

Research Article

# GOCI-II 대기상한 반사도와 기계학습을 이용한 남한 지역 시간별 에어로졸 광학 두께 산출

양세영 <sup>1)\*</sup> · 최현영 <sup>1)\*</sup> · 임정호 <sup>2)†</sup>

## Retrieval of Hourly Aerosol Optical Depth Using Top-of-Atmosphere Reflectance from GOCI-II and Machine Learning over South Korea

Seyoung Yang <sup>1)\*</sup> · Hyunyoung Choi <sup>1)\*</sup> · Jungho Im <sup>2)†</sup>

**Abstract:** Atmospheric aerosols not only have adverse effects on human health but also exert direct and indirect impacts on the climate system. Consequently, it is imperative to comprehend the characteristics and spatiotemporal distribution of aerosols. Numerous research endeavors have been undertaken to monitor aerosols, predominantly through the retrieval of aerosol optical depth (AOD) via satellite-based observations. Nonetheless, this approach primarily relies on a look-up table-based inversion algorithm, characterized by computationally intensive operations and associated uncertainties. In this study, a novel high-resolution AOD direct retrieval algorithm, leveraging machine learning, was developed using top-of-atmosphere reflectance data derived from the Geostationary Ocean Color Imager-II (GOCI-II), in conjunction with their differences from the past 30-day minimum reflectance, and meteorological variables from numerical models. The Light Gradient Boosting Machine (LGBM) technique was harnessed, and the resultant estimates underwent rigorous validation encompassing random, temporal, and spatial N-fold cross-validation (CV) using ground-based observation data from Aerosol Robotic Network (AERONET) AOD. The three CV results consistently demonstrated robust performance, yielding  $R^2=0.70-0.80$ ,  $RMSE=0.08-0.09$ , and within the expected error (EE) of 75.2–85.1%. The Shapley Additive exPlanations (SHAP) analysis confirmed the substantial influence of reflectance-related variables on AOD estimation. A comprehensive examination of the spatiotemporal distribution of AOD in Seoul and Ulsan revealed that the developed LGBM model yielded results that are in close concordance with AERONET AOD over time, thereby confirming its suitability for AOD retrieval at high spatiotemporal

Received September 30, 2023; Revised October 14, 2023; Accepted October 20, 2023; Published online October 31, 2023

<sup>1)</sup> 울산과학기술원 지구환경도시건설공학과 석·박사과정생(Combined MS/PhD Student, Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea)

<sup>2)</sup> 울산과학기술원 지구환경도시건설공학과 교수(Professor, Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea)

\* The first two authors equally contributed to the paper.

† Corresponding author: Jungho Im (ersgis@unist.ac.kr)

resolution (i.e., hourly, 250 m). Furthermore, upon comparing data coverage, it was ascertained that the LGBM model enhanced data retrieval frequency by approximately 8.8% in comparison to the GOCI-II L2 AOD products, ameliorating issues associated with excessive masking over very illuminated surfaces that are often encountered in physics-based AOD retrieval processes.

**Keywords:** Aerosol optical depth, GOCI-II, Machine learning, Air quality

**요약:** 대기 중 에어로졸은 인체에 악영향을 끼칠 뿐 아니라 기후 시스템에도 직간접적인 영향을 미치므로 에어로졸의 특성과 시공간적인 분포에 대한 이해는 매우 중요하다. 이를 위해 위성기반 관측을 통해 에어로졸 광학 두께(Aerosol Optical Depth, AOD)를 산출하여 에어로졸을 모니터링하는 다양한 연구가 수행되어 왔다. 하지만 이는 주로 조건표를 활용한 역 산출 알고리즘에 기반하여 이루어지기 때문에 많은 계산량을 요구하며 불확실성이 존재한다. 따라서, 본 연구에서는 Geostationary Ocean Color Imager-II (GOCI-II)의 대기상반반사도와 30일 동안의 대기상반반사도 중 최솟값과 관측 시점 값의 차이 값, 수치 모델 기반 기상학적 변수 등을 활용하여 기계학습 기반 고해상도 AOD 직접 산출 알고리즘을 개발하였다. Light Gradient Boosting Machine (LGBM) 기법이 사용되었으며, 추정된 결과는 지상 관측 자료인 Aerosol Robotic Network (AERONET) AOD를 활용하여 랜덤, 시간 및 공간별 N-fold 교차검증을 통해 검증되었다. 세 가지 교차검증 결과  $R^2=0.70-0.80$ ,  $RMSE=0.08-0.09$ , 기대오차(Expected Error, EE) 안에 있는 비율은 75.2-85.1% 수준으로 안정적인 성능을 보였다. Shapley Additive exPlanations (SHAP) 분석에서는 반사도 관련 변수들이 기여도의 상위권 대부분을 차지하고 있는 것을 통해 반사도 자료가 AOD 추정에 많은 기여를 하는 것을 확인하였다. 서울과 울산 지역에 대한 시간별 AOD의 공간 분포를 분석한 결과, 개발된 LGBM 모델은 시간의 흐름에 따라 AERONET AOD 값과 유사한 수준으로 AOD를 추정하고 있었다. 이를 통해 높은 시공간 해상도(i.e., 시간별, 250 m)에서의 AOD 산출이 가능함을 확인하였다. 또한, 산출 커버리지 비교에서 LGBM 모델의 평균 산출 빈도가 GOCI-II L2 AOD 산출물 대비 8.8%가량 증가한 것을 통해 기존 물리모델기반 AOD 산출 과정에서 발생하던 밝은 지표면에 대한 과도한 마스크의 문제점을 개선시킨 것을 확인하였다.

**주요어:** 에어로졸 광학 두께, GOCI-II, 기계학습, 대기질

## 1. 서론

에어로졸은 대기 중에 부유하는 고체 또는 액체 상태의 작은 입자(약 0.001-100  $\mu\text{m}$ )들의 총칭으로, 발생원의 종류에 따라 크게 자연적 발생과 인위적 발생으로 구분된다. 해염 입자, 광진, 바이오 에어로졸 등은 자연적 발생원에 의해 생성되며, 블랙 카본(black carbon), 황산화물, 질소산화물, 그리고 휘발성 유기화합물 등은 주로 인위적 활동을 통해 대기 중으로 배출된다(Kang et al., 2022). 에어로졸은 인체 건강상의 유해성으로 공중보건 문제에도 중요한 역할을 할 뿐 아니라 기후 시스템에도 직간접적으로 영향을 미친다(Mao et al., 2021; Noh, 2021). 직접적으로는 빛의 산란, 흡수를 통해 복사 강제력을 변화시켜 복사 평형에 영향을 미치며, 간접적으로는 구름 응결핵으로 작용하여 알베도(albedo)를 변화시키는 역할을 한다(Charlson et al., 1992). 따라서, 에어로

졸의 광학 및 물리적 특성과 시공간적 분포를 이해하는 것은 매우 중요하다.

에어로졸 광학 두께(Aerosol Optical Depth, AOD)는 대기 중 에어로졸의 흡수와 산란으로 인한 태양 복사 에너지의 감소 정도를 나타내는 값으로, 대기 오염도를 정량화하여 나타내는 지표로 활용된다(Chen et al., 2022b; Wei et al., 2020). AOD는 주로 태양 광도계(sunphotometer), 복사계(radiometer) 등 복사관측장비를 활용한 지상 관측과 항공 관측, 위성 관측을 통해 산출된다(Antuña-Marrero et al., 2018). 지상 관측 장비의 경우 설치 및 관리에 한계가 있어 미관측 지역에 대한 공간적인 불확실성이 존재하기 때문에, 최근 넓은 지역에 대하여 공간적으로 연속적인 정보를 제공할 수 있는 위성기반 AOD 자료가 널리 활용되고 있다(Cao et al., 2023; Kim et al., 2021; Kang et al., 2022; Ranjan et al., 2021).

위성기반 AOD 관측은 극궤도 위성이나 정지궤도

위성 등에 탑재된 위성 센서들을 통해 이루어져왔다. 위성기반 AOD 산출은 주로 조건표(Look-Up Tables, LUTs)를 활용한 역 산출 알고리즘에 기반하여 이루어진다(Lim et al., 2016). LUTs는 특정 에어로졸 및 환경 조건의 범위에 따라 복사전달모델(Radiative Transfer Model, RTM)을 통해 미리 계산하여 데이터화 한 표를 의미한다. 이후, 실시간으로 관측되는 위성 영상의 대기상한 반사도가 입력되었을 때 미리 작성된 LUTs와의 비교를 통해 오차를 최소화하는 최적의 파라미터를 찾고, 이 정보를 활용하여 AOD 값을 산출하게 된다(Levy et al., 2009).

