $AE_{440\_675}$ 에 대한 확률론적 변화를 분석하였다. 이 과정에서 사용된 머신러닝기반의 시계열 예측 모델인 ARIMA, SARIMA, SARIMAX를 이용하여 개별 입력자료와 모델 수행조건으로부터 최적의 예측 모델과 미래 예측값을 제시하였다.

AOD<sub>500</sub>와 AE<sub>440\_675</sub>에 시계열은 뚜렷한 계절적 변화를 나타내며 연도별 변화 트렌드를 가지는 동적 변화양상을 나타냈다. 따라서 각 자료에 대한 정상성을 확보하기 위해 계절별 차분 변환된 시계열 자료의 ACF와 PACF를 추정하였으며, 유의한 시차를 사용하여 모델의 최적 매개변수(p, q, r)의 조합을 확보하였다. 모델 파라미터에 따라 ARIMA(1, 0, 0) × (2, 1, 2)<sub>12</sub> 모형이 가장 잘 맞는 모형으로 에어로솔 매개변수 예측에 적합한 것으로 나타났다. 또한, 이렇게 최적화된 예측 모형의 유효성을 테스트하기 위해 각 모형의 실제값과 예측값 사이의 통계적 검증지수(RMSE, MAE, MAPE, MASE)를 비교한 결과는 SARIMAX(1, 0, 0) × (2, 1, 2)<sub>12</sub> 모형이 가장 오차가 적은 것으로 나타나 이 모델도 예측 목적에 적합하다는 것을 알 수 있었다.

최적의 예측 모델로 선정된 SARIMA를 이용하여 2023년부터 2028년까지 예측을 수행한 결과는 주간 AOD<sub>500</sub> = 0.2072 ± 0.0835, 주야간 AOD<sub>500</sub> = 0.2156 ± 0.0829, 주간 AE<sub>440\_675</sub> = 1.3203 ± 0.1441, 주야간 AE<sub>440\_675</sub> = 1.3393 ± 0.1473이 될 것으로 예측되었다. 이번에 확인된 최적 모델은 기후학, 기후변화 등 다른 연구 분야에도 적용하여 미래 값을 예측하는 데 활용할 것을 권장한다.

그러나 인공지능 기반의 대기질 예측에는 다음과 같은 한계점이 존재한다. 입력자료의 퀄리티, 대기의 다양성, 모델 계산 시 사용되는 가정 등을 포함하는 불균질한 가변성으로 인해 정밀한 예측이 어렵다. 또한 시계열 변화가 가지는 정상성의 확보가 불가능한 경우에는 예측 품질의 저하는 필연적이다. 이러한 한계점에도 불구하고 인공지능 기반의 대기질 예측은 대기오염에 대한 현황 분석과 비교적 빠르고 효과적인 모델링이 가능한 도구이므로 대기질 연구에서 중요한 역할을 담당할 수 있을 것이다.

## 감사의 글

이 논문은 2019년도 정부(교육부)의 재원으로 한국

연구재단의 지원을 받아 수행된 기초연구사업입니다 (NRF-2019R1I1A3A01062804).

