Direct retrievals of aerosol optical depth using MODIS reflectance data and machine learning over East Asia

EGU General Assembly 2023

Eunjin Kang[1], Seonyoung Park[2], Miae Kim[3], Cheolhee Yoo[4], Jungho Im[1]*

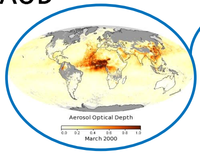
1. Department of Urban & Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea
2. Department of Applied Artificial Intelligence, Seoul National University of Science and Technology, Seoul, Republic of Korea
3. Prediction Research Department Climate Services and Research Division APEC Climate Center (APCC), Busan, South Korea
4. Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung 12 Hom, Kowloon, Hong Kong

(Under Review: ATMENV-D-23-00272)

Background

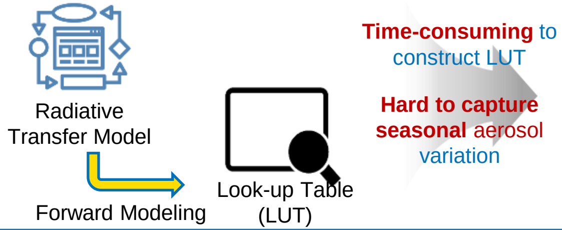
- How to monitor AOD

Ground-based stations

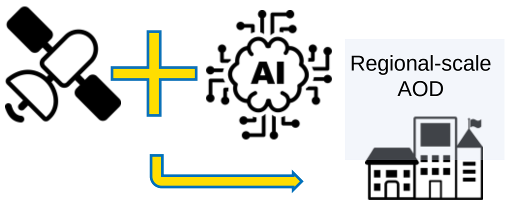


Satellite product

- Previous AOD retrieval



- Proposed AOD retrieval



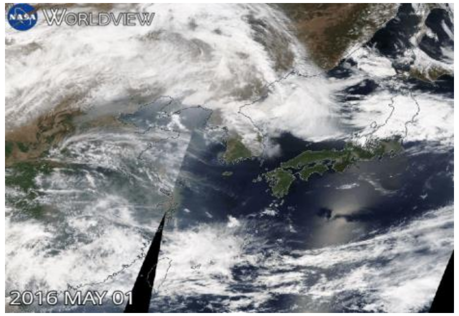
Conclusion

✓ **High-resolution AODs** were retrieved directly using **MODIS TOA reflectance data and machine learning** with **excellent quality** and **reduced missing rates**.

✓ **Sensor zenith angle** and **TOA blue reflectance** were identified as key contributors through **XAI**.

Methodology

- Study Area & Period



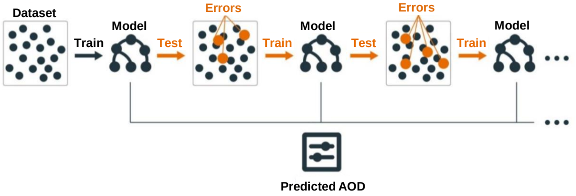
: East Asia, including northeastern China, Korea, and Japan during 2016-2019

- Data

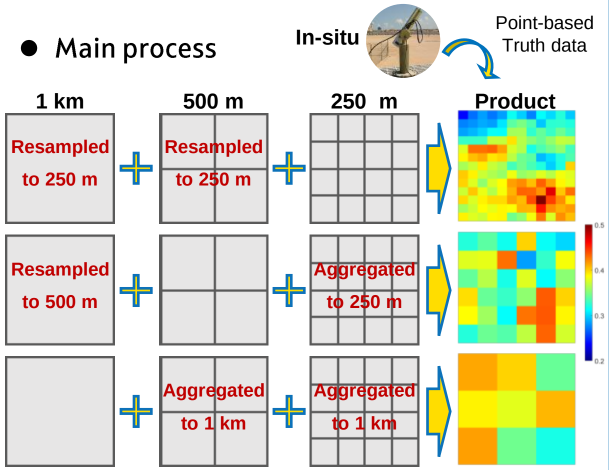
Scale	Source	Name
250 m	MOD02QKM	Band 1,2
500 m	MOD02HKM	Band 3,4,7
1 km	MOD021KM	Band 8, 17, 18, 19, 26
	MOD03	SEA, SEZ, SOA, SOZ

- Machine learning

: Light Gradient Boosting Model (LightGBM)



- Main process



Result & Discussion

- Significant accuracy compared to MAIAC!

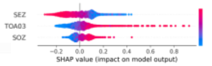
Independent test: Korea-United States Air Quality campaign data (May to June 2016)

N = 558	250 m	500 m	1 km	MAIAC
R	0.86	0.87	0.84	0.88
RMSE	0.084	0.083	0.095	0.097
Within EE (%)	77.8	76.3	76.5	70.8
IOA	0.93	0.93	0.91	0.93

- Explainable AI to analyze variable

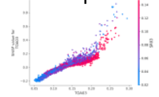
SHAP

SHAP summary plot



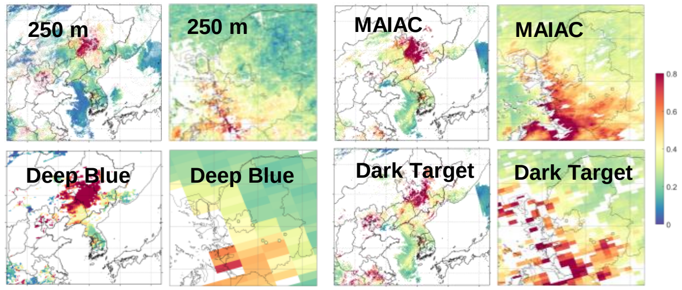
For variable contribution to the model

SHAP dependency plot

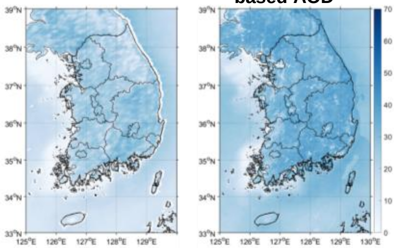


For interaction between variables

- Detailed spatial resolution of 250 m product



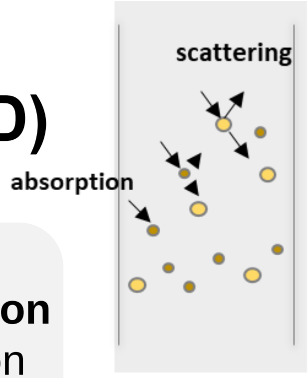
- Higher frequency than MAIAC AOD



Introduction

● Aerosol Optical Depth (AOD)

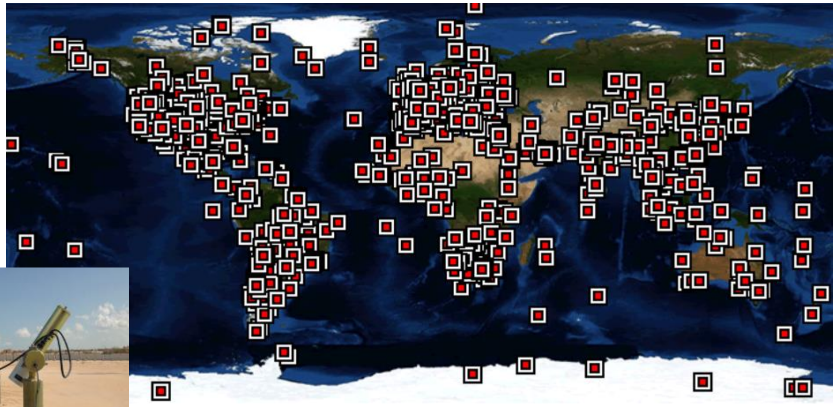
A measure of **vertically integrated extinction (scattering + absorption)** of solar radiation by atmospheric aerosol particles