그러나 이와 같은 물리 모델 기반 AOD 산출 알고리즘은 LUTs가 구축될 때 고려된 특정 조건 및 가정에 기반하기 때문에 이 외의 상황에서는 정확도가 떨어질 수 있다는 한계를 가진다. 특히, 육상에서는 지표의 다양한 특성에 따른 반사도 차이가 AOD 산출에 불확실성을 초래한다고 알려져 있다(Kittaka et al., 2011; Remer et al., 2005; Tao et al., 2015). 또한, RTM과 LUTs를 기반으로 한 AOD 산출의 경우 상대적으로 계산량이 많고 불확실성을 내포한다는 한계점을 가지고 있기 때문에, 최근 기계학습을 활용한 AOD를 산출하고자 하는 다양한 시도가 이루어지고 있으며 보다 빠른 계산 속도와 우수한 성능을 보였다(Chen et al., 2020; Kang et al., 2023; Kang et al., 2022). 예를 들어, Kang et al. (2022)은 Geostationary Ocean Color Imager (GOCI) 대기상한 반사도를 활용하여 동아시아 지역에 대하여 기계학습 기반 6 km 해상도의 AOD 산출 알고리즘을 개발하고 이를 기존의 물리모델 기반으로 산출된 MODerate resolution Imaging Spectrometer (MODIS) Dark Target (DT) 및 Deep Blue (DB) 알고리즘 및 GOCI L2 산출물과의 비교를 수행하였다. 그 결과, 제안된 기계학습 기반 알고리즘이 기존 산출물 대비  $R^2$  값을 3-26% 향상시키며 기계학습 기법이 AOD 산출의 성능 향상에 기여할 수 있음을 보여주었다.

그러나 높은 시공간 해상도에서의 기계학습 기반 AOD 산출 및 분석은 여전히 부족한 실정이다. 현재까지의 선행 연구들은 대부분 극궤도 위성을 활용하여 1 km 미만의 높은 공간해상도의 결과를 산출하지만 일 1~2회의 낮은 시간 해상도를 가지거나, 정지궤도 위성을 활용하여 높은 시간해상도로 결과를 산출하지만 5 km 이상의 낮은 공간해상도 특성을 가진다(Chen et al.,

2023). AOD의 시공간적 분포는 대기 중 에어로졸과 밀접한 관련이 있으며, 정확한 대기오염 추정에 고해상도의 에어로졸 정보가 요구되므로 고해상도의 AOD 산출이 중요시되고 있다(Zhao et al., 2019). 따라서, 본 연구에서는 정지 궤도 위성인 GOCI-II 위성의 대기상한 반사도 자료를 활용하여 기계학습 기반의 AOD 직접 산출 알고리즘을 개발하고자 한다. 본 연구는 GOCI-II 위성 자료를 활용하여 선행 연구 대비 높은 시공간 해상도(i.e., 시간별, 250 m)의 AOD 산출 기계학습 모델을 개발하고, 이를 바탕으로한 남한 지역에 대하여 보다 정확한 대기 중 에어로졸 농도 감시를 목표로 한다. 또한, 제안하는 모델의 시공간적 안정성 및 일반화 가능성을 검증하고 모델 내에서 각 입력 변수들의 기여도를 분석하여 AOD와 입력 변수들 간의 상관관계에 대해 확인하고자 한다.

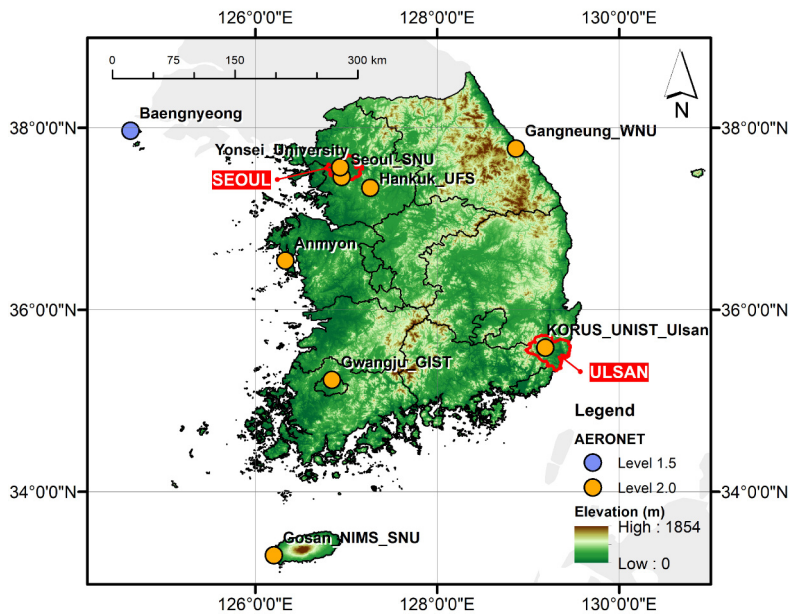
## 2. 연구지역 및 연구자료

### 2.1. 연구지역

본 연구는 남한 지역(33-39°N, 124-131°E)을 대상으로 진행되었으며(Fig. 1), 연구 기간은 2021년부터 2022년까지로 선정하였다. 해당 지역은 지난 수 십년간 급속한 산업화와 경제 성장으로 인한 대기오염물질의 인위적 배출량 급증으로 심각한 대기오염 문제를 겪고 있는 것으로 알려져 있다(Choi et al., 2023). 이에 따라, 최근 이 지역에서 에어로졸의 시공간적 모니터링에 대한 관심 또한 증가하고 있는 추세이다(Lee et al., 2023; Lee et al., 2022b). Fig. 1은 연구 지역과 함께 본 연구에서 사용된 에어로졸 지상 관측소인 Aerosol Robotic Network (AERONET)의 분포를 나타낸다. 본 연구에서는 대기오염물질의 배출량이 높은 도시 두 곳(서울 및 울산)을 추가로 선정하여 분석에 사용하였다. 서울은 대한민국의 수도로서 인구나 경제활동의 집중으로 타도시에 비해 많은 에너지 소비가 이루어지는 곳이다. 울산은 국내 최대 산업도시로서 화석 연료 사용으로 인한 대기오염물질이 대량 배출되고 있어 대기오염이 심각한 지역 중 하나이다.

**Table 1.** Summary of input variables used in the AOD retrieval algorithm proposed in this study

Source	Variables	Abbreviation	Spatial resolution	Temporal resolution
GOCI-II	TOA Reflectance @ 380, 412, 443, 490, 510, 555, 620, 660, 680, 709, 745, 865 nm	TOA <sub>380, 412, 443, 490, 510, 555, 620, 660, 680, 709, 745, 865 nm</sub>	0.25 km	10 times / day
	Differences between reflectance of target day and its minimum value over the past 30 days @ 380, 412, 443, 490, 510, 555, 620, 660, 680, 709, 745, 865 nm	Diff <sub>380, 412, 443, 490, 510, 555, 620, 660, 680, 709, 745, 865 nm</sub>		
	Observation Angles (Satellite Azimuth, Satellite Zenith, Solar Azimuth, Solar Zenith)	PHV, THV, SOLA, SOLZ		
UM-LDAPS	Surface Temperature	Tsrf	1.5 km	8 times / day
	Temperature	Temp		
	1 Hour Maximum Temperature	Tmax		
	1 Hour Minimum Temperature	Tmin		
	Dew point Temperature	Dew		
	Relative Humidity	RH		
	Cosine Value of Wind Direction	Wcos		
	Sine Value of Wind Direction	Wsin		
	1 Hour Maximum Wind Speed	MaxWS		
	Surface Pressure	Psrf		
	Planetary Boundary Layer Height	PBLH		
GPM	Visibility	Visibility	0.1° (~ 10 km)	30 min
	Precipitation	Precip		



**Fig. 1.** Study area with *in-situ* observation (i.e., AERONET) sites. The color-coded dots are indicative of the quality level of the AERONET data, while the background image provides an elevation profile (m) of the study area. The red lines denote the major cities (Seoul and Ulsan) used for the subsequent analysis.

## 2.2. 연구자료

남한 지역에 대하여 시간별 AOD를 산출하기 위해 지상관측소 기반 AOD 자료, 수치 모델 기반 기상자료, 다양한 위성 산출물 등이 함께 활용되었다. 지상관측 AOD 값은 기계학습 모델의 종속 변수(목적 변수)로 사용되었으며, 나머지 자료들은 독립변수(입력 변수)로 활용되었다. 총 41 개의 변수가 AOD 산출 알고리즘 개발을 위해 고려되었다(Table 1).

### 2.2.1. 지상관측 자료

본 연구에서는 남한 지역 내에 존재하는 에어로졸 지상 관측소인 AERONET 관측 AOD 자료를 모델의 종속 변수로 활용하였다(Chen et al., 2022a). AERONET은 National Aeronautics and Space Administration (NASA)와 PHOTométrie pour le Traitement Opérationnel de Normalisation Satellitaire (PHOTONS)이 설립한 지상 기반 원격 감지 에어로졸 네트워크 연합으로 태양 광도계 관측을 통해 에어로졸의 광학적 특성을 산출하여 관측 자료에 대한 보정 및 품질 관리를 거친 데이터를 제공한다(<https://aeronet.gsfc.nasa.gov/>). AERONET은 Level 1.0 (미검수), Level 1.5 (구름 스크리닝 및 품질 관리), 그리고 Level 2.0 (품질 보증)과 같이 세 가지의 데이터 품질 수준에서 AOD 자료를 제공하고 있다. 본 연구에서는 품질 보증이 이루어진 Level 2.0 자료를 우선적으로 사용하였고, 해당 품질 수준의 자료가 제공되지 않는 관측소와 기간에 한해 충분한 데이터 확보를 위해 Level 1.5 자료로 대체 활용하여 남한 내의 총 9 곳의 관측소 자료를 사용하였다(Fig. 1). GOCI-II는 550 nm 파장에서 AOD를 제공하지만, AERONET은 해당 파장에서 AOD 값을 제공하지 않기 때문에, 적절한 비교를 위해 500 nm 파장에서의 AERONET AOD를 다음 식을 통해 보정 과정을 거쳐 사용하였다(Bibi et al., 2015).

$$AOD_c = AOD_a \left(\frac{c}{a}\right)^{-\alpha} \quad (1)$$

식(1)에서  $AOD_c$ 는 보정된 AOD 값,  $AOD_a$ 는 AERONET의 500 nm AOD 값이며  $c$ 는 550 nm,  $a$ 는 500 nm의 파장 값이다.  $\alpha$ 는 440–870 nm 파장대의 옹스트롬 지수(Ångström exponent)를 나타낸다.

