Every Hour, Across East Asia: 24-hour, Full-Coverage Ground-level Ozone Estimation Techniques using Machine Learning

S15

Yejin Kim[1], Seohui Park[2,3], Hyunyoung Choi[1], Jungho Im[1]

[1] Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, 44919, Republic of Korea
[2] Goddard Earth Sciences Technology and Research (GESTAR) Ⅱ, Morgan state university, Baltimore, MD 21251, USA
[3] NASA Goddard Space Flight Center (GSFC), Greenbelt, MD 20771, USA

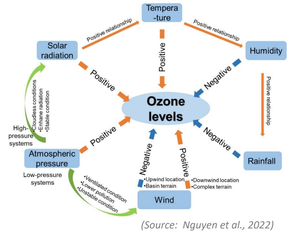
Contact: happykimmy@unist.ac.kr

105TH ANNUAL MEETING 12–16 JANUARY NEW ORLEANS & ONLINE AMERICAN METEOROLOGICAL SOCIETY

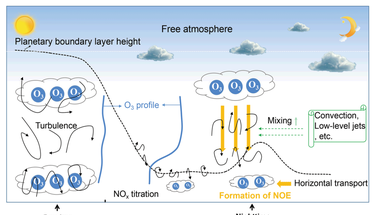
Introduction

Background

- Unlike ozone in the ozone layer, **ground-level ozone can be harmful to our health even when exposed to high concentrations for a short time**.
- Ground-level ozone is a secondary pollutant in the atmosphere, **produced by photochemical reactions of primary pollutants such as NOx and VOCs**, and is affected by varying weather conditions *(Nguyen et al., 2022)*.
- Ozone usually peaks during the day and diminishes after sunset, but **events of intensified ozone levels at night are frequently observed**. Elevated nighttime ozone levels can impact the next day's atmospheric chemical composition *(He et al., 2022)*.



(Source: Nguyen et al., 2022)



(Source: He et al., 2022)

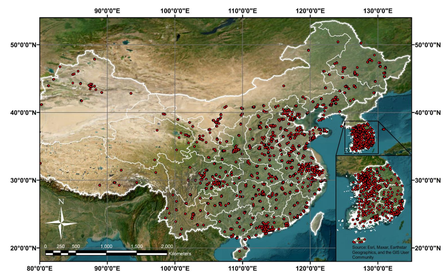
Objectives

- Development of a **machine learning-based model for 24-hour ground-level ozone estimation** utilizing multiple data sources.
- Achieving **seamless estimation of ground-level ozone spatial distribution** to overcome the limitations of traditional monitoring station-based observations.

Study domain & Data

Study domain

- **Study area**: Korean Peninsula, China (18–54°N, 80–135°E)
- **Study period:** 2021.01.01 – 2022.12.12
- 2536 observation stations