Low concentration of AOD

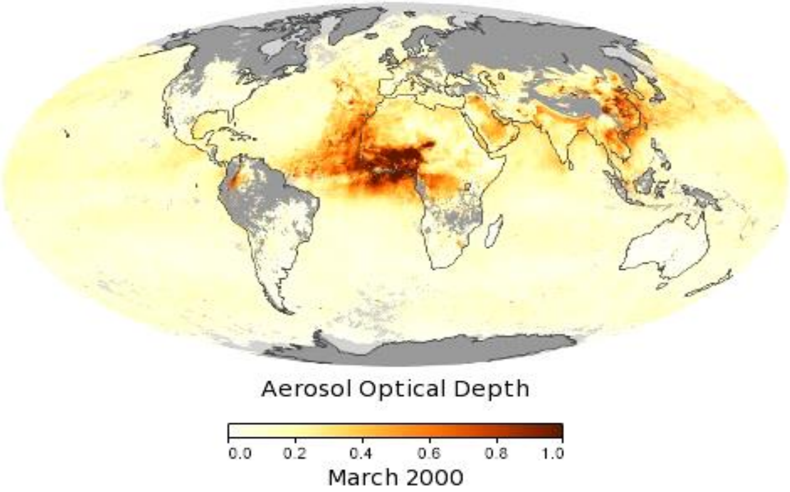
High concentration of AOD

1. Ground-based Measurements



Spatial distribution of AOD ground monitoring station (AERONET)

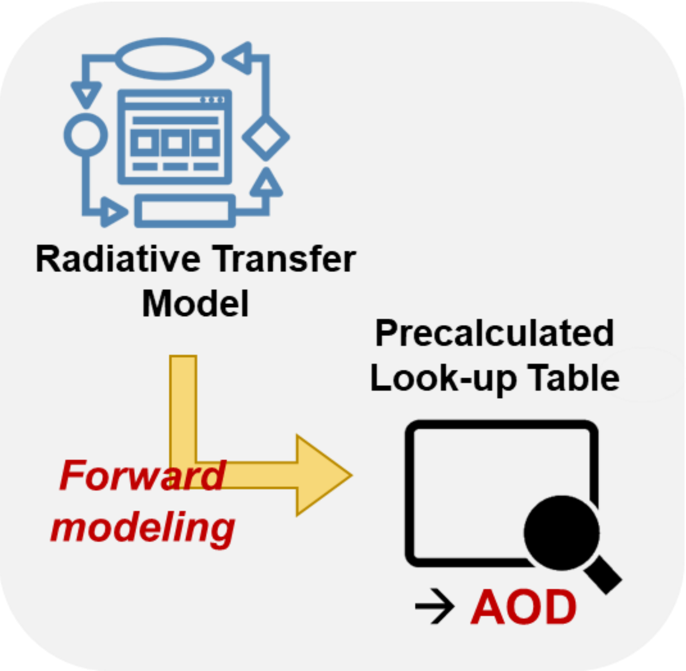
2. Satellite product (Applying Look-up Table & Radiative Transfer Model)



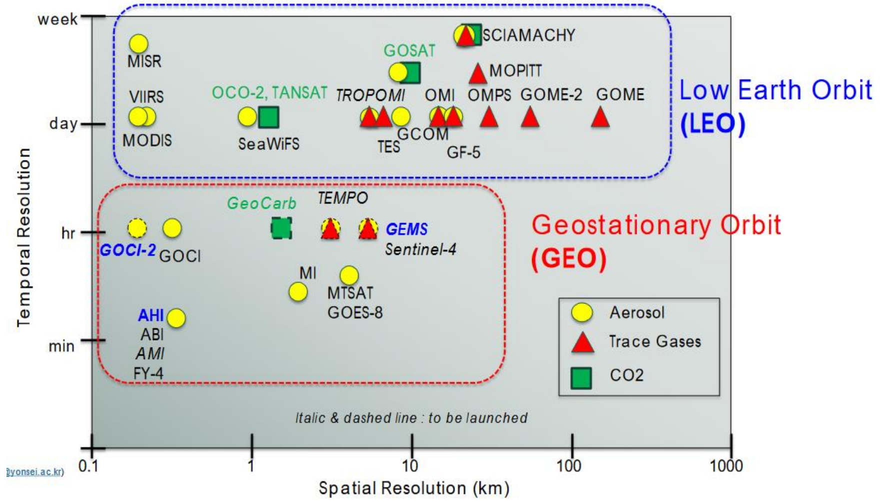
Contact me: jek0420@unist.ac.kr

Introduction

- AOD retrieval process

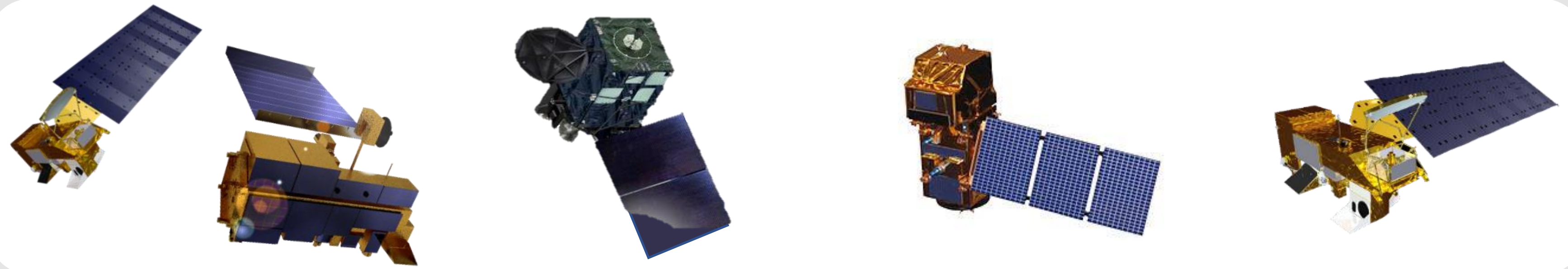


● AOD Satellite Monitoring over East Asia



(Provided by Jhoon Kim (jkim2@yonsei.ac.kr))

Satellite & Sensor



Terra & Aqua
MODIS AOD : 10km, 3km
MAIAC AOD : 1km

Himawari-8 AHI
Level 3 AOD : 5km

Sentinel AOT
: 20m

COMS- GOCI AOD
: 6km

Contact me: jek0420@unist.ac.kr

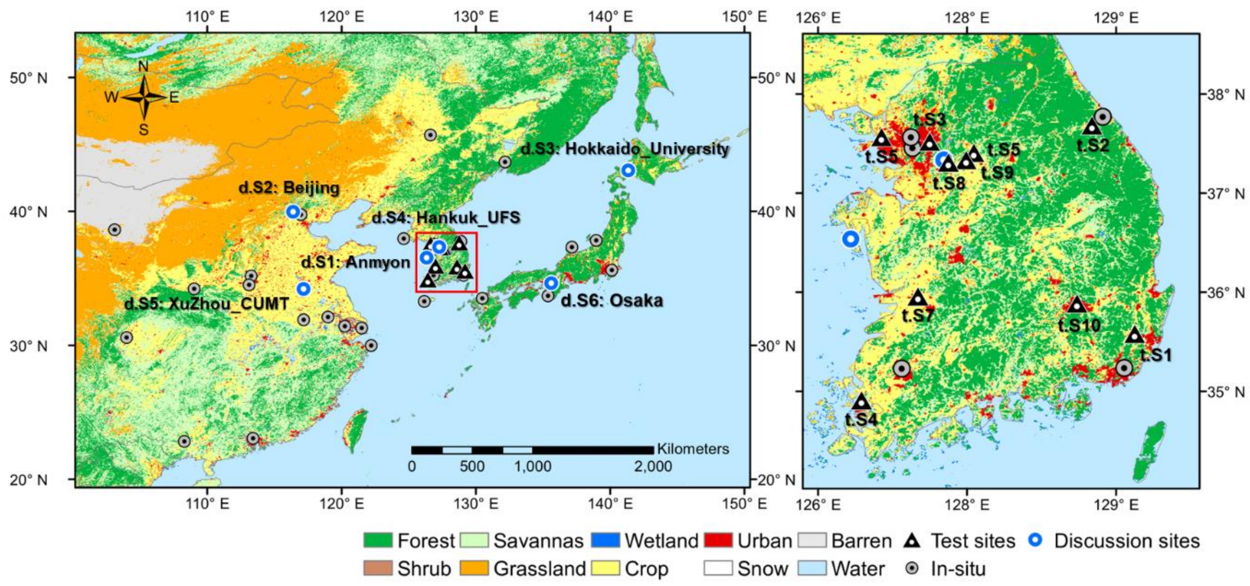
Input variables & Method

In-situ / level	Data Quality
1.0	unscreened
1.5	cloud-screened and quality controlled
2.0	quality-assured