### 2.2.2. 위성기반 자료

정지궤도 복합위성 천리안2B호(Geo-Kompsat2B, GK-2B)에 탑재된 정지궤도해양위성 2호(Geostationary Ocean Color Imager-II, GOCI-II)는 GOCI의 임무를 승계 받아, 250 m 해상도로 하루에 10회 한 시간 간격(08–17 KST)으로 동북아시아 지역에 대한 관측을 수행한다. 지역 관측은 총 12개의 슬롯(slot)으로 구분되는데, 본 연구에서는 남한 지역이 속한 슬롯 7을 사용하였다. GOCI-II는 370–865 nm의 분광 범위에서 UV 밴드 1개, 가시광선 밴드 8개, 근적외선 밴드 3개, 광대역폭 밴드 1개, 총 13개의 밴드를 가진다. GOCI-II 관측 자료는 국립해양조사원 국가위성센터(National Ocean Satellite Center, NOSC)에서 다운받을 수 있으며 본 연구에서는 광대역폭 밴드를 제외한 12개 밴드의 대기상한반사도(*i.e.*, 레일리 산란 보정 반사도(Rayleigh-corrected reflectance,  $R_{rc}$ )) 자료를 사용했다(<https://www.nosc.go.kr/main.do>).

또한, 반사도에 영향을 끼치는 태양 천정각, 태양 방위각, 위성 천정각, 위성 방위각 4개의 기하각 정보도 입력 변수로 사용하였다. GOCI-II에는 SWIR 채널이 부족하므로 SWIR 범위의 대기상한반사도에서 가시 범위의 표면 반사도를 계산할 수 없다. 따라서 표면 반사도를 결정하기 위해 최소 반사도 기법을 적용하여 12개 밴드의 반사도에 대해서 관측 시점 이전의 30일 동안 관측된 대기상한반사도 중 최소값과 관측 시점의 대기상한 반사도의 차이 값을 추가로 입력 변수로 사용하였다(Choi et al., 2016). 30일 동안의 최소값이 가지는 의미는 매우 맑은 날 대기 상한의 반사도를 의미하며, 에어로졸과 구름이 증가함에 따라 자체 반사도 증가가 생기기 때문에 차이 값을 통해 에어로졸의 영향을 간접적으로 학습시켜주고자 하였다.

추가로 강수에 의한 에어로졸 습성 침적 효과를 고려하기 위해 Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG)의 강수량 자료를 이용하였다. GPM IMERG는  $0.1 \times 0.1^\circ$ 의 해상도를 가진 Level 3 데이터로 Goddard Earth Science Data and Information Service Center (GES DISC)에서 제공받았다(<https://disc.gsfc.nasa.gov/>). 또한 구름 탐지를 위해 국가기상위성센터에서 제공하는 2분 간격의 2 km 해상도를 지닌 천리안위성 2A호(Geo-KOMPSAT-2A; GK2A)의 차세대 기상영상기(Advanced Meteorological



Imager; AMI)의 구름 탐지 자료가 사용되었다(<https://nmssc.kma.go.kr/homepage/html/main/main.do>).

### 2.2.3. 수치모델 자료

기상학적 요인을 고려하기 위해 기상청에서 제공하는 국지예보모델(Local Data Assimilation and Prediction System, LDAPS)의 자료를 사용하였다. LDAPS는 1.5 km의 공간해상도로 연직으로 약 40 km까지 70 개 층의 정보를 제공하며, 3시간 간격으로 전지구모델로부터 경계장을 제공받아 하루에 8회 예측을 수행한다. 본 연구에서는 선행 연구를 참고하여 지표면온도(Tsrf), 기온(Temp), 최고 및 최저 기온(Tmax, Tmin), 이슬점(Dew), 상대습도(RH), 풍향의 cos, sin값(Wcos, Wsin), 시간 별 최대 풍속(MaxWS), 해면기압(Psrf), 대기경계층 높이(PBLH), 가시성(Visibility) 총 12개의 기상변수를 고려하였다(Che et al., 2019; Kang et al., 2022). 해당 자료는 기상청 기상자료개방포털에서 다운받아 사용하였다(<https://data.kma.go.kr/cmmn/main.do>).

### 2.2.4. 자료 전처리 및 준비

입력 변수들의 시공간 해상도가 다양하기 때문에 GOCI-II의 시공간 해상도로 모델 결과를 생성하기 위해 GOCI-II의 시공간 해상도에 맞춰 리샘플링을 수행하였다. GOCI-II의 공간 해상도와 다른 LDAPS와 GPM 자료는 이중선형 보간법(bilinear interpolation)을 통해 250 m로 리샘플링 해주었다. LDAPS 자료는 세 시간 간격으로 하루 8회 제공되기 때문에 linear 내삽을 이용하여 시간 별 자료를 산출하였다.

본 연구에서 주된 자료로 활용하는 GOCI-II의 경우 열적의 채널이 부재하기 때문에 열린 구름이 고농도 AOD로 오 탐지되는 경우가 발생한다(Lee et al., 2021a). 따라서 보다 정밀한 구름 탐지를 위해 기상위성인 GK2A AMI의 cloud mask를 이용하여 구름으로 탐지된 픽셀들을 직접 제거해주었다. GK2A AMI는 구름 탐지 결과를 픽셀별로 구름(cloud: 0), 구름 확률 높음(probably cloud: 1), 청천(clear: 2)으로 구분하고, 2 km의 해상도로 2 분 간격 자료를 제공한다. 구름 탐지 결과, 구름(0)과 구름 확률 높음(1)으로 판별된 픽셀을 모두 구름이라고 정의한 뒤, 연구 지역(i.e., GOCI-II 슬롯(slot) 7)에서 GOCI-II의 촬영 시간을 고려하여 매시간 25분에서 35분 사이의 자

료 중 한 번이라도 구름으로 탐지된 픽셀은 구름으로 간주하여 제거하였다.

## 3. 연구방법

본 연구는 시간별 250 m의 고해상도로 AOD 농도를 추정하기 위해, 독립 변수로는 GOCI-II, GPM 그리고 LDAPS의 자료를 사용하고 종속 변수로 AERONET의 AOD를 사용하였다. AOD 농도를 추정하기 위해 Light Gradient Boosting Machine (LGBM)이 사용되었으며, 추정된 결과는 N-fold 교차검증을 통해 검증되었다. 또한 각 입력 변수가 모델 결과에 미치는 기여도에 대한 분석과 GOCI-II L2 AOD 산출물과 LGBM 추정 AOD간의 시공간적 분포와 테이터 커버리지 정도에 대한 비교 분석을 수행하였다(Fig. 2).

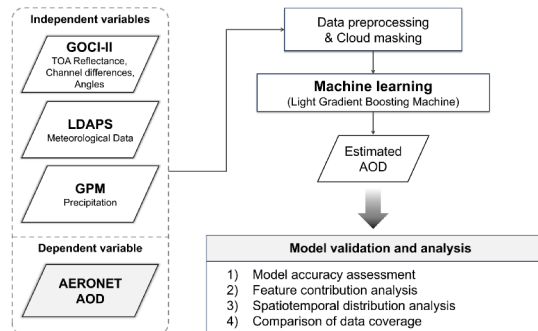


Fig. 2. Process flow diagram of the machine learning-based AOD retrieval proposed in this study.

### 3.1. 기계학습 알고리즘

LGBM은 훈련 데이터를 오류를 줄이는 방향으로 학습해 의사결정 나무를 구성하고 다수의 의사결정 나무들을 결합하여 최종 예측을 얻는 경사 부스팅 알고리즘의 하나이다(Kang et al., 2021). 기존 부스팅 방법인 Gradient Boosting Decision Tree (GBDT) 알고리즘은 입력 자료의 종류와 수가 증가하면 순차적 연산으로 인해 계산 시간이 길어지며 모델의 효율성이 감소하게 된다. LGBM은 이러한 한계를 극복하기 위해 Gradient-based One-Side Sampling (GOSS) 부스팅 방법의 도입과 Exclusive Feature Bundling (EFB)의 적용을 통해 모델의 복잡성을 줄이고 훈련의 효율성을 증가시킨 모델이다(Ke et al., 2017).

