Efficient generation of native resolution NO_2 a priori profiles for TEMPO retrievals using machine learning

Jennifer Grant

Qindan Zhu

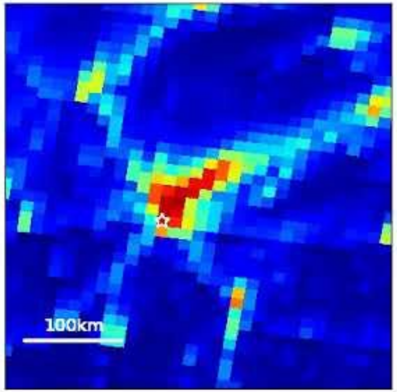
Hannah Kenagy

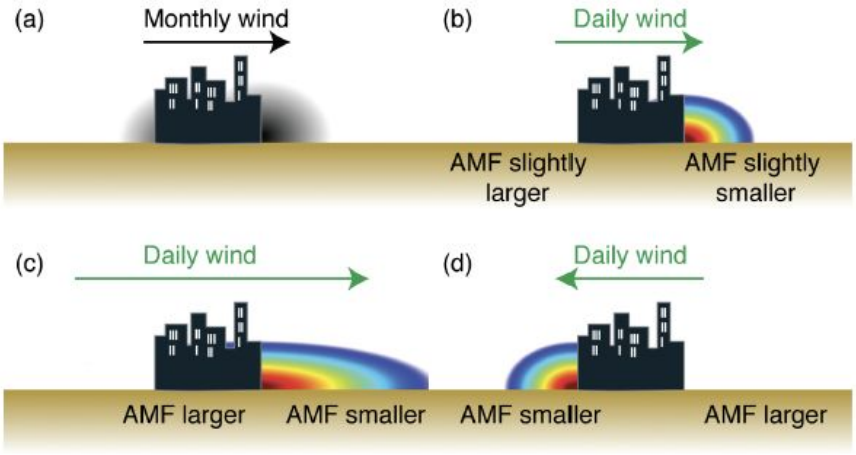
Ron Cohen

University of California, Berkeley

Accurately interpreting TEMPO NO2 observations requires *a priori* profiles at both high spatial and high temporal resolution

The *a priori* NO_2 profiles are the largest contributor to uncertainty in satellite retrievals. Incorrect profiles lead to systematic biases in estimating NO_x lifetime and emissions from satellite retrievals.

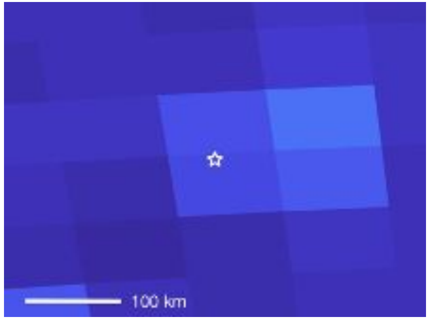




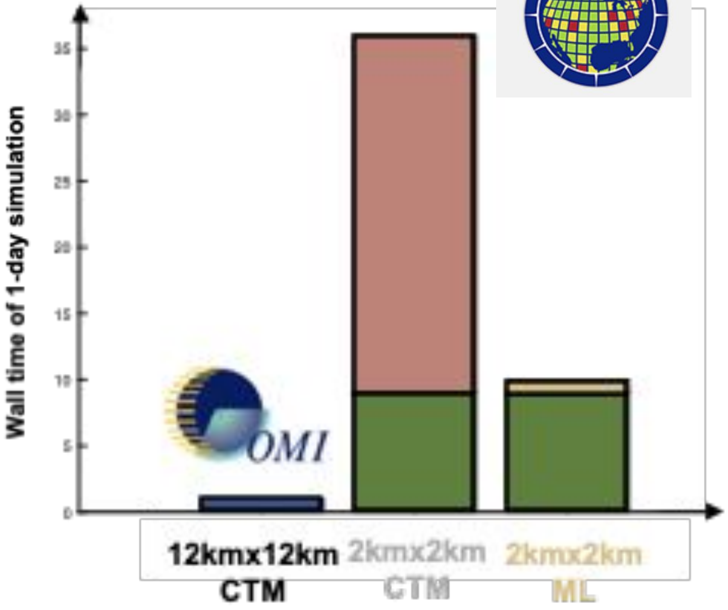
Schematic from Laughner et al., 2016. (a) Monthly average a priori profiles. (b) A case when the daily wind is similar to the monthly average wind. (c) A case where the daily wind is significantly faster than the average but blows in the same direction. (d) A case where the daily wind direction is different from the monthly average.