- **Study purpose**

Retrieval of 250 m, 500 m, 1 km AOD using only satellite data and machine learning over East Asia

- **Study Area & Period: East Asia (2016-2019)**



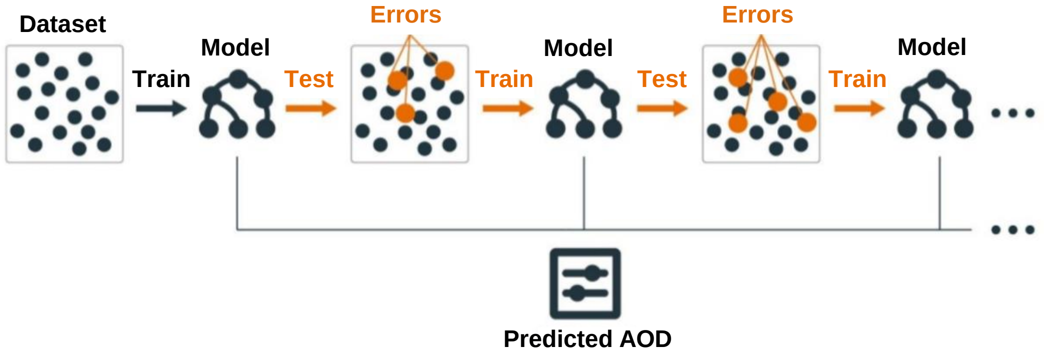
Satellite / Source	Variables	Spatial resolution	Wavelength (nm)
MXD02QKM	Top of Atmosphere reflectance Band 1, 2	250m	B1: 620-270 B2: 841-876 B3: 459-479 B4: 545-565 B7: 2105-2155 B8: 405-420 B17: 890-920 B18: 931-941 B19: 915-965 B26: 1360-1390
MXD02HKM	Top of Atmosphere reflectance Band 1, 2, 3, 4, 7	500m	
MXD021KM	Top of Atmosphere reflectance Band 1, 2, 3, 4, 7, 8, 17, 18, 19, 26	1km	
MXD09	Surface reflectance Band 1, 2	250m	
	Surface reflectance Band 1, 2, 3, 4, 7	500m	
	Surface reflectance Band 1, 2, 3, 4, 7	1km	
	b_r, r_s b_s, b_g	250m 500m 1km	b_r = B3/B1 r_s = B1/B7 b_s = B3/B7, b_g = B3/B4
MXD03	Sensor Azimuth Angle (SEA) Sensor Zenith Angle (SEZ) Solar Azimuth Angle (SOA) Solar Zenith Angle (SOZ)	1km	-