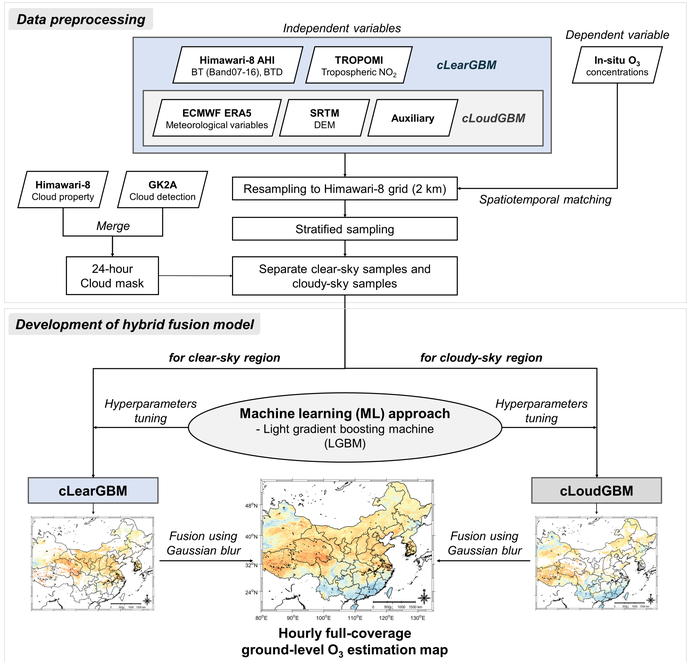


Data

Type	Source	Variables(Abbreviation)	Spatial Resolution	Temporal Resolution
In-situ	AirKorea, CNEMC	Ground-based O_3 measurements	Point	Hourly
Satellite	Himawari-8/AHI	BTs of Bands 07, 11 – 16 BTD (difference between BT13 and BT15) Solar zenith angle	2 km	10 min
	S5P TROPOMI	Tropospheric NO_2	5.5 x 3.5 km	~ Daily
	S5P TROPOMI	Tropospheric NO_2	5.5 x 3.5 km	~ Daily
	MODIS	NDVI	1 km	16days
	GHS	Built-up surface (Built)	1 km	5 years
	SRTM	DEM	3 arc-second (~ 0.09 km)	-
Numerical Model	ERA5	Dew, Temp, SSR, RH, WS	0.1°	Hourly
		PBLH	0.25°	
Auxiliary	GPW	Population density (PopDens)	1 km	5 years
	Time	DOY converted into a sine function (DOY)	-	daily

Methodology

Framework for Hybrid fusion model



Validation method

➢ **K-fold Cross-validation**
- Random 10-fold cross-validation (**RDCV**): divided randomly
- Spatial 10-fold cross-validation (**SPCV**): divided by station
- Temporal 10-fold cross-validation (**TPCV**): divided by doy

➢ **Independent spatial validation**
: One station selected for each of China and South Korea

Conclusion

This study developed a **hybrid fusion model to provide high-resolution (2 km) and full-coverage ground-level O_3 estimates**, enabling **detailed hourly diurnal analysis across both daytime and nighttime**. The model **captured distinct diurnal patterns, with urban areas showing elevated O_3 concentrations during the day and rural areas exhibiting higher levels at night**. By addressing the limitations of previous studies focused primarily on daytime, this research highlights its novelty in delivering seamless 24-hour O_3 monitoring, offering valuable insights for advancing air quality management and public health strategies.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (RS-2024-00408905)

Nguyen, D. H., Lin, C., Vu, C. T., Cheruiyot, N. K., Nguyen, M. K., Le, T. H., ... & Bui, X. T. (2022). Tropospheric ozone and NOx: A review of worldwide variation and meteorological influences. Environmental Technology & Innovation, 102809.

He, C., Lu, X., Wang, H., Wang, H., Li, Y., He, G., ... & Fan, S. (2022). The unexpected high frequency of nocturnal surface ozone enhancement events over China: characteristics and mechanisms. Atmospheric Chemistry and Physics, 22(23), 15243-15261.

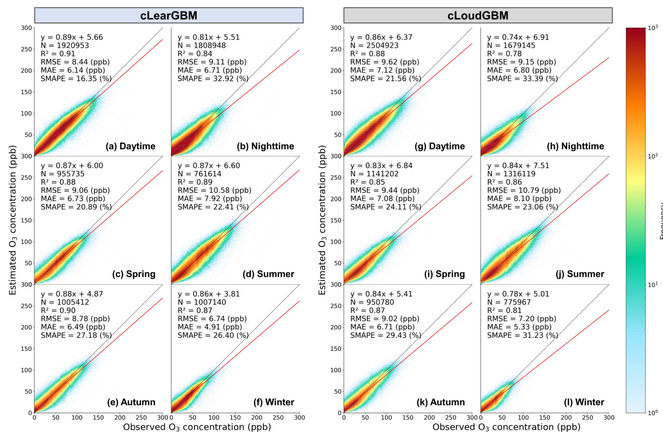
Results & Discussion

Overall validation results

Models		cLearGBM	cLoudGBM
Cross-Validation	N	3,729,901	4,184,068
RDCV	R^2	**0.90**	**0.87**
	RMSE	**8.77**	**9.44**
	MAE	**6.42**	**6.99**
SPCV	R^2	0.82	0.77
	RMSE	11.60	12.53
	MAE	8.50	9.33
TPCV	R^2	0.80	0.72
	RMSE	12.41	13.92
	MAE	9.10	10.28