BEHR (BErkeley High spatial resolution OMI satellite NO_2 Retrievals) product uses daily modeled *a priori* profiles at 12 km x 12 km to capture the daily variation in AMF.

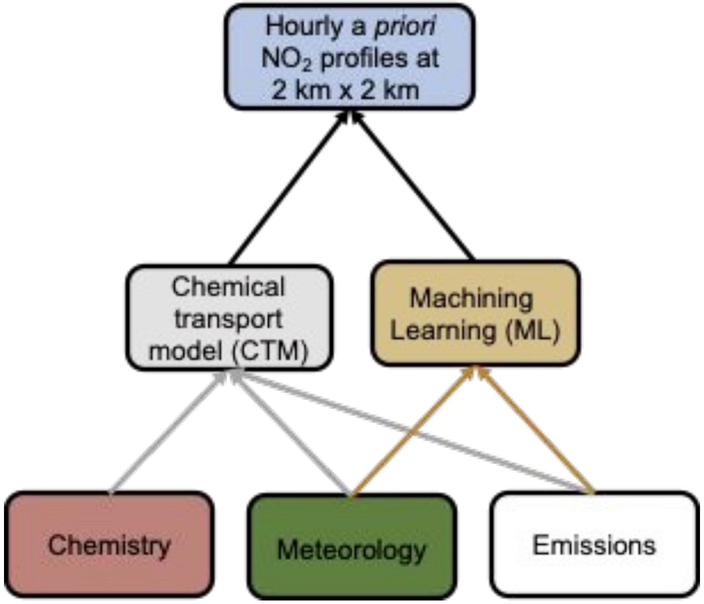
NASA OMI operational product v4.0 utilizes 1 deg x 1.25 deg GMI model based monthly *a priori* NO_2 profile shapes.



Chemical transport model (CTM) is computationally expensive: Machine learning (ML) is an efficient alternative



Back-of-the-envelope estimate of computation time running WRF-Chem on 12 nodes (Skylake processor (2x16 @ 2.1 GHz) 96 GB RAM, 32 cores per node, Berkeley High Performance Computing) for CONUS.



Goal: produce NO_2 *a priori* profiles at native TEMPO resolution using **machine learning on meteorological forecast/analysis and a high resolution emissions inventory.**

Case Study - Atlanta, GA, USA

Method

Meteorology

NO Emissions (anthropogenic & lightening)

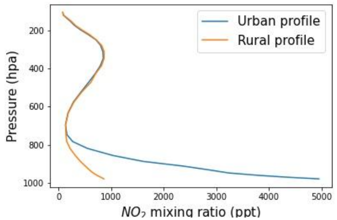
35 km Surrounding NO Emissions

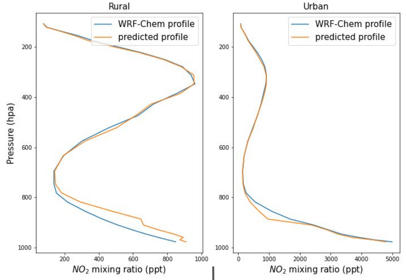
6 years
3.45 million observations (~75%)