본 연구는 Python 3 scikit-learn 환경에서 수행되었고 'lightgbm' 패키지의 LGBMRegressor를 활용하였다. LGBM 모델의 하이퍼파라미터는 베이지안 최적화 (Bayesian Optimization) 알고리즘을 기반으로 한 Python 3.6.11 scikit-learn 모듈에 제공되는 'hyperopt' 라이브러리를 통해 최적화되었다. 그 결과, GOSS 기법의 부스팅 방식을 사용되었으며, max\_depth=9, n\_estimators=900, num\_leaves=400, colsample\_bytree=0.4, 그리고 learning\_rate=0.04의 하이퍼파라미터가 선정되었다.

### 3.2. 검증 방법

본 연구에서 개발된 모델을 평가하기 위해 결정 계수 ( $R^2$ ), 평균 제곱근 편차(Root-Mean-Square-Error; RMSE), 기대오차(Expected Error; EE)의 오차 지표가 사용되었다. 각 오차 지표를 계산하는 수식은 식(2)-(4)에 나타내었다.

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

$$EE_{MODISDT} = \pm (0.15 \times AERONET AOD + 0.05) \quad (4)$$

수식에서  $n$ 은 전체 자료의 개수,  $y_i$ 는 측정값,  $\hat{y}_i$ 는 예측값,  $\bar{y}$ 는 평균값을 의미한다. EE는 AOD 값의 불확실성을 의미하는 값으로, 본 연구에서는 MODIS DT 팀의 기준 값을 사용하여 EE를 계산하였다(Levy et al., 2013; Remer et al., 2005). EE 범위 안에 들어오는 데이터의 비율(i.e., within EE)을 백분율로 나타내었으며, 값이 높을수록 모델의 안정성이 높다고 해석할 수 있다.

본 연구에서는 개발된 모델의 시공간적 안정성을 평가하기 위해 무작위 교차검증(random cross-validation, RDCV), 공간적 교차검증(spatial cross-validation, SPCV), 시간적 교차검증(temporal cross-validation, TPCV) 세 가지의 N-fold 교차 검증을 수행하였다. N-fold 교차검증은 전체 데이터를 N개의 서브 그룹으로 나눈 후, N-1개의 서브 그룹으로 모델을 훈련하고 나머지 1개의 서브 그룹에 대해 검증을 하는 방식을 N번 반복하여 그 결과를 도출하는 검증 방법이다(Choi et al., 2023). RDCV의 경우 전체 학습자료를 무작위로 10개의 서브 그룹으로 나누고, TPCV의 경우 GOCI-II의 관측 시간대별(i.e., 08-17 KST)로 전체 샘플을 10개의 서브 그룹으로 구분

하였다(10-fold CV). SPCV의 경우 활용된 AERONET 관측소별로 9개의 서브 그룹으로 나눠 9-fold CV를 수행하였다.

개발된 모델에서 각 변수의 기여도를 확인하기 위해 설명 가능한 인공지능 (eXplainable Artificial Intelligence, XAI)의 일종인 SHAP (Shapley Additive exPlanations) 분석을 적용하였다. SHAP은 각각의 입력 변수에 대한 SHAP value를 계산함으로써 입력 변수와 모델의 결과 값 사이의 관계를 분석하며, SHAP value는 기계학습 모델의 조건부 기대 값 함수에 대한 샐플리(Shapley) 값으로 정의된다(Lundberg and Lee, 2017; Shapley, 1953). 샐플리 값은 각 특성(입력 변수)에 대한 중요도를 알기 위해 여러 특성들의 조합을 구성하고 해당 특성의 유무에 따른 평균적인 기여도 변화를 통해 얻어낸 값으로, 이를 통해 모델의 예측에 대한 입력 변수의 기여도를 설명하는 것이 가능하다. SHAP은 각 입력 변수들의 중요도를 확인할 수 있을 뿐 아니라 결과에 미치는 영향력의 방향성과 크기까지 확인할 수 있다는 장점이 있다(Jang et al., 2022).

## 4. 연구결과 및 토의

### 4.1. 모델 성능 평가

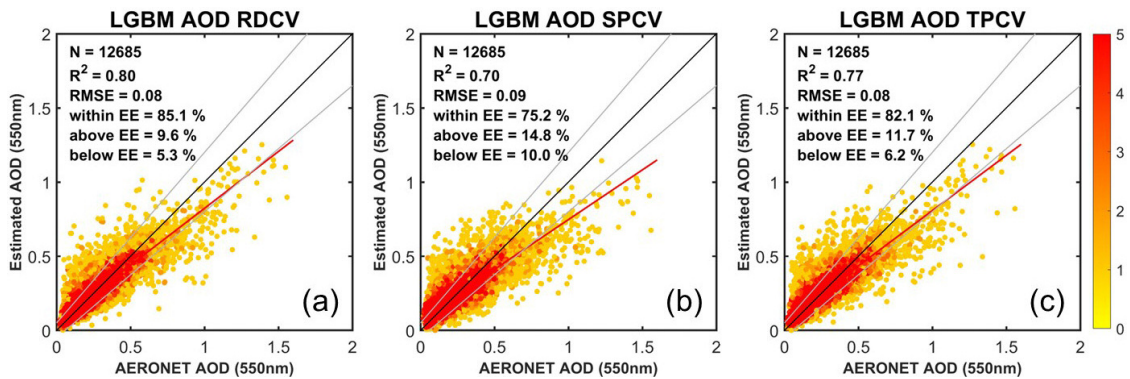
Fig. 3은 2021-2022년에 대하여 구축된 LGBM 모델의 N-fold 교차검증 결과를 나타낸다. LGBM 모델을 통해 추정된 AOD 값과 현장관측자료인 AERONET AOD를 비교한 결과, 세 가지 검증(RDCV, SPCV, TPCV) 측면에서  $R^2=0.70-0.80$ ,  $RMSE=0.08-0.09$ , 그리고 기대오차(EE) 안에 있는 비율이 75.2-85.1% 수준의 정확도를 보이며 현장관측 값과의 높은 일치도를 보였다. 기대오차(EE) 안에 있는 비율 측면에서 Levy et al. (2013)은 MODIS AOD 검증 결과 육지에 대하여 69.4%, 해양에 대하여 76.2%의 결과를 보였으며, Choi et al. (2018)은 GOCI Yonsei Aerosol Retrieval (YAER) AOD 검증 결과 육지에 대하여 60%, 해양에 대하여 71%의 결과를 보였다. MODIS TOA 반사도를 직접 활용하여 기계학습 기반의 AOD를 산출한 Kang et al. (2023)에서는 250 m 공간 해상도로 AOD 산출 시 RDCV, SPCV, TPCV 결과에 따라 69.2-76.9% 수준의 정확도를 보였다. 선행 연구별로

검증을 위해 사용된 데이터와 모델 산출물의 시공간 해상도가 다르기 때문에 직접적인 정확도 비교를 하기에는 어려움이 있지만, 본 연구에서 제안한 모델이 기존 연구들과 유사하거나 더 나은 결과를 보이며 유의미한 AOD 값을 산출함을 확인할 수 있었다.

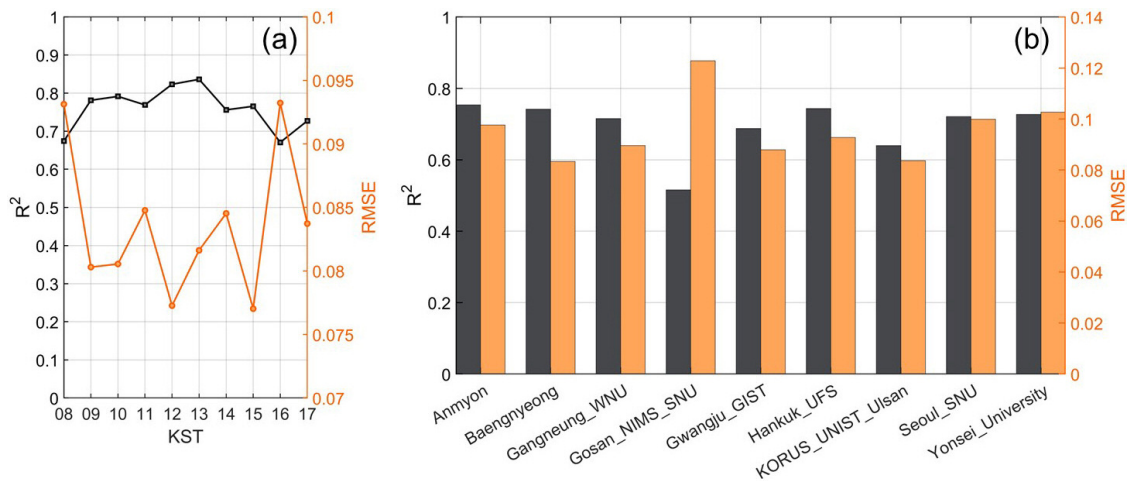
LGBM 모델의 일반화 성능을 평가하기 위해 세 가지 교차검증 방법에 따른 결과를 분석하였을 때, RDCV가 가장 우수한 성능을 나타냈으며 TPCV와 SPCV가 그 뒤를 이었다(Fig. 3). 이는 RDCV의 경우 각 관측소와 시간대 별로 명백한 분리 없이 무작위로 서브 그룹을 나누어 훈련에 활용되어, 훈련 자료에 인근하는 시공간 정보가 검증 자료에도 포함될 수 있기 때문에 나타나는 결과로 해석된다. 반면, TPCV와 SPCV는 각각 시간대와

관측소를 기준으로 분리된 서브 그룹으로 훈련 및 검증 데이터셋이 구성되기 때문에 검증 정확도가 상대적으로 감소하는 것을 확인할 수 있다. 이와 같이 훈련에 포함되지 않은 시간 혹은 공간에 대하여 정확도가 감소하는 경향은 기계학습 기반의 모델에서 흔히 나타나는 특성이다(Choi et al., 2021; Huang et al., 2018).