Contact me: jek0420@unist.ac.kr

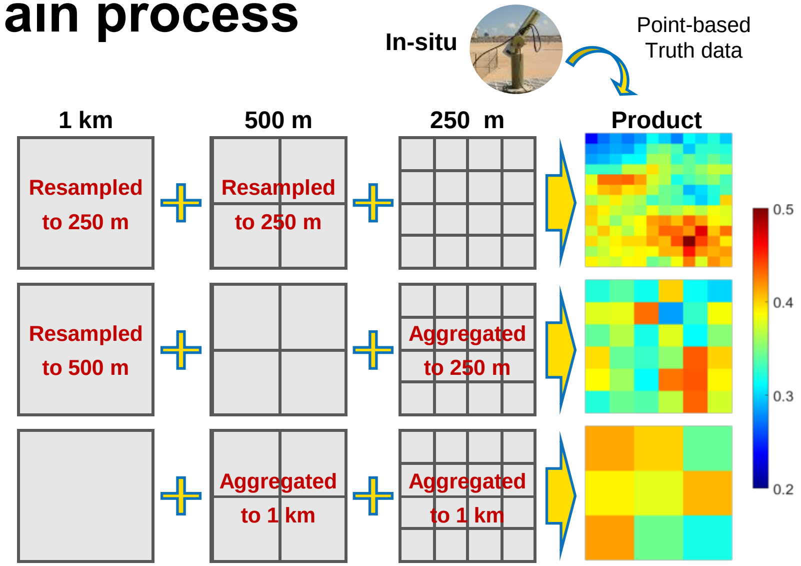
Method

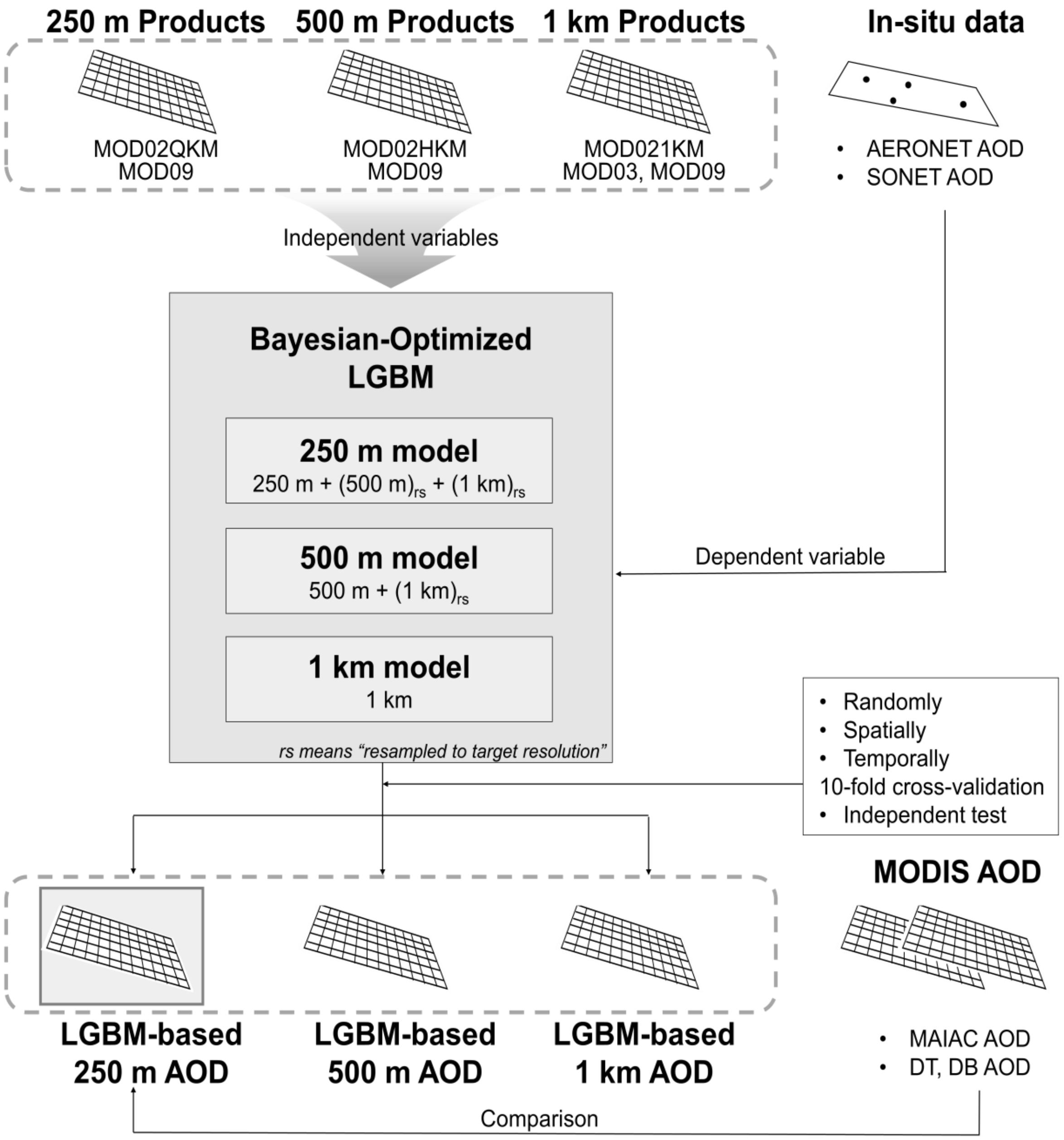
Applied Machine learning

Light Gradient Boosting Model (LGBM)



Main process

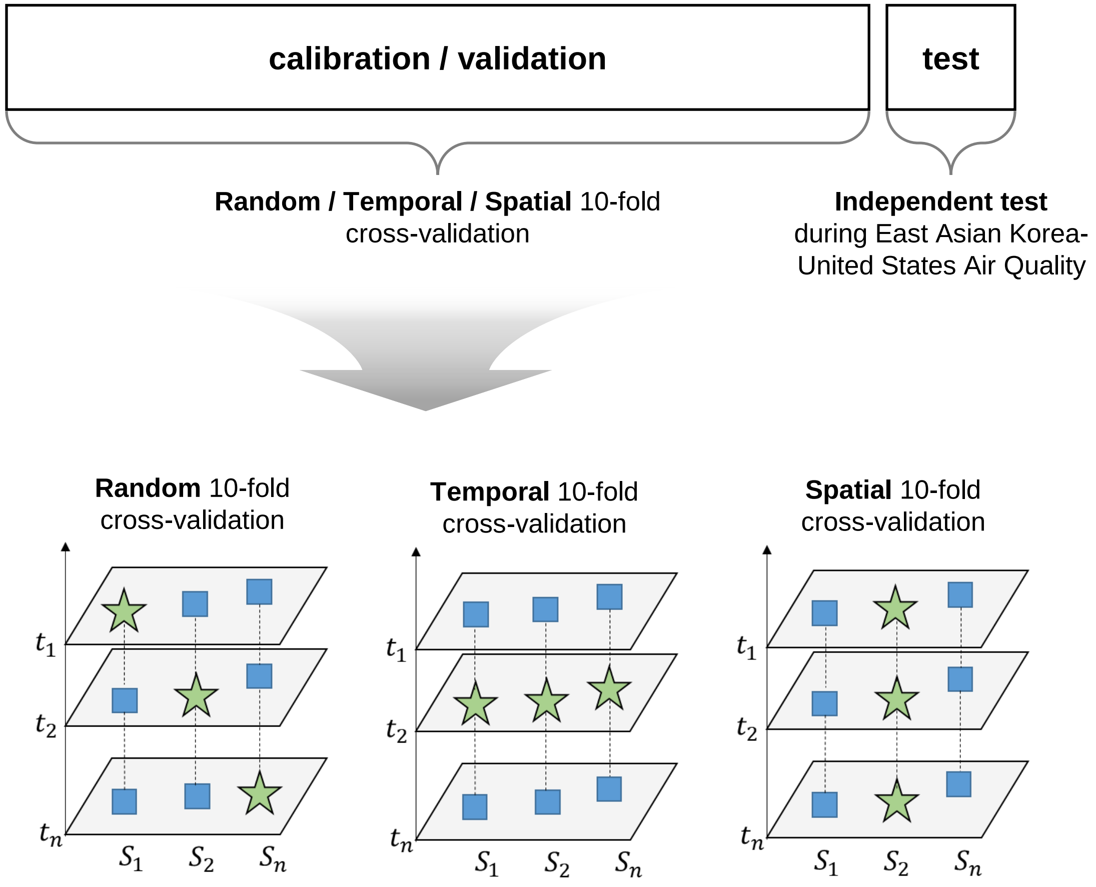




Contact me: jek0420@unist.ac.kr

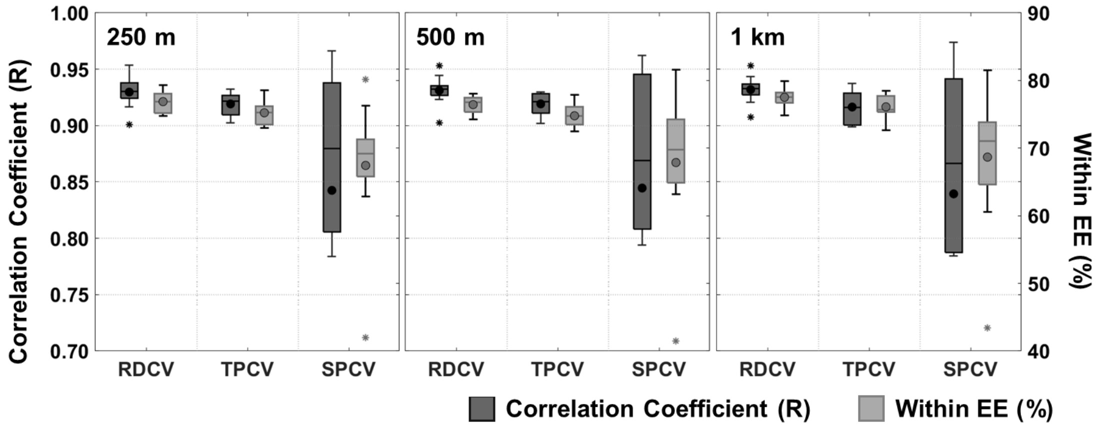
Result & Discussion

● Modeling performance evaluation



- **Medians of the 10-fold results for each cross-validation** for the proposed 250 m, 500 m, and 1 km models.

CV	RDCV			TPCV			SPCV		
Resolution	250 m	500 m	1 km	250 m	500 m	1 km	250 m	500 m	1 km
R	0.93	0.93	0.93	0.92	0.92	0.92	0.88	0.87	0.87
RMSE	0.099	0.098	0.097	0.107	0.106	0.101	0.110	0.114	0.116
MAE	0.064	0.064	0.062	0.066	0.066	0.066	0.077	0.078	0.073
Within EE (%)	76.9	76.8	77.5	75.3	74.8	75.7	69.2	69.8	71.0
IOA	0.96	0.96	0.96	0.95	0.95	0.95	0.93	0.93	0.92



- Boxplots for the Pearson correlation (R) and within EE (%) values of **each cross-validation for different AOD spatial resolutions.**

