TG2024

Estimates of diurnal hydroxyl radical (OH) columns in East Asia using GEMS observations

Sang-Ik Oh[1)], Rokjin J. Park[1)], Hyeji Cha[2)], Jhoon Kim[2)]

1) Seoul National University, 2) Yonsei University

✉: ohsangik98@snu.ac.kr

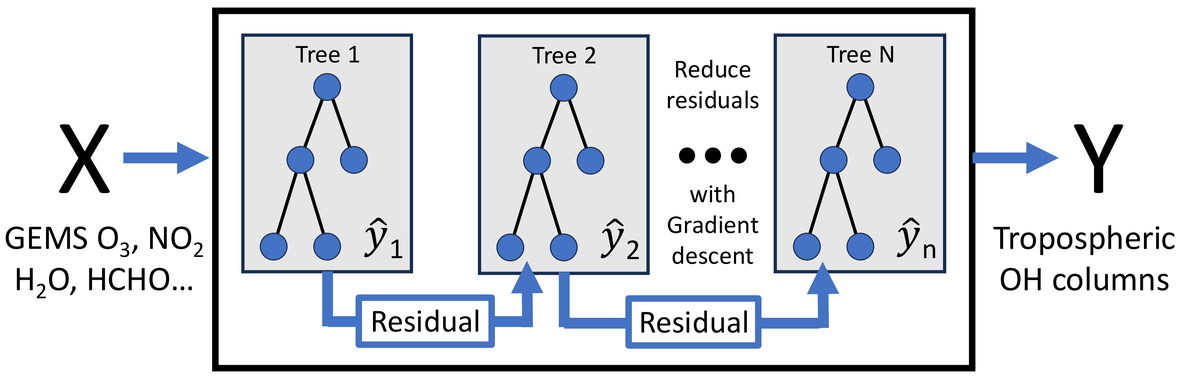
Introduction

- **The hydroxyl radical (OH)** is a pivotal oxidant in the troposphere, **determining the lifetime of pollutants** such as CO, CH4, and NOx.
- However, **direct monitoring of OH through satellites is unavailable** due to its low concentration and very short lifetime.
- D. C. Anderson et al. (2023) introduced a **machine learning** methodology to infer daily-averaged tropospheric OH columns in the tropics, utilizing products from low Earth orbit satellites.
- **Here, we estimate tropospheric OH column amounts in East Asia using GEMS observations and a machine learning method**.

Method

Machine learning methodology (supervised learning)

- Supervised learning framework
  : Trained with given pairs of features (X) and labels (Y), the model predicts labels with unseen features.
- **XGBoost** (eXtreme Gradient Boosting) algorithm
  : High performance in dealing with non-linear problems.



Data : Satellite & Reanalysis products

GEMS							TROPOMI
NO_2	HCHO	O_3	H_2O	Cloud	O_3	Geo-info	CO
Column amounts				Fraction, Center P	Profile	Tropopause, Surface pressure	Column amount

+ Lat, Lon, VZA, SZA (calculated off-line)

XGBoost model

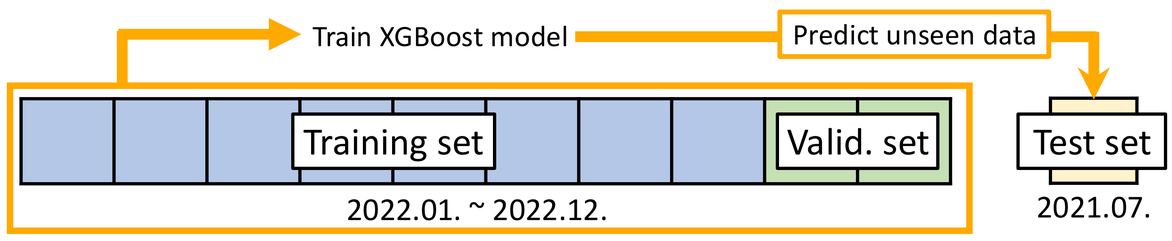
ECMWF Atmospheric Composition reanalysis 4

Tropospheric column OH (TCOH)

- **GEMS** (The Geostationary Environment Monitoring Spectrometer)
  : Hourly observations of trace gases (NO_2, SO_2, O_3, HCHO, CHOCHO)
- EAC4 (ECMWF Atmospheric Composition reanalysis 4)
  - Spatiotemporal resolution : 0.75°×0.75°, 25 layers, 3-hourly
  - Trace gases concentration reanalysis field