또한, TPCV와 SPCV에 대하여 각각 N-fold 교차검증에 사용된 서브 그룹별 정확도 비교를 통해 시간대별, 그리고 관측소별 모델 성능을 비교하였다(Fig. 4). 시간대별 결과의 경우 정오 근처에서 가장 높은 정확도를 보였으며 아침 또는 저녁으로 갈수록 정확도가 떨어지는 경향을 보였다. 이는 아침 또는 저녁의 낮은 태양 고도 조건에서는 정지궤도 위성의 관측 반사도 값의 오차를



**Fig. 3.** Scatter plots between *in-situ* observations (x-axis) and AODs (y-axis) estimated by the LGBM model for each cross-validation method: (a) RDCV, (b) SPCV, and (c) TPCV. The black lines represent the one-to-one correspondence, and the grey lines denote the expected error ranges of MODIS DT AOD. The color gradation from yellow to red serves to highlight variations in data density.



**Fig. 4.** Variations of  $R^2$  (black) and RMSE (orange) of each fold within (a) TPCV and (b) SPCV.



증가시켜 입력 변수의 불확실성이 모델 성능 저하로 이어진 것으로 볼 수 있다(Feng et al., 2020; Lee et al., 2021b). 관측소별 결과의 경우 제주도에서 위치한 ‘Gosan\_NIMS\_SNU’ 관측소에서 눈에 띄는 성능 저하가 발견되었는데, 이는 관측소의 위치가 다른 관측소들과 달리 섬 지역에 위치하여(Fig. 1) 국지 오염원의 방해가 적고 인구밀도가 낮은 청정지역이기 때문에 이러한 지리적인 위치에 따른 요인으로볼 수 있다(Lee et al., 2010). 또한, 본 연구에서 활용된 9개의 AERONET 관측소가 일부 지역에 제한적으로 분포되어 있기 때문에 연구 지역의 다양한 공간적인 변동성을 정확히 훈련하지 못해 특정 관측소에서의 정확도가 저하될 수 있다.

#### 4.2. SHAP 분석

Fig. 5는 본 연구에서 제안된 LGBM 모델에 사용된 입력 변수들이 AOD 예측에 미치는 기여도와 방향성을 나타낸다. 변수기여도가 높았던 상위 10 개 변수 중 대부분이 대기상한반사도 혹은 30일 최소반사도와의 차이 값인 것을 통해 위성의 반사도 자료가 에어로졸 광학적 특성을 나타내는 AOD를 추정하는 데 많은 영향을

끼친다는 것을 확인할 수 있다. 특히, 반사도 관련 변수 중 가장 높은 기여도를 보였던  $Diff_{443nm}$ 의 경우 높은 값을 가지는 경우 AOD 예측 값을 증가시키는 방향으로, 낮은 값을 가지는 경우 예측 값을 감소시키는 방향으로 기여하였다. 이는 해당 변수가 에어로졸 및 구름의 증가에 따른 지표면 반사도의 증가, 즉 에어로졸에 의한 반사를 간접적으로 의미하는 변수로써  $Diff_{443nm}$  값이 클 때 AOD 예측 값이 높아지는 방향으로 영향을 미쳤다고 판단할 수 있다. 또한,  $TOA_{380nm}$  변수가 뒤따라 높은 변수기여도를 보였는데 이는 380 nm가 자외선 관측 밴드로 황사 혹은 블랙 카본 등의 흡수성 에어로졸에 대한 민감도가 증가한다는 사실과 일치하는 결과를 보인다(Lee et al., 2021c; Li et al., 2012). 높은 지표면 반사도 환경에서는 AOD 산출의 오차가 커지기 때문에, 눈과 얼음을 제외한 지표에서 낮은 반사도를 갖는 자외선 영역의 관측은 에어로졸 탐지에 이점을 가져 모델에 높은 기여도를 보였을 것이라 판단된다(Penning de Vries et al., 2009). 기상학적 요인으로는 Visibility 변수 값이 낮을 때 AOD 예측 값을 증가시키는 방향으로, 높을 때 AOD 예측 값을 감소시키는 방향으로 작용하며 기상 변수 증가

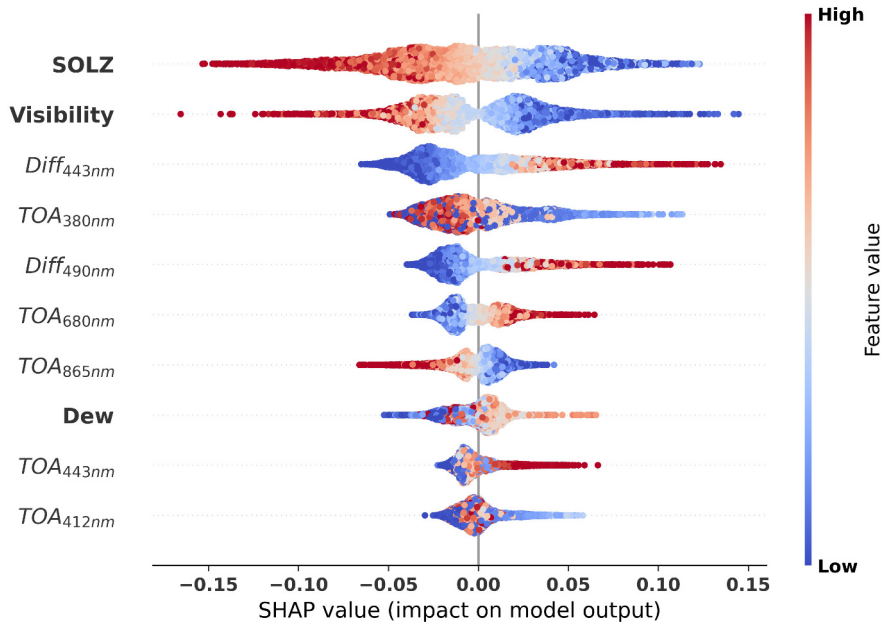


Fig. 5. SHAP summary plots of input variables, ranked by their global feature importance in the LGBM model (Only the top 10 variables of significance are displayed). The x-axis shows the SHAP value that is related to the impact of a variable for each sample on model prediction. Each point represents a specific sample, with varying degrees of feature values illustrated by a color scheme ranging from high (red) to low (blue) values.

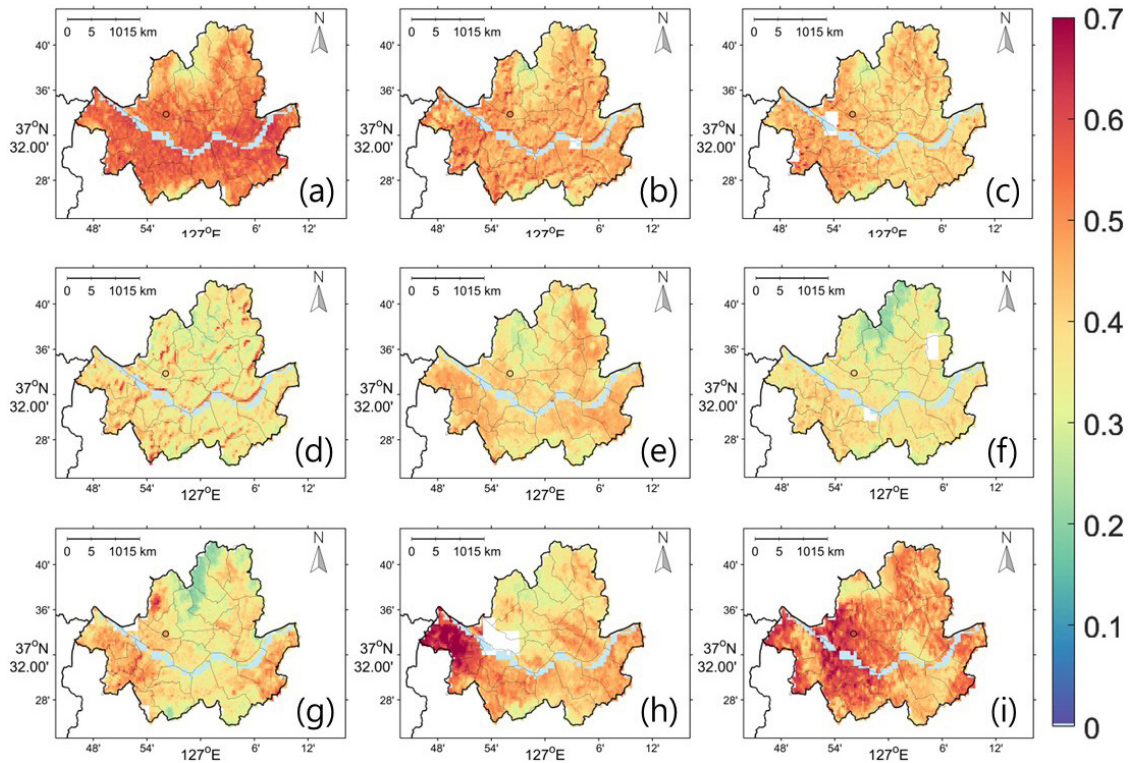
장 높은 기여도를 보여주었다. 에어로졸은 햇빛을 산란 및 흡수시키면서 Visibility를 감소시키므로, 가시성이 낮을 때 AOD 예측 값이 높다는 SHAP의 해석과 일치한다.