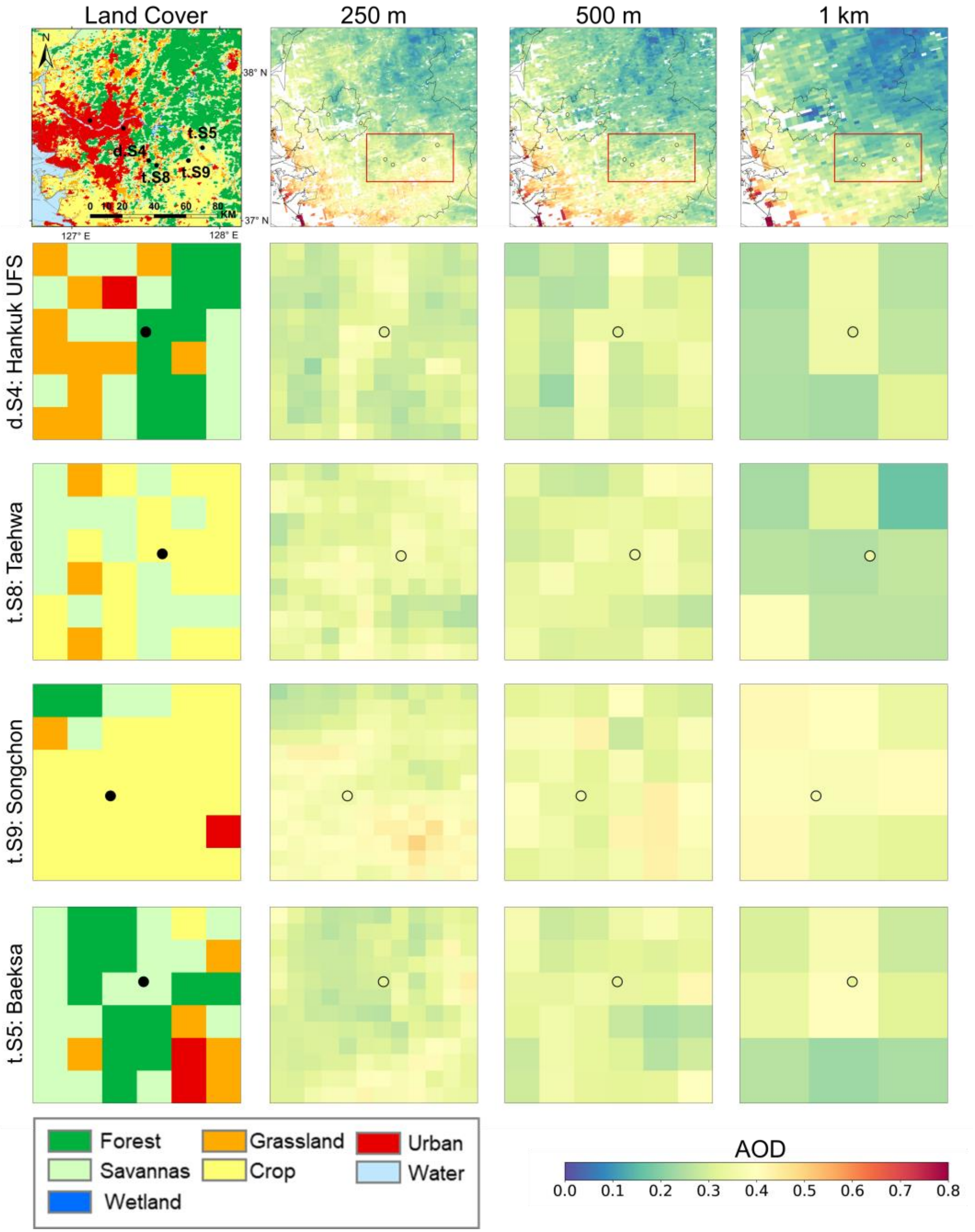
- **Results of the test set** for the proposed 250 m, 500 m, 1 km models and MAIAC AOD

N = 558	250 m	500 m	1 km	MAIAC
R	0.86	0.87	0.84	0.88
RMSE	0.084	0.083	0.095	0.097
MAE	0.059	0.058	0.065	0.068
Within EE (%)	77.8	76.3	76.5	70.8
IOA	0.93	0.93	0.91	0.93

Contact me: jek0420@unist.ac.kr

Result & Discussion

Spatial scale of different resolution model



- MODIS land cover map for 2016 and the images for the 250 m, 500 m and 1 km models are shown from the Aqua overpass on June 5, 2016.
- Aerosol characteristics were mainly affected by land cover parameters. Low AOD values were estimated for forests, while high aerosol loading was estimated near urban pixels (Qin et al., 2021)
- 1 km model tends to underestimate the AOD while the 250 m and 500 m models well captured the high/low aerosol fluctuations.

Contact me: jek0420@unist.ac.kr

Result & Discussion

● Variable analysis through SHAP

SHAP

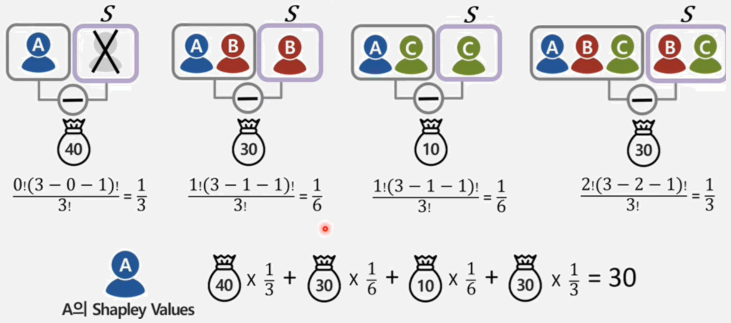
SHapley Additive exPlanations

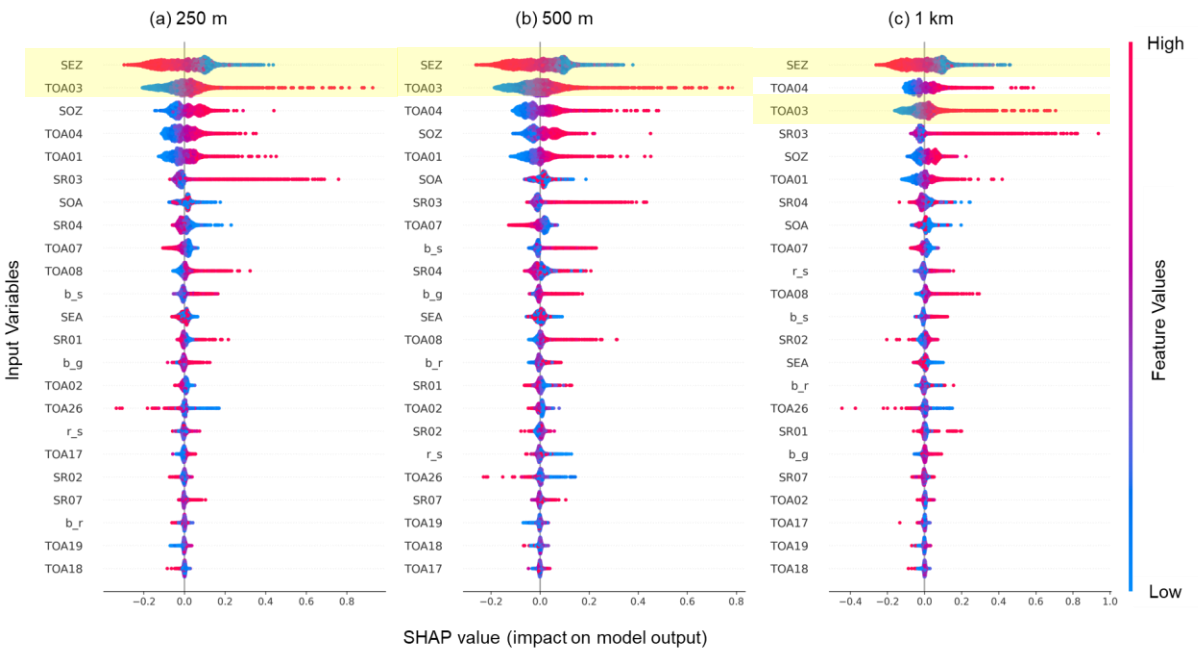
a method based on **game theory** and shows the **contribution or the importance of each feature** on the prediction of the model

$$\text{SHAP values} = \sum_{S \subseteq F \setminus \{A\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{A\}}(x_{S \cup \{A\}}) - f_S(x_S) \right]$$