2 Years
1.15 million observations (~25%)

Train

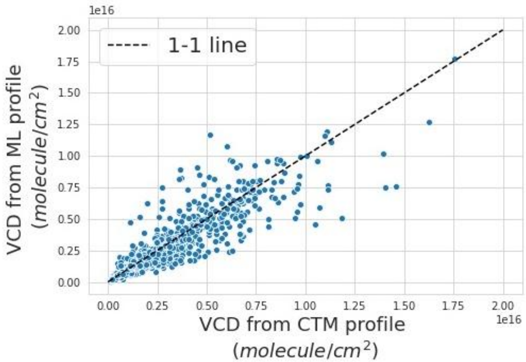
Predict & Evaluate

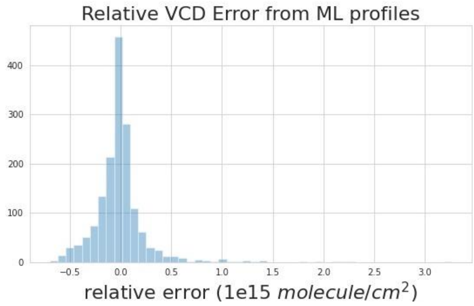


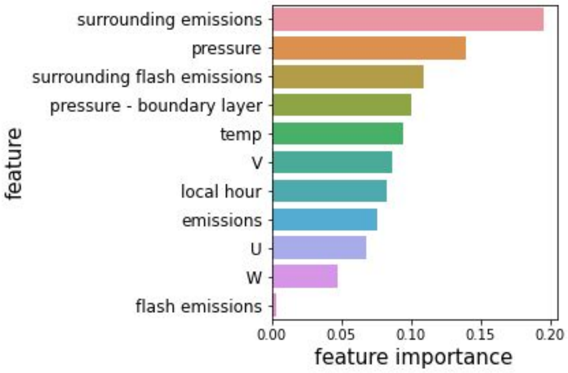


VCD

Results







Summary

Machine learning is a computationally efficient method for producing NO2 vertical profiles at native TEMPO resolution in **real time**.

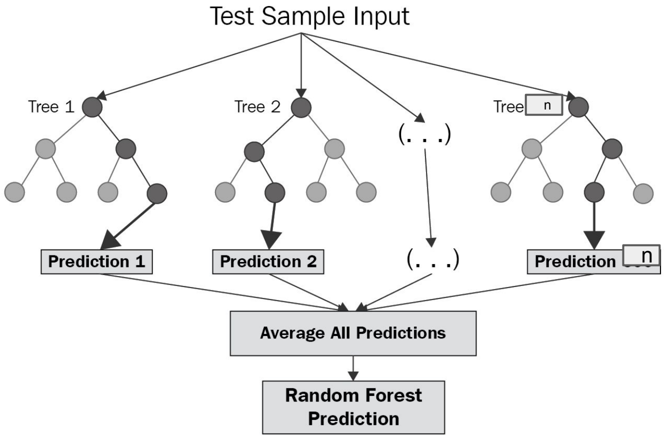
Our model is generalizable because it can capture variability of urban and rural domain

Future Work/Next Steps

Scale up to CONUS - 13.8 billion observations/year!

Supplement

Random Forest Ensemble



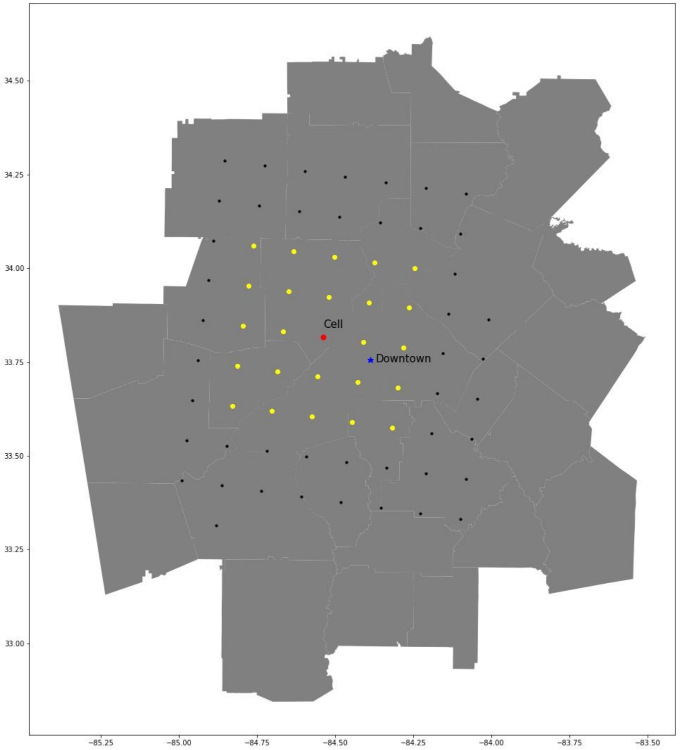
https://towardsdatascience.com/random-forest-and-its-implementation-71824ced454f

Library

scikit learn

Parameters

Num trees: 100
Criterion for split: MSE
Max features: 10
Bootstrap: True



Range of Data

<u>Years:</u>
2005
2007-2009
2011-2014

<u>Month:</u>
July

<u>Time:</u>
5am - 5pm
(SZA < 80°)
Hourly

<u>Vertical Layers:</u>
29

<u>Geo Coordinates</u>
Urban Dataset:
Lon [-85, -84]
Lat [33.3, 34.3]

Rural Dataset:
Lon [-84, -86]
Lat [33, 34]

List of features

<u>Each Observation:</u>
Local Hour
Pressure
Temperature
U, V, W
Flash Emissions
(CG & IC)
Pressure - PBL
NO Emissions

Surrounding Features:
35km Flash Emissions
(CG & IC)
35km NO Emissions