### 4.3. 시공간 분포 분석

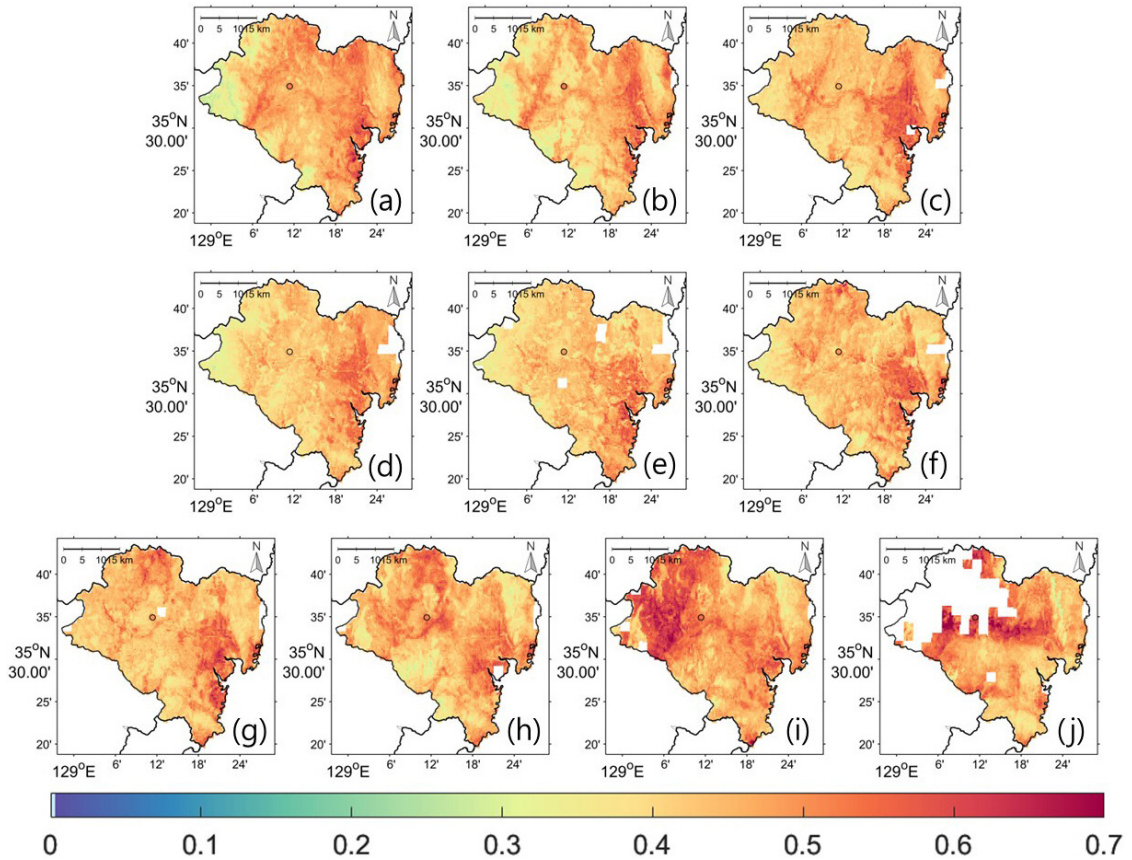
Fig. 6은 2021년 3월 10일 서울 지역에 대하여 본 연구를 통해 추정된 시간별 AOD 농도 분포를 나타낸다. 극 궤도 위성의 에어로졸 관측 횟수가 일 1-2회인 것에 비해 정지궤도 위성인 GOCI-II는 일 10회 관측을 수행하여 하루 동안의 에어로졸 변화를 더 높은 시간 해상도로 분석할 수 있다. 서울 지역은 대한민국의 수도로 인구 집중과 도시화로 인한 대기오염물질이 다량 배출되어 에어로졸의 농도가 다른 지역에 비해 높은 경향이 있다. 시공간 분포 분석을 위해 선택된 날은 평균 0.5의 높은 AOD 값을 보이며 아침 시간대에 높은 값을 보이다 점심 시간대에 낮아진 뒤 다시 높아지는 분포를 보였다. 해당 지역 및 일자에 존재하는 AERONET 관측소인

‘Yonsei\_University’ AOD 관측값과도 일치하며 개발된 모델이 높은 시간 해상도로 AOD 값을 잘 모의하고 있음을 확인하였다.

Fig. 7은 2022년 5월 24일 울산 지역에 대해 LGBM 모델이 추정된 AOD 값의 분포를 나타낸다. 울산은 많은 공장 및 산업단지가 조성되어 있는 지역으로, 대기오염이 심각하여 시간 별 에어로졸 모니터링이 중요한 곳이다. 분석에 활용된 날은 AOD가 평균 0.46으로 높은 농도를 보이며, 17:00 KST에는 0.6 이상의 높은 AOD를 갖는다. Fig. 6에서와 마찬가지로 ‘KORUS\_UNIST\_Ulsan’ 현장 관측소의 AOD값과 비교했을 때 시간의 흐름에 따라 변화하는 AOD 값을 잘 추정하고 있다. 서울과 울산 두 가지 사례를 통해 개발된 모델이 AOD 추정을 안정적으로 수행하고 있음을 확인할 수 있었으며, 높은 시공간 해상도의 AOD 산출이 가능하다는 것 또한 확인하였다.



**Fig. 6.** Spatial distribution of hourly estimated AOD derived by the LGBM model for Seoul during 08:00-17:00 (KST) on March 10, 2021 (There is a data gap at 09:00 (KST) due to a wheel off-loading issue). The area marked with a circle is the ‘Yonsei\_University’ station of AERONET. The date was selected for its low cloud contamination and high AOD concentration.



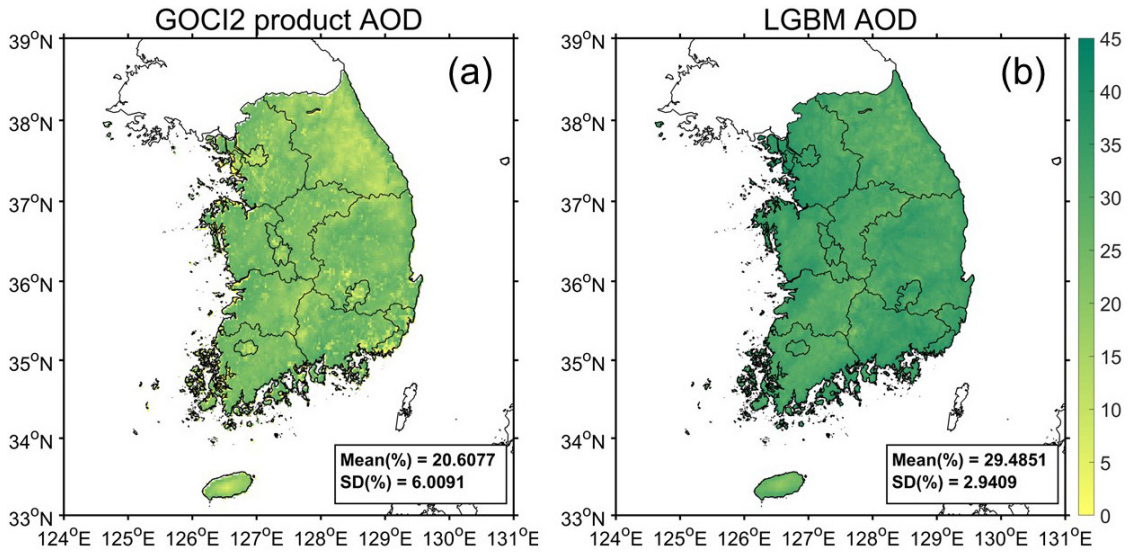
**Fig. 7.** Spatial distribution of hourly estimated AOD derived by the LGBM model for Ulsan during 08:00-17:00 (KST) on May 24, 2022. The area marked with a circle is the 'KORUS\_UNIST\_Ulsan' station of AERONET. The date was selected for its low cloud contamination and high AOD concentration.