Training configuration

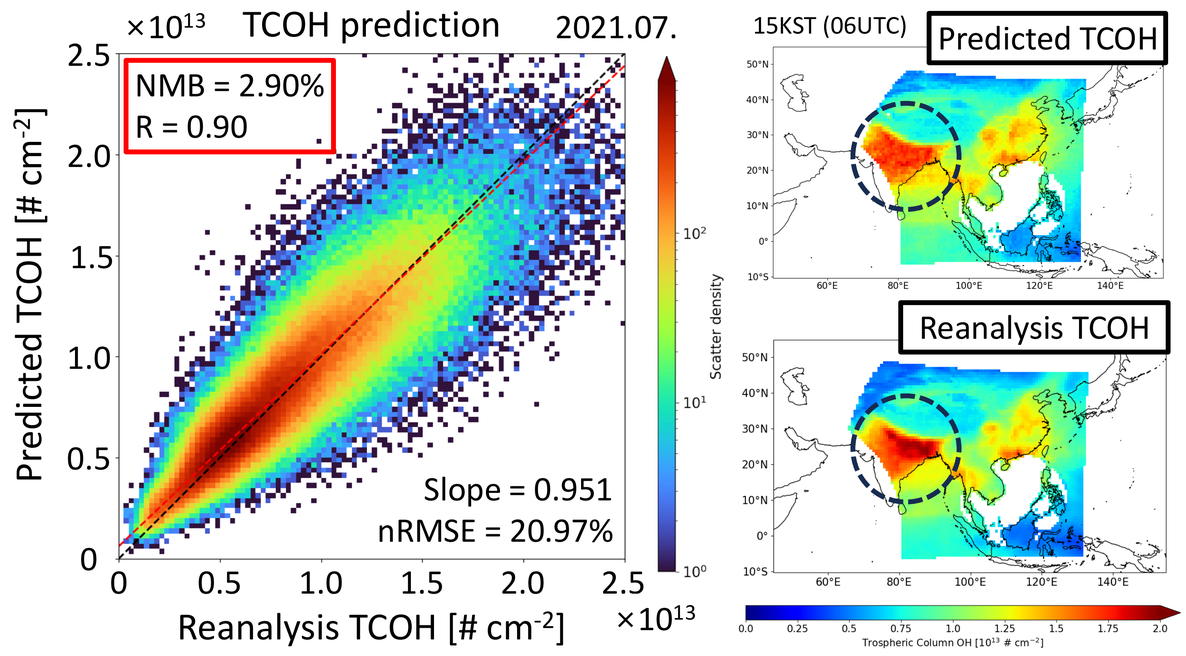
- Training set : 2022.01. ~ 2022.12. (1 year, total 2,644,776 grids)
  Test set : 2021.07. (1 month, total 250,362 grids)
- Learning rate : 0.06, Max tree depth : 19, Boosting round : 500
  Hyperparameters are tuned under grid-search with 5-fold cross validation.



Acknowledgment

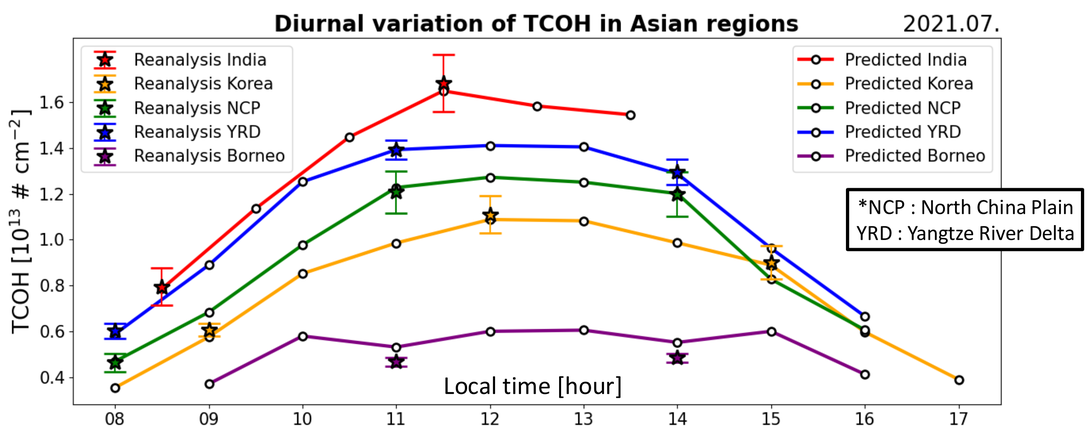
This research was supported by a grant from the National Institute of Environmental Research (NIER) funded by the Korea Ministry of Environment (MOE) of the Republic of Korea (NIER-2024-04-02-028).