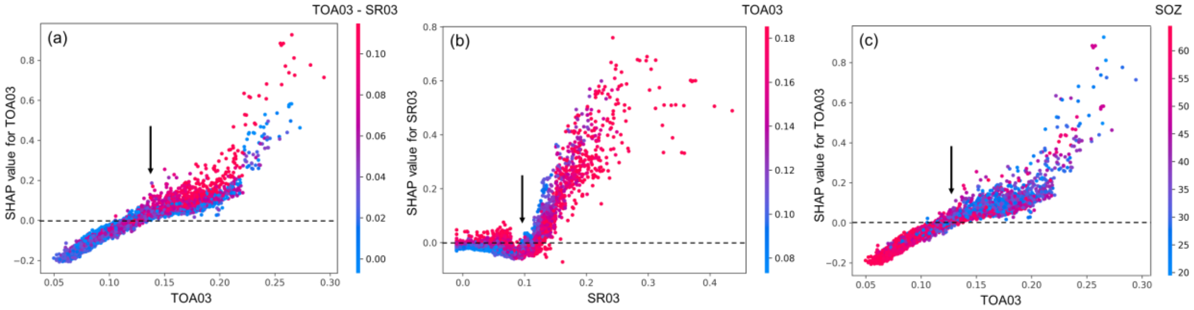
- $\boldsymbol{f}$: the prediction (estimation) model
- $\boldsymbol{F}$: the set of all input features
- $\boldsymbol{S}$: one of all possible subsets excluding i,
- $\frac{|S|!(|F|-|S|-1)!}{|F|!}$: normalized weight sum of perturbations of the sets exist in the case of no i.

Examples:





- **SHAP summary plots** of all input variables ranked by **global feature importance** in the LGBM-based 250 m, 500 m, and 1 km models



- **SHAP dependency plots** of LGBM-based 250 m model indicating the **contribution** of **TOA03** and **SR03 (459–479 nm)** to the AOD predictions and comparisons based on the **difference between TOA03 and SR03** and **SOZ**.

Contact me: jek0420@unist.ac.kr

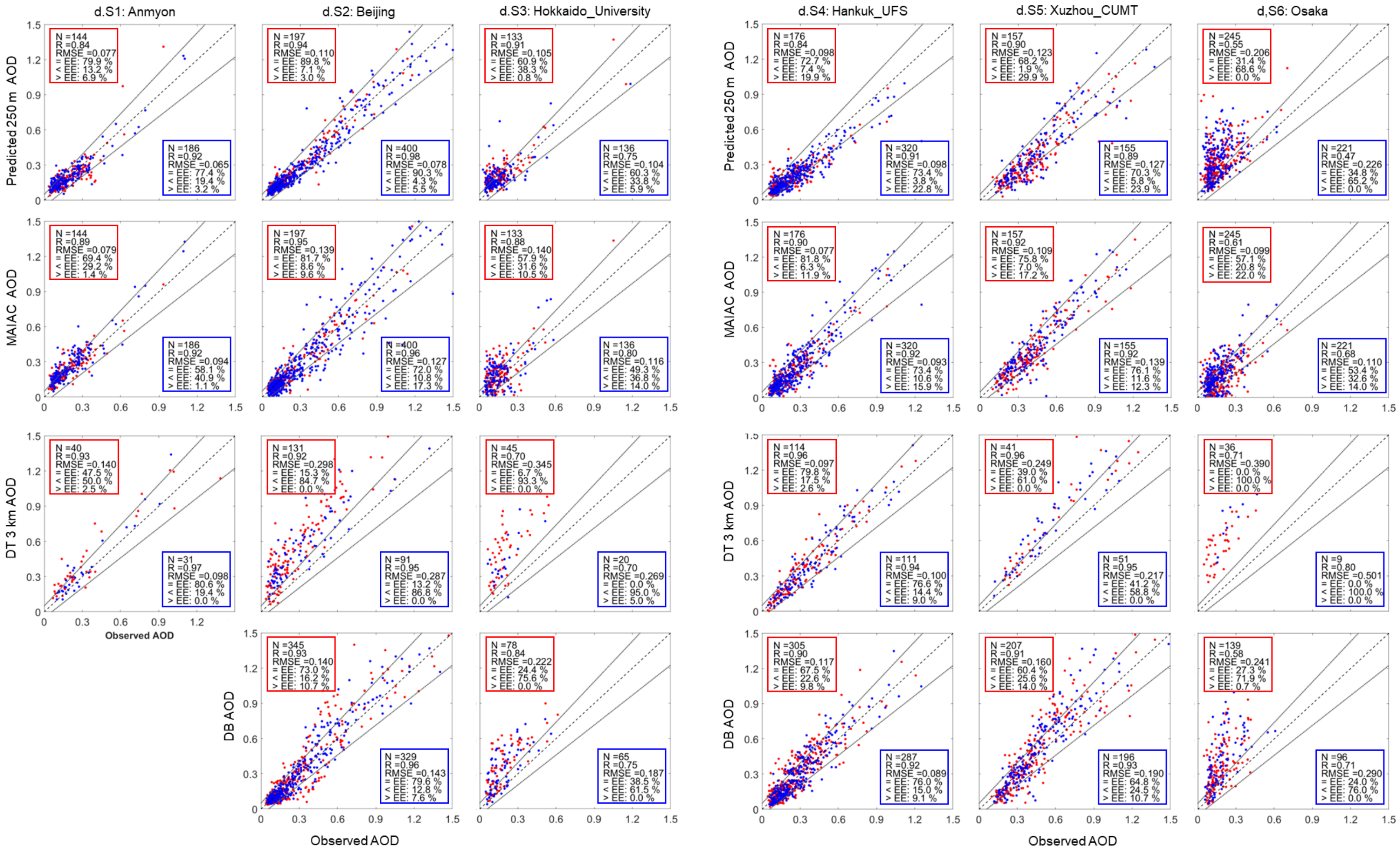
Result & Discussion

● Spatiotemporal analysis of MODIS AOD

- Accuracy metrics RDCV, TPCV, and SPCV for the Osaka station

	RDCV		TPCV		SPCV	
Satellite	Terra	Aqua	Terra	Aqua	Terra	Aqua
R	0.65	0.69	0.63	0.64	0.55	0.47
RMSE	0.105	0.099	0.107	0.101	0.206	0.226
Within EE (%)	63.3	63.3	60.0	66.1	31.4	34.8
Above EE (%)	32.2	34.8	35.5	31.7	68.6	65.2
Below EE (%)	4.5	1.8	4.5	2.3	0.0	0.0

- **Better performance** compared to MODIS AODs
- **Lower performance** than MAIAC, but higher than the others



The SPCV results of the 250 m model were used to evaluate the ability to retrieve AOD from unseen data.