#### 4.4. 공간 커버리지 분석

본 연구에서는 위성의 대기상한반사도를 활용하여 LUTs와 RTM을 통한 물리모델기반방법 대신 기계학습 기법을 적용하여 AOD 산출 알고리즘을 개발하였다. 이 과정에서 기상위성인 GK2A 구름 탐지 자료를 활용하여 구름 지역을 마스킹 함으로써 기존의 물리모델기반 AOD 산출 시 밝은 표면(e.g, built-up 또는 사막) 혹은 높은 대기오염 사례에 대하여 과도하게 제거되어 AOD 값이 산출되지 못하는 한계점을 개선시킬 수 있었다. Fig. 8은 물리모델에 기반하여 산출된 GOCI-II L2 AOD 산출물과 본 연구에서 제안된 LGBM 모델을 통해 산출된 AOD의 픽셀 별 산출 커버리지의 공간적 비교를 나타낸다. 산출 커버리지는 전체 연구 기간 내 GOCI-II 관

측이 이루어지는 전체 시간과 각 픽셀 별 AOD 값이 산출되는 시간의 비율을 통해 계산되었다. 전체 연구 지역에 대한 평균 산출 빈도는 GOCI-II L2 AOD 산출물의 빈도인 20.6%에서 29.4%로 8.8%가량 증가한 것을 확인할 수 있었다. 공간적으로는 서울, 대구, 부산 등 대도시 지역에서 크게 향상되었으며 이를 통해 건물 등이 밀집되어 있는 도시 지역에서의 밝은 지표면 반사도에 의해 과도하게 마스킹되는 문제가 개선된 것을 확인할 수 있었다. 이는 도시 지역에서 위성기반 AOD 산출 빈도를 높임으로써 대기질 관련 연구 뿐 아니라 정부의 대기오염 관리 정책에도 유용하게 활용될 수 있음을 시사한다.





**Fig. 8.** Spatial distribution maps representing the coverage (%) of AOD data and standard deviation (SD) based on (a) GOCI-II AOD product and (b) LGBM-derived AOD by pixel from 2021 to 2022. The spatial resolution of the GOCI-II L2 AOD product is 2.5 km, while that of the LGBM-derived AOD is 250 m.

### 5. 결론

본 연구에서는 2021-2022년 남한 지역에 대하여 정지 궤도 위성인 GOCI-II 대기상반사도 자료를 포함한 위성 자료와 수치모델기반 기상학적 변수들을 융합 활용하여 기계학습 기반의 고해상도 AOD 직접 산출 알고리즘을 개발하였다. 기계학습 알고리즘으로는 LGBM 모델이 활용되었으며, 기존 연구 대비 높은 시공간 해상도인 250 m 해상도의 시간별 AOD 분포가 산출되었다. 개발된 모델의 안정성 및 일반화 성능 평가를 위해 랜덤, 공간적, 시간적 N-fold 교차검증을 통해 AERONET AOD와의 비교 결과, 세 가지 검증 측면에서  $R^2=0.70-0.80$ ,  $RMSE=0.08-0.09$ , 그리고 기대오차(EE) 안에 있는 비율이 75.2-85.1% 수준의 정확도를 보이며 본 연구에서 개발된 모델이 유의미한 AOD 분포를 산출할 수 있다는 것을 확인하였다.

SHAP 알고리즘을 활용하여 각 입력 변수가 예측 값에 미치는 영향을 분석한 결과 주로 대기상반사도 관련 변수가 높은 변수기여도를 보였으며, 특히  $Diff_{443nm}$ 는 에어로졸에 의한 반사를 간접적으로 나타내는 변수로써 양의 방향으로 예측에 높은 영향을 미쳤다. 시공간적 고해상도(i.e., 시간별, 250 m) AOD 산출 결과의 이

점을 확인하기 위해 서울 및 울산 지역에 대한 시공간 분포 분석을 수행한 결과 관측 값과 유사한 분포를 보이며 AOD의 공간적인 분포를 파악하는 데에 큰 기여를 할 수 있을 것으로 판단된다. 시공간 해상도 측면 뿐만 아니라 기계학습 기반의 AOD 직접 산출을 통해 공간 산출 커버리지 역시 도시 지역을 중심으로 8.8% 증가 시키며 인간 활동이 활발한 지역에서의 에어로졸 모니터링에 유용하게 활용될 수 있음을 나타냈다.

그러나 위성 관측을 통한 AOD 산출은 여전히 구름에 의한 결측 비율이 높으며 추후 딥러닝 기법 등의 적용을 통한 보간 시 AOD 결측 값 역시 추정이 가능할 것이라 기대된다(Lee et al., 2022a; Zhang et al., 2022). 또한 개발된 모델에 대한 세 가지 교차검증 결과, 무작위 검증 대비 시간 및 공간적 검증 결과가 소폭 감소하는 경향을 보였으며, 추후 개선 가능 여지가 남아있다. 시간적인 측면에서는 GOCI-II 반사도 산출물의 버전 업데이트를 통해 개선될 가능성이 있으며, 공간적인 측면에서는 지상 참조자료로 활용된 AERONET 관측소가 부족하기 때문이므로 추후 연구 지역을 동남아시아 지역까지 확대시키거나 연구 기간을 확장시킴으로써 개선 가능할 것으로 판단된다.



## 사사

본 논문은 환경부의 재원으로 국립환경과학원의 지원을 받아 수행하였습니다(NIER-2023-01-02-131).

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## References

- Antuña-Marrero, J. C., Cachorro Revilla, V., García Parrado, F., de Frutos Baraja, Á., Rodríguez Vega, A., Mateos, D. et al., 2018. Comparison of aerosol optical depth from satellite (MODIS), sun photometer and broadband pyrhelimeter ground-based observations in Cuba. *Atmospheric Measurement Techniques*, 11(4), 2279–2293. <https://doi.org/10.5194/amt-11-2279-2018>
- Bibi, H., Alam, K., Chishtie, F., Bibi, S., Shahid, I., and Blaschke, T., 2015. Intercomparison of MODIS, MISR, OMI, and CALIPSO aerosol optical depth retrievals for four locations on the Indo-Gangetic plains and validation against AERONET data. *Atmospheric Environment*, 111, 113–126. <https://doi.org/10.1016/j.atmosenv.2015.04.013>
- Cao, Z., Luan, K., Zhou, P., Shen, W., Wang, Z., Zhu, W. et al., 2023. Evaluation and comparison of multi-satellite aerosol optical depth products over East Asia ocean. *Toxics*, 11(10), 813. <https://doi.org/10.3390/toxics11100813>
- Charlson, R. J., Schwartz, S. E., Hales, J. M., Cess, R. D., Coakley Jr, J. A., Hansen, J. E. et al., 1992. Climate forcing by anthropogenic aerosols. *Science*, 255(5043), 423–430. <https://doi.org/10.1126/science.255.5043.423>
- Che, H., Gui, K., Xia, X., Wang, Y., Holben, B. N., Goloub, P. et al., 2019. Large contribution of meteorological factors to inter-decadal changes in regional aerosol optical depth. *Atmospheric Chemistry and Physics*, 19(16), 10497–10523. <https://doi.org/10.5194/acp-19-10497-2019>
- Chen, A., Yang, J., He, Y., Yuan, Q., Li, Z., and Zhu, L., 2023. High spatiotemporal resolution estimation of AOD from Himawari-8 using an ensemble machine learning gap-filling method. *Science of the Total Environment*, 857, 159673. <https://doi.org/10.1016/j.scitotenv.2022.159673>
- Chen, B., Ye, Y., Tong, C., Deng, J., Wang, K., and Hong, Y., 2022a. A novel big data mining framework for reconstructing large-scale daily MAIAC AOD data across China from 2000 to 2020. *GIScience & Remote Sensing*, 59(1), 670–685. <https://doi.org/10.1080/15481603.2022.2051382>
- Chen, X., Leeuw, G., Arola, A., Liu, S., Liu, Y., Li, Z. et al., 2020. Joint retrieval of the aerosol fine mode fraction and optical depth using MODIS spectral reflectance over northern and eastern China: Artificial neural network method. *Remote Sensing of Environment*, 249, 112006. <https://doi.org/10.1016/j.rse.2020.112006>
- Chen, Y., Li, D., Karimian, H., Wang, S., and Fang, S., 2022b. The relationship between air quality and MODIS aerosol optical depth in major cities of the Yangtze River Delta. *Chemosphere*, 308, 136301. <https://doi.org/10.1016/j.chemosphere.2022.136301>
- Choi, H., Kang, Y., and Im, J., 2021. Estimation of TROPOMI-derived ground-level SO<sub>2</sub> concentrations using machine learning over East Asia. *Korean Journal of Remote Sensing*, 37(2), 275–290. <https://doi.org/10.7780/KJRS.2021.37.2.8>
- Choi, H., Park, S., Kang, Y., Im, J., and Song, S., 2023. Retrieval of hourly PM<sub>2.5</sub> using top-of-atmosphere reflectance from geostationary ocean color imagers I and II. *Environmental Pollution*,