## References

- Barreto, Á., Cuevas, E., Damiri, B., Guirado, C., Berkoff, T., Berjón, A.J., Hernández, Y., Almansa, F., Gil, M. (2013) A New Method for Nocturnal Aerosol Measurements with a Lunar Photometer Prototype, *Atmospheric Measurement Techniques*, 6, 585-598. <https://doi.org/10.5194/amt-6-585-2013>
- Barreto, A., Cuevas, E., Granados-Munoz, M.-J., Alados-Arboledas, L., Romero, P.M., Grobner, J., Kouremeti, N., Almansa, A.F., Stone, T., Toledano, C., Roman, R., Sorokin, M., Holben, B., Canini, M., Yela, M. (2016) The New Sun-sky-lunar CimelCE318-T Multiband Photometer - a Comprehensive Performance Evaluation, *Atmospheric Measurement Techniques*, 9, 631-654. <https://doi.org/10.5194/amt-9-631-2016>
- Barreto, Á., Román, R., Cuevas, E., Pérez-Ramírez, D.J., Berjón, A., Kouremeti, N., Kazadzis, S., Gröbner, J., Mazzola, M., Toledano, C., Benavent-Oltra, J.A., Doppler, L., Juryšek, J., Almansa, A.F., Victori, S., Maupin, F., Guirado-Fuentes, C., González, R., Vitale, V., Goloub, P., Blarel, L., Alados-Arboledas, L., Woolliams, E., Greenwell, C., Taylor, S., Antuña, J.C., Yela, M. (2019) Evaluation of night-time aerosols measurements and lunar irradiance models in the frame of the first multi-instrument nocturnal intercomparison campaign, *Atmospheric Environment*, 202, 190-211. <https://doi.org/10.1016/j.atmosenv.2019.01.006>
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control*, 3<sup>rd</sup> Edition, Prentice-Hall, Englewood Cliffs, NJ.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015) *Time Series Analysis: Forecasting and Control*, 5th Edition, John Wiley & Sons.
- Cadenas, E., Rivera, W. (2010) Wind Speed Sorecasting in Three Different Regions of Mexico, Using a Hybrid ARIMA-ANN Model, *Renewable Energy*, 35, 2732-2738. <https://doi.org/10.1016/j.renene.2010.04.022>
- Chattopadhyay, G., Chattopadhyay, S. (2009) Autoregressive Forecast of Monthly Total Ozone Concentration: a Neurocomputing Approach, *Computational Geosciences*, 35, 1925-1932. <https://doi.org/10.1016/j.cageo.2008.11.007>