Terra

Aqua

Contact me: jek0420@unist.ac.kr

Result & Discussion

● Spatiotemporal analysis of MODIS AOD

- **Local pollution (May 17–22)**
: low AOD levels were found under stagnant conditions after Asian dust from May 4 to 7

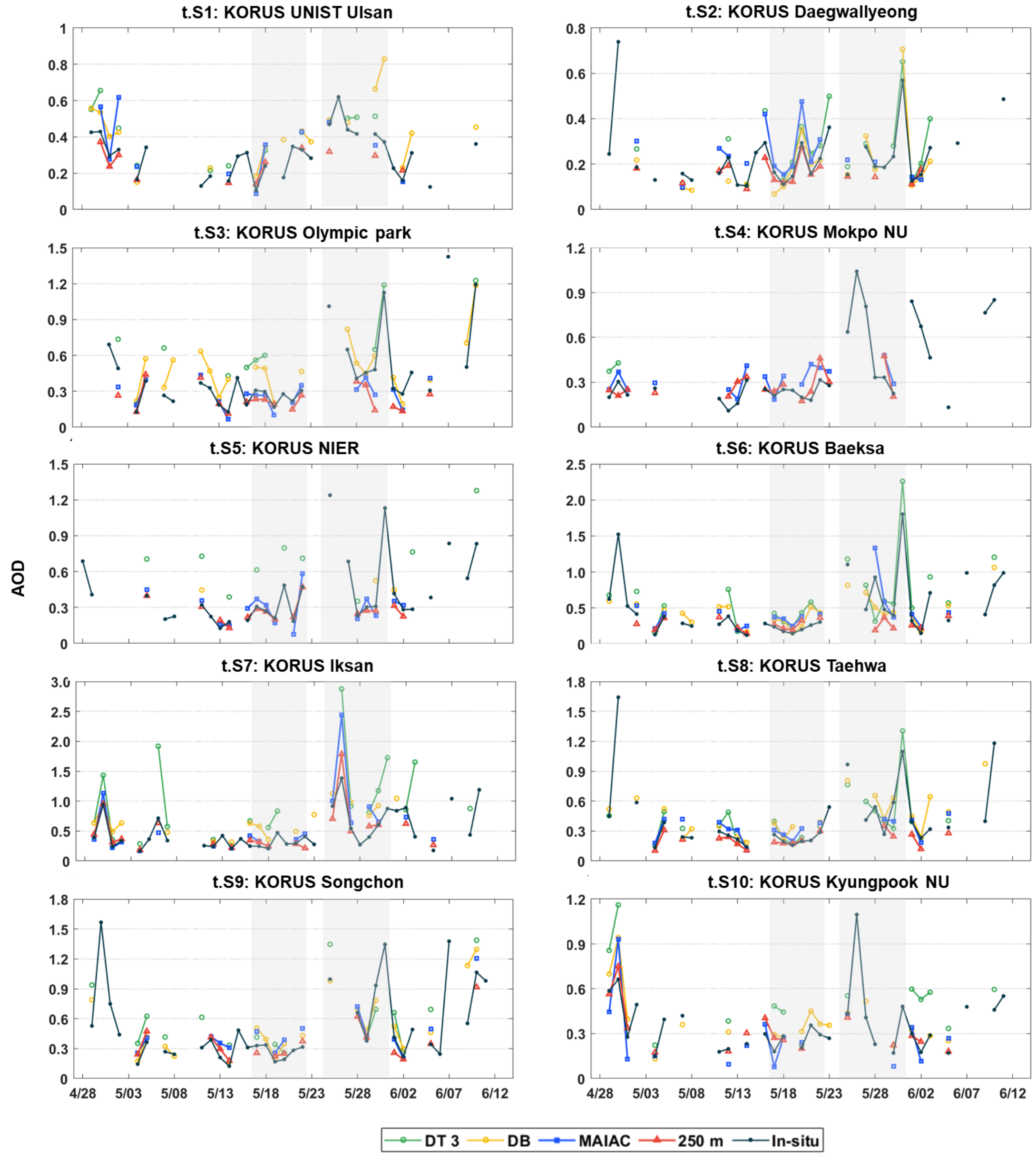
- **Long-range transportation (May 25–31)**
: peak values from a highly polluted air mass from China in the western part of Korea were observed at all sites.

(Choi et al., 2021)

- Accuracy metrics from the predicted 250 m, MAIAC, DT, and DB AODs during KORUS-AQ (May 1 to June 10, 2016) and local pollution (May 17 to 22, 2016).

5/1-6/10	250 m	MAIAC	DT	DB	In-situ
mean	**0.259**	**0.350**	**0.414**	**0.373**	**0.287**
max	0.917	1.336	1.387	1.295	1.065
RMSE	0.120	0.124	0.201	0.144	-
MBE	-0.028	0.064	0.128	0.086	

5/17-5/22	250 m	MAIAC	DT	DB	In-situ
mean	**0.230**	**0.292**	**0.345**	**0.313**	**0.221**
max	0.365	0.475	0.600	0.508	0.330
RMSE	0.051	0.102	0.156	0.112	-
MBE	0.009	0.071	0.124	0.092	



- **Daily time series** of 550 AOD from the predicted 250 m, MAIAC, DT, and DB AODs during the KORUS-AQ (May 1 to June 10, 2016).

Contact me: jek0420@unist.ac.kr

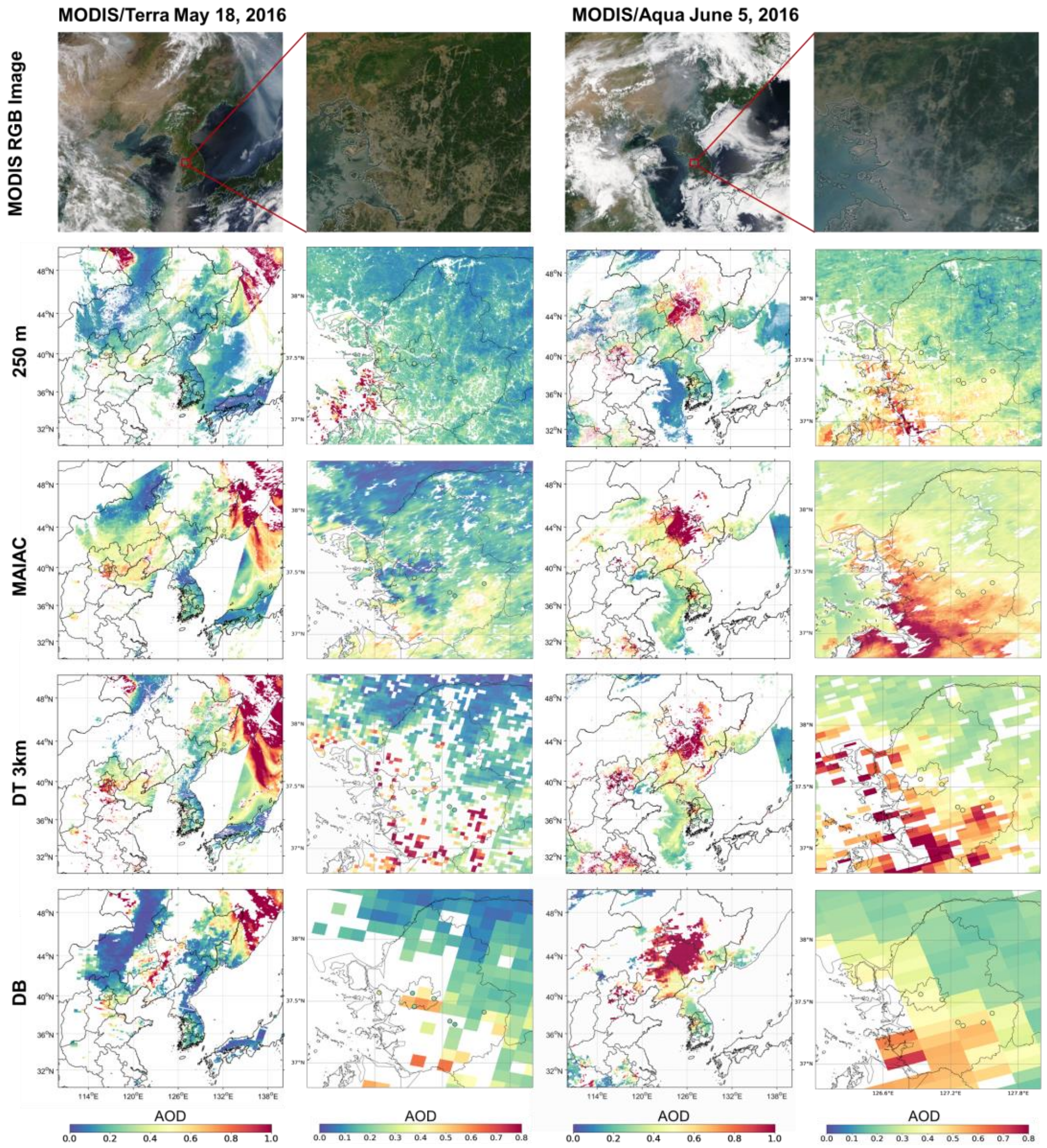
Result & Discussion

Spatiotemporal analysis of MODIS AOD

- **May 18, 2016**
: high AOD occurred caused by Russian forest fire

- **June 5, 2016**
: local emissions in Korea

(Choi et al., 2019; Kang et al., 2022)



- **AOD maps of the MODIS true color image, proposed 250 m AOD, MAIAC (1 km), Deep Blue (DB), and Dark Target (DT) algorithms** over East Asia (left) and the expanded Seoul metropolitan area (right) on May 18, 2016 (Terra) and June 5, 2016 (Aqua).