- 323, 121169. <https://doi.org/10.1016/j.envpol.2023.121169>
- Choi, M., Kim, J., Lee, J., Kim, M., Park, Y. J., Holben, B. et al., 2018. GOCI Yonsei aerosol retrieval version 2 products: An improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia. *Atmospheric Measurement Techniques*, 11(1), 385–408. <https://doi.org/10.5194/amt-11-385-2018>
- Choi, M., Kim, J., Lee, J., Kim, M., Park, Y. J., Jeong, U. et al., 2016. GOCI Yonsei aerosol retrieval (YAER) algorithm and validation during the DRAGON-NE Asia 2012 campaign. *Atmospheric Measurement Techniques*, 9(3), 1377–1398. <https://doi.org/10.5194/amt-9-1377-2016>
- Feng, J., Chen, H., Zhang, H., Li, Z., Yu, Y., Zhang, Y. et al., 2020. Turbidity estimation from GOCI satellite data in the turbid estuaries of China's coast. *Remote Sensing*, 12(22), 3770. <https://doi.org/10.3390/rs12223770>
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A. et al., 2018. Predicting monthly high-resolution PM<sub>2.5</sub> concentrations with random forest model in the North China Plain. *Environmental Pollution*, 242, 675–683. <https://doi.org/10.1016/j.envpol.2018.07.016>
- Jang, E., Kim, Y. J., Im, J., Park, Y. G., and Sung, T., 2022. Global sea surface salinity via the synergistic use of SMAP satellite and HYCOM data based on machine learning. *Remote Sensing of Environment*, 273, 112980. <https://doi.org/10.1016/j.rse.2022.112980>
- Kang, E., Park, S., Kim, M., Yoo, C., Im, J., and Song, C. K., 2023. Direct aerosol optical depth retrievals using MODIS reflectance data and machine learning over East Asia. *Atmospheric Environment*, 309, 119951. <https://doi.org/10.1016/j.atmosenv.2023.119951>
- Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C. K. et al., 2021. Estimation of surface-level NO<sub>2</sub> and O<sub>3</sub> concentrations using TROPOMI data and machine learning over East Asia. *Environmental Pollution*, 288, 117711. <https://doi.org/10.1016/j.envpol.2021.117711>
- Kang, Y., Kim, M., Kang, E., Cho, D., and Im, J., 2022. Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over East Asia. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, 253–268. <https://doi.org/10.1016/j.isprsjprs.2021.11.016>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W. et al., 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 2017 Advances in Neural Information Processing Systems*, Long Beach, CA, USA, Dec. 4–9, pp. 3146–3154. <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>
- Kim, G., Lee, S., Im, J., Song, C. K., Kim, J., and Lee, M. I., 2021. Aerosol data assimilation and forecast using geostationary ocean color imager aerosol optical depth and in-situ observations during the KORUS-AQ observing period. *GIScience & Remote Sensing*, 58(7), 1175–1194. <https://doi.org/10.1080/15481603.2021.1972714>
- Kittaka, C., Winker, D. M., Vaughan, M. A., Omar, A., and Remer, L. A., 2011. Intercomparison of column aerosol optical depths from CALIPSO and MODIS-Aqua. *Atmospheric Measurement Techniques*, 4(2), 131–141. <https://doi.org/10.5194/amt-4-131-2011>
- Lee, C., Lee, K., Kim, S., Yu, J., Jeong, S., and Yeom, J., 2021a. Hourly ground-level PM<sub>2.5</sub> estimation using geostationary satellite and reanalysis data via deep learning. *Remote Sensing*, 13(11), 2121. <https://doi.org/10.3390/rs13112121>
- Lee, K. S., Ahn, J. H., and Park, M. S., 2021b. GOCI-II capability of improving the accuracy of ocean color products through fusion with GK-2A/AMI.

- Korean Journal of Remote Sensing*, 37(5–2), 1295–1305. <https://doi.org/10.7780/kjrs.2021.37.5.2.7>
- Lee, K., Kim, M., Choi, M., Kim, J., Choi, Y., Jeong, J. et al., 2022a. Fast and operational gap filling in satellite-derived aerosol optical depths using statistical techniques. *Journal of Applied Remote Sensing*, 16(4), 044507. <https://doi.org/10.1117/1.JRS.16.044507>
- Lee, S. B., Kang, C. H., Jung, D. S., Ko, H. J., Kim, H. B., Oh, Y. S. et al., 2010. Composition and pollution characteristics of TSP, PM<sub>2.5</sub> atmospheric aerosols at Gosan site, Jeju Island. *Analytical Science and Technology*, 23(4), 371–382. <https://doi.org/10.5806/AST.2010.23.4.371>
- Lee, S. J., Lee, H. Y., Kim, S. J., Kang, H. J., Kim, H., Seo, Y. K. et al., 2023. Pollution characteristics of PM<sub>2.5</sub> during high concentration periods in summer and winter in Ulsan, the largest industrial city in South Korea. *Atmospheric Environment*, 292, 119418. <https://doi.org/10.1016/j.atmosenv.2022.119418>
- Lee, S., Kim, J., Ahn, J. H., Lim, H., and Cho, Y., 2021c. Exploiting GOCI-II UV channel to observe absorbing aerosols. *Korean Journal of Remote Sensing*, 37(6–1), 1697–1707. <https://doi.org/10.7780/kjrs.2021.37.6.1.17>
- Lee, S., Park, S., Lee, M. I., Kim, G., Im, J., and Song, C. K., 2022b. Air quality forecasts improved by combining data assimilation and machine learning with satellite AOD. *Geophysical Research Letters*, 49(1), e2021GL096066. <https://doi.org/10.1029/2021GL096066>
- Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F. et al., 2013. The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 6(11), 2989–3034. <https://doi.org/10.5194/amt-6-2989-2013>
- Levy, R. C., Remer, L. A., Tanré, D., Mattoo, S., and Kaufman, Y. J., 2009. Algorithm for remote sensing of tropospheric aerosol over dark targets from MODIS: Collections 005 and 051: Revision 2 (Product ID: MOD04/MYD04). National Aeronautics and Space Administration. [https://modis-images.gsfc.nasa.gov/\\_docs/ATBD\\_MOD04\\_C005\\_rev2.pdf](https://modis-images.gsfc.nasa.gov/_docs/ATBD_MOD04_C005_rev2.pdf)
- Li, Q., Li, C., and Mao, J., 2012. Evaluation of atmospheric aerosol optical depth products at ultraviolet bands derived from MODIS products. *Aerosol Science and Technology*, 46(9), 1025–1034. <https://doi.org/10.1080/02786826.2012.687475>
- Lim, H., Choi, M., Kim, M., Kim, J., and Chan, P. W., 2016. Retrieval and validation of aerosol optical properties using Japanese next generation meteorological satellite, Himawari-8. *Korean Journal of Remote Sensing*, 32(6), 681–691. <https://doi.org/10.7780/kjrs.2016.32.6.12>
- Lundberg, S. M., and Lee, S. I., 2017. A unified approach to interpreting model predictions. In *Proceedings of the 2017 Advances in Neural Information Processing Systems*, Long Beach, CA, USA, Dec. 4–9, pp. 4765–4774. <https://dl.acm.org/doi/pdf/10.5555/3295222.3295230>
- Mao, W., Jiao, L., Wang, W., Wang, J., Tong, X., and Zhao, S., 2021. A hybrid integrated deep learning model for predicting various air pollutants. *GIScience & Remote Sensing*, 58(8), 1395–1412. <https://doi.org/10.1080/15481603.2021.1988429>
- Noh, Y., 2021. A Long-term trend of annual average aerosol optical properties in the Korean peninsula using AERONET sun/sky radiometer data. *Journal of Korean Society for Atmospheric Environment*, 37(3), 456–465. <https://doi.org/10.5572/KOSAE.2021.37.3.456>
- Penning de Vries, M. J. M., Beirle, S., and Wagner, T., 2009. UV aerosol indices from SCIAMACHY: Introducing the scattering Index (SCI). *Atmospheric Chemistry and Physics*, 9(24), 9555–9567. <https://doi.org/10.5194/acp-9-9555-2009>
- Ranjan, A. K., Patra, A. K., and Gorai, A. K., 2021. A

- review on estimation of particulate matter from satellite-based aerosol optical depth: Data, methods, and challenges. *Asia-Pacific Journal of Atmospheric Sciences*, 57, 679–699. <https://doi.org/10.1007/s13143-020-00215-0>
- Remer, L. A., Kaufman, Y. J., Tarré, D., Mattoo, S., Chu, D. A., Martins, J. V. et al., 2005. The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences*, 62(4), 947–973. <https://doi.org/10.1175/JAS3385.1>
- Shapley, L. S., 1953. A value for n-person games. In: Kuhn, H., and Tucker, A. (eds.), *Contributions to the theory of games (AM-28), Volume II*, Princeton University Press, pp. 307–318. <https://doi.org/10.1515/9781400881970-018>
- Tao, M., Chen, L., Wang, Z., Tao, J., Che, H., Wang, X. et al., 2015. Comparison and evaluation of the MODIS collection 6 aerosol data in China. *Journal of Geophysical Research: Atmospheres*, 120(14), 6992–7005. <https://doi.org/10.1002/2015JD023360>
- Wei, X., Chang, N. B., Bai, K., and Gao, W., 2020. Satellite remote sensing of aerosol optical depth: Advances, challenges, and perspectives. *Critical Reviews in Environmental Science and Technology*, 50(16), 1640–1725. <https://doi.org/10.1080/10643389.2019.1665944>
- Zhang, T., Zhou, Y., Zhao, K., Zhu, Z., Asrar, G. R., and Zhao, X., 2022. Gap-filling MODIS daily aerosol optical depth products by developing a spatiotemporal fitting algorithm. *GIScience & Remote Sensing*, 59(1), 762–781. <https://doi.org/10.1080/15481603.2022.2060596>
- Zhao, C., Liu, Z., Wang, Q., Ban, J., Chen, N. X., and Li, T., 2019. High-resolution daily AOD estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region. *Atmospheric Environment*, 203, 70–78. <https://doi.org/10.1016/j.atmosenv.2019.01.045>