- Chaudhuri, S., Dutta, D. (2014) Manne Kendall Trend of Pollutants, Temperature and Humidity Over an Urban Station of India with Forecast Verification Using Different ARIMA Models, *Environmental Monitoring and Assessment*, 186(8), 4719-4742. <https://doi.org/10.1007/s10661-014-3733-6>
- Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Laveanu, F., Jankowiak, I., Smirnov, A. (1998) AERONET- A Federated Instrument Network and Data Archive for Aerosol Characterization, *Remote Sensing of Environment*, 66, 1-16. [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5)
- Huang, B., Liu, M., Ren, Z., Bi, X., Zhang, G., Sheng, G., Fu, J. (2013) Chemical Composition, Diurnal Variation and Sources of PM<sub>2.5</sub> at Two Industrial Sites of South China, *Atmospheric Pollution Research*, 4(3), 298-305. <https://doi.org/10.5094/apr.2013.033>
- Intergovernmental Panel on Climate Change (IPCC) (2014) Anthropogenic and Natural Radiative Forcing. In *Climate Change 2013 - The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 659-740. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.018>
- Kim, K.-J., Lee, S.-H., Chung, Y. (1988) A Time Series Analysis for the Monthly Variation of SO<sub>2</sub> in the Certain Areas of Seoul Using ARIMA Model, *Journal of Korean Society for Atmospheric Environment*, 4(2), 72-81.
- Kim, Y.-I., Lee, K.-H., Lee, K.-T. (2022) Evaluation and Prediction of Column Aerosol by Using the Time Series Machine Learning Technique, *Journal of Korean Society for Atmospheric Environment*, 38(1), 57-73, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2022.38.1.57>
- Kim, Y.-I., Lee, K.-H., Park, S.-H. (2023) Application and Evaluation of Machine Learning Techniques for Real-time Short-term Prediction of Air Pollutants, *Journal of Korean Society for Atmospheric Environment*, 39(1), 107-127, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2023.38.1.107>
- Kripalani, R.H., Kulkarni, A. (2001) Monsoon Rainfall Variations and Teleconnections Over South and East Asia, *International Journal of Climatology*, 21(5), 603-616. <https://doi.org/10.1002/joc.625>
- Lee, K.-H., Lee, K.-T. (2020) Characteristics of Atmospheric Aerosols Based on Column Measurements by Using Machine Learning Clustering, *Journal of Korean Society for Atmospheric Environment*, 36(5), 608-619. <https://doi.org/10.5572/KOSAE.2020.36.5.608>
- Leiterer, U., Naebert, A., Naebert, T., Alekseeva, G. (1995) A New Star Photometer Developed for Spectral Aerosol Optical Thickness Measurements in Lindenberg, *Contributions to Atmospheric Physics*, 68, 133-141.
- Lelieveld, J., Barlas, C., Giannadaki, D., Pozzer, A. (2019) Model Calculations of Aerosol Effects on Local Precipitation: An Analysis for the Region of Cologne, Germany, *Journal of Geophysical Research: Atmospheres*, 124 (5), 2728-2751.
- Manigandan, P., Alam, M.D., Alharthi, M., Khan, U., Alagirisamy, K., Pachiyappan, D., Rehman, A. (2021) Forecasting Natural Gas Production and Consumption in United States-Evidence from SARIMA and SARIMAX Models, *Energies*, 14, 6021. <https://doi.org/10.3390/en14196021>
- Pérez-Ramírez, D., Lyamani, H., Olmo, F.J., Whiteman, D.N., Alados-Arboledas, L. (2012) Columnar Aerosol Properties from Sun-and-Star Photometry: Statistical Comparisons and Day-to-Night Dynamic, *Atmospheric Chemistry and Physics*, 12, 9719-9738. <https://doi.org/10.5194/acp-12-9719-2012>
- Perrone, M.R., Lorusso, A., Romano, S. (2022) Diurnal and Nocturnal Aerosol Properties by AERONET Sun-sky-lunar Photometer Measurements Along Four Years, *Atmospheric Research*, 265, 105889. <https://doi.org/10.1016/j.atmosres.2021.105889>
- Perrone, M.R., Romano, S., Orza, J.A.G. (2015) Columnar and Ground-Level Aerosol Optical Properties: Sensitivity to the Transboundary Pollution, Daily and Weekly Patterns, and Relationships, *Environmental Science and Pollution Research*, 22, 16570-16589. <https://doi.org/10.1007/s11356-015-4850-7>
- Pyo, S.-H., Lee, K.-H. (2022) Estimation of Aerosol Radiative Forcing Using Deep Learning Technique, *Journal of Korean Society for Atmospheric Environment*, 38(5), 669-686, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2022.38.5.669>
- Pyo, S.-H., Lee, K.-H., Lee, K.-T. (2021) Estimation of Column Aerosol Contribution in Seoul and Gangneung Using Machine Learning Clustering Technique, *Journal of Korean Society for Atmospheric Environment*, 37(6), 931-945, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2021.37.6.931>
- Roman, R., Gonzalez, R., Toledano, C., Barreto, A., Perez-Ramírez, D., Benavent-Oltra, J.A., Olmo, F.J., Cachorro, V.E., Alados-Arboledas, L., de Frutos, A.M. (2020) Correction of a Lunar-Irradiance Model for Aerosol Optical Depth

- Retrieval and Comparison with a Star Photometer, *Atmospheric Measurement Techniques*, 13, 6293-6310. <https://doi.org/10.5194/amt-13-6293-2020>
- Snipes, M., Taylor, D.C. (2014) Model Selection and Akaike Information Criteria: An Example from Wine Ratings and Prices, *Wine Economics and Policy*, 3(1), 3-9. <https://doi.org/10.1016/j.wep.2014.03.001>
- Taddy, M. (2019) *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate and Accelerate Business Decisions*, McGraw-Hill Education.
- Tularam, G.A., Ilahee, M. (2010) Time Series Analysis of Rainfall

and Temperature Interactions in Coastal Catchments, *Journal of Mathematics and Statistics*, 6(3), 372-380. <https://doi.org/10.3844/jmssp.2010.372.380>

### Authors Information

이권호 (국립강릉원주대학교 복사위성연구소, 대기환경과학과 교수) (kwonho.lee@gmail.com)