(* RMSE, MAE unit: ppb)

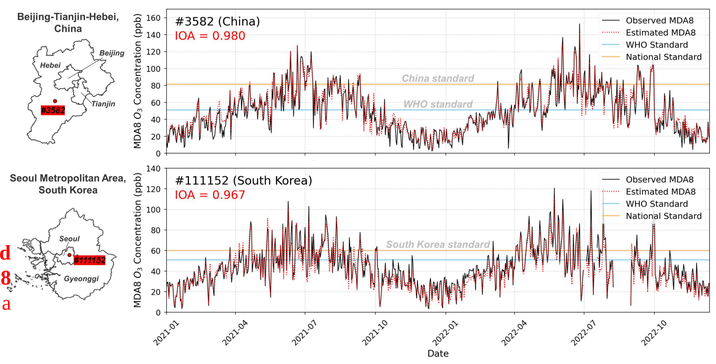
Temporal-scale accuracy



Scalability of Hybrid fusion model

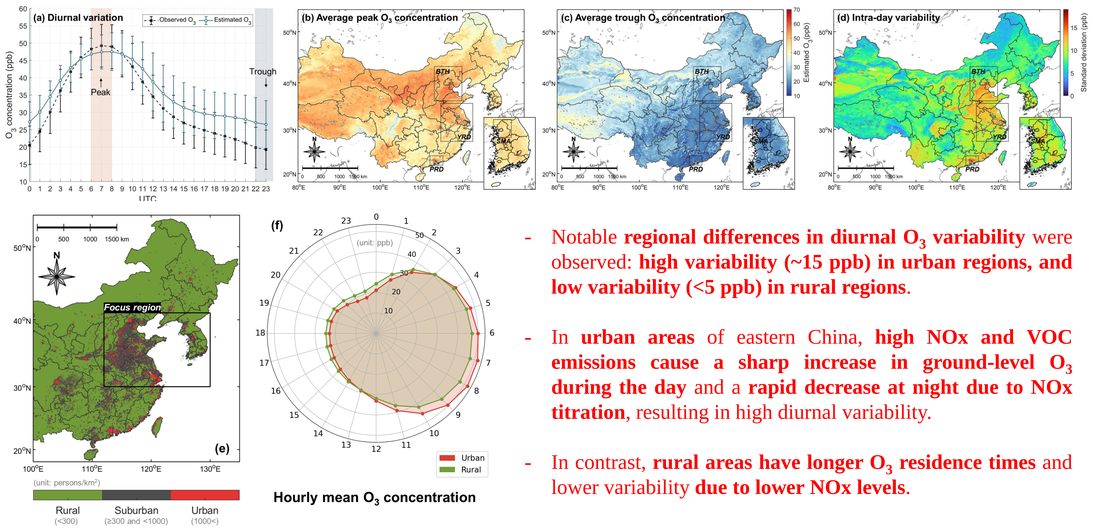
Test station	Metric	Hourly scale	MDA8 scale
China (#3582)	R^2	0.86	**0.93**
	RMSE	9.87	**7.18**
	MAE	7.27	**5.40**
South Korea (#111152)	R^2	0.85	**0.88**
	RMSE	8.04	**6.86**
	MAE	5.83	**5.13**

(* RMSE, MAE unit: ppb)



- The **robust performance of the hybrid fusion model at both hourly and MDA8 time scales** highlights its potential as a reliable tool for air quality management.

Diurnal variation



- Notable **regional differences in diurnal O_3 variability** were observed: **high variability (~15 ppb) in urban regions, and low variability (<5 ppb) in rural regions**.
- In **urban areas** of eastern China, **high NOx and VOC emissions cause a sharp increase in ground-level O_3 during the day** and a **rapid decrease at night due to NOx titration**, resulting in high diurnal variability.
- In contrast, **rural areas have longer O_3 residence times** and lower variability **due to lower NOx levels**.