Model evaluation



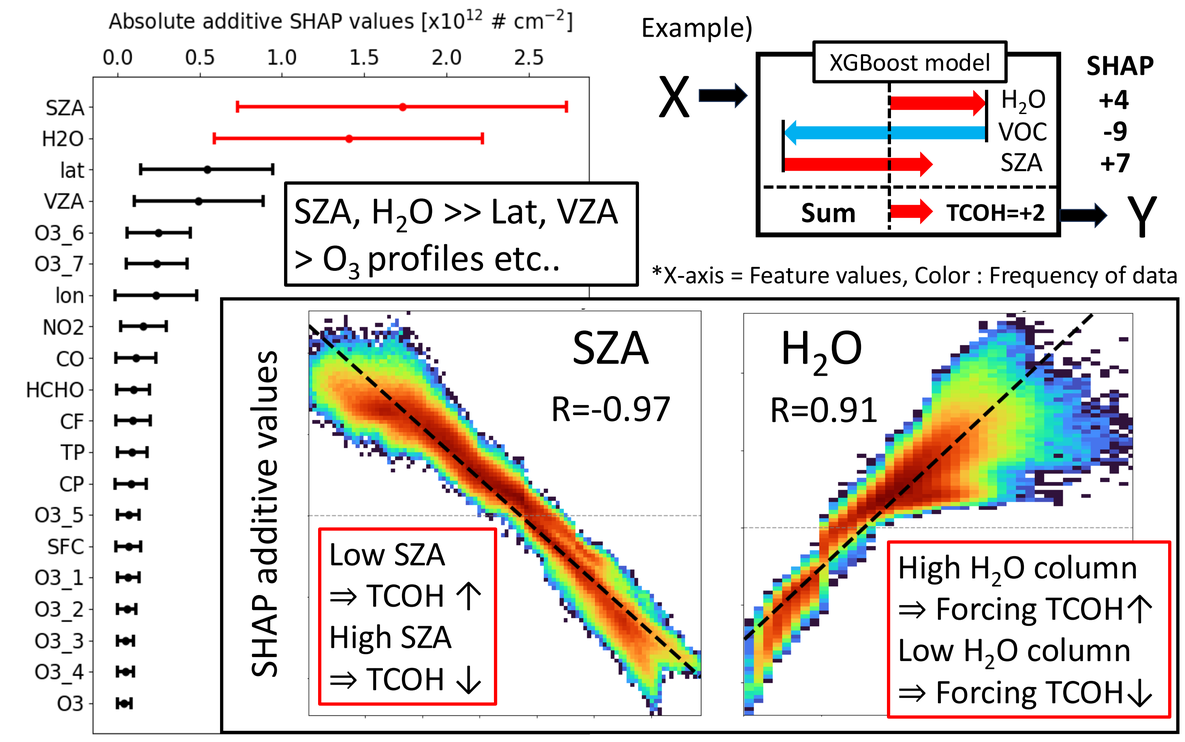
- The predicted TCOH from the XGBoost model is highly correlated with the original TCOH from the reanalysis data.
- However, the model predicted more homogeneous distributions than the reanalysis TCOH due to a shortage of high TCOH data.
  *e.g., underestimation in India.

Diurnal variation of TCOH in Asia



• Predicted TCOH shows reasonable spatiotemporal variations in Asia.
: Higher TCOH is found in polluted regions and at high solar latitudes.

Feature contribution : SHAP analysis



Summary

- This study uses GEMS products with a machine learning model to estimate diurnal variation of tropospheric column OH in Asia.
- The machine learning model successfully predicts diurnally varying TCOH compared to the reanalysis TCOH (R=0.90).
- SZA and H_2O columns are the two most dominant features in predicting TCOH, showing highly correlated SHAP values.