- Accuracy metrics from the predicted 250 m, MAIAC, DT, and DB AODs on May 18, 2016 (Terra) and June 5, 2016 (Aqua) over Seoul metropolitan area.

May 18	250 m	MAIAC	DT	DB
RMSE	0.058	0.097	0.384	0.246
MBE	-0.049	0.002	0.384	0.207

June 5	250 m	MAIAC	DT	DB
RMSE	0.039	0.129	0.208	0.110
MBE	0.007	0.126	0.166	0.108

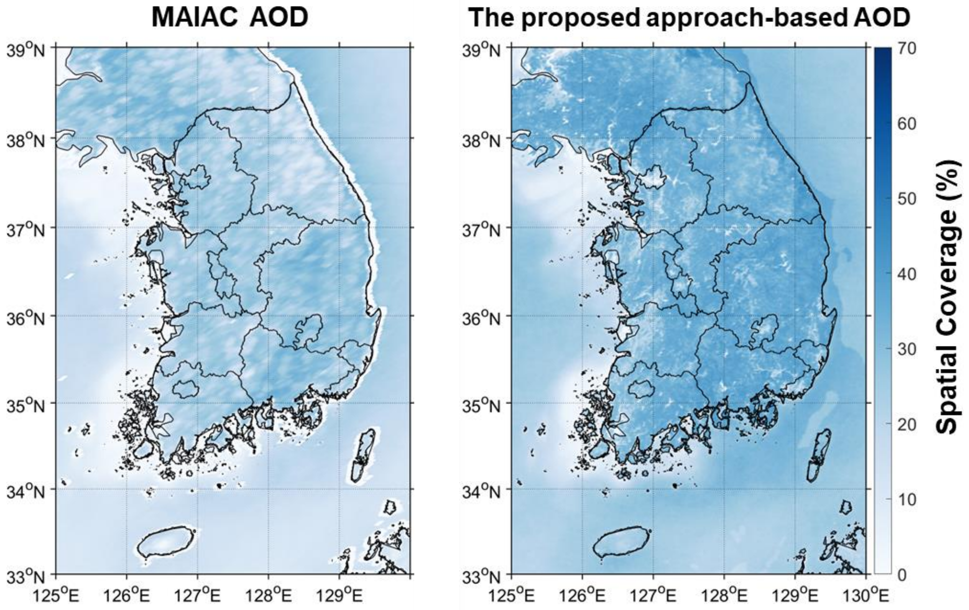
Contact me: jek0420@unist.ac.kr

Result & Discussion

- **Spatial coverage comparison of the proposed approach-based AOD with MAIAC AOD**

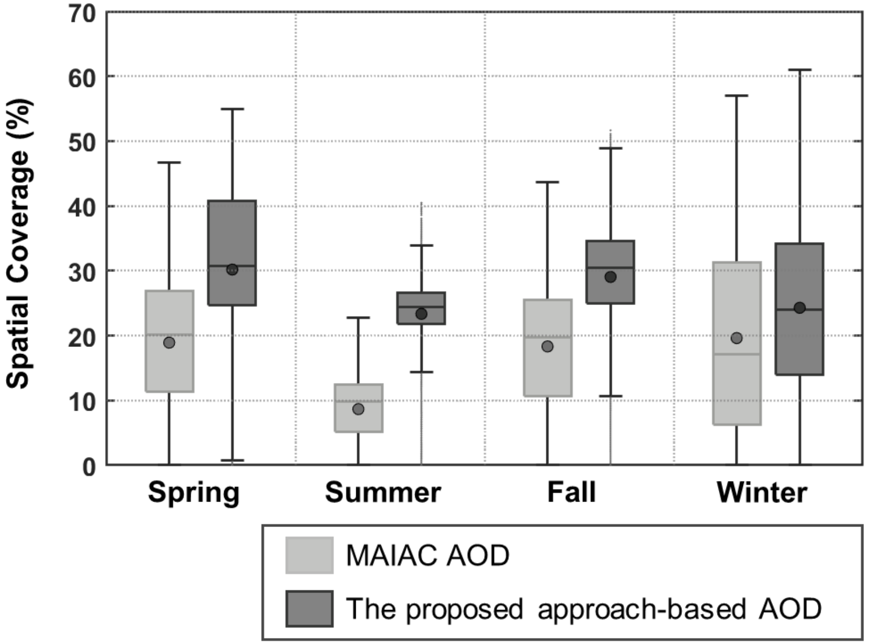
Proposed approach-based AOD applied two cloud masking methods

1. 95% clear sky for the MODIS cloud mask
2. red/SWIR threshold (<0.78) suggested by Jethva et al (2009)



- **Spatial coverage maps of the proposed approach-based AOD and MAIAC AOD** by pixel around South Korea. Each pixel indicates the number of valid AODs during the entire study period (2016-2019) using Terra data.

→ 1.5 higher frequency than MAIAC AOD



- **Seasonal spatial coverage (%) boxplot of the MAIAC AOD and the proposed approach-based AOD** (i.e., red/SWIR under 95% clear sky conditions) for each pixel over South Korea. The analysis was based on Terra overpass data from 2016 to 2019, which were divided into spring (March-May), summer (June-August), autumn (September-November), and winter (December-February).

Contact me: jek0420@unist.ac.kr

Conclusion

- 250 m findings demonstrated high and consistent accuracy compared to those of the 500 m and 1 km models despite the dynamic atmospheric condition.
- The SHAP analysis demonstrated high contributions of SEZ and TOA03 to the predicted AOD, and it is comparable to the physical process of AOD retrieval.
- The frequency rate of the proposed approach-based AOD by pixels was approximately 1.5 times higher than that of MAIAC AOD showing significant accuracy in regions where MAIAC pixels are not provided.
- Despite the limitation, of no sufficiently trained aerosol characteristics due to limited ground observations, the findings indicate that satellite data can be applied for machine learning-based high-resolution AOD.

Reference

Choi, M., Lim, H., Kim, J., Lee, S., Eck, T.F., Holben, B.N., Garay, M.J., Hyer, E.J., Saide, P.E. and Liu, H., 2019. Validation, comparison, and integration of GOCI, AHI, MODIS, MISR, and VIIRS aerosol optical depth over East Asia during the 2016 KORUS-AQ campaign. *Atmospheric Measurement Techniques*, *12*(8), pp.4619-4641.

Jethva, H., Satheesh, S.K., Srinivasan, J. and Moorthy, K.K., 2009. How good is the assumption about visible surface reflectance in MODIS aerosol retrieval over land? A comparison with aircraft measurements over an urban site in India. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), pp.1990-1998.

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*, pp.253-268.