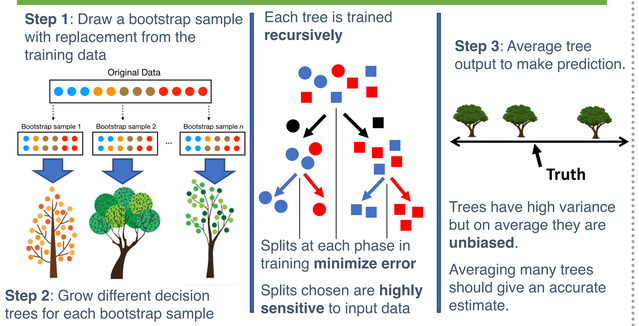
Using machine learning and the GOCI satellite instrument to map fine particulate matter air quality in East Asia

Drew C. Pendergrass[1], Shixian Zhai[1], Jhoon Kim[2,3], Ja-Ho Koo[2], Seoyoung Lee[2], Minah Bae[4], Soontae Kim[4], Hong Liao[5], and Daniel J. Jacob[1]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, Mass., USA [2]Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea [3]Particulate Matter Research Institute, Samsung Advanced Institute of Technology (SAIT), Suwon, South Korea [4]Department of Environmental and Safety Engineering, Ajou University, Suwon., South Korea [5]Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China

Abstract. We use 2011-2019 aerosol optical depth (AOD) observations from the Geostationary Ocean Color Imager (GOCI) instrument over East Asia to infer 24-h daily surface fine particulate matter ($PM_{2.5}$) concentrations at continuous 6x6 km^2 resolution over eastern China, South Korea, and Japan. This is done with a random forest (RF) algorithm applied to the gap-filled GOCI AODs and other data, including information encoded in GOCI AOD retrieval failure, and trained with $PM_{2.5}$ observations from the three national networks. The predicted 24-h GOCI $PM_{2.5}$ concentrations for sites entirely withheld from training in a ten-fold crossvalidation procedure correlate highly with network observations ($R^2 = 0.89$) with single-value precision of 26-32% depending on country. The area-weighted and population-weighted trends of GOCI $PM_{2.5}$ concentrations for eastern China, South Korea, and Japan show steady 2015-2019 declines consistent with surface networks. Further examination of GOCI $PM_{2.5}$ fields for South Korea identifies hotspots where surface network sites were initially lacking and shows 2015-2019 $PM_{2.5}$ decreases across the country except for flat concentrations in the Seoul metropolitan area.

The random forest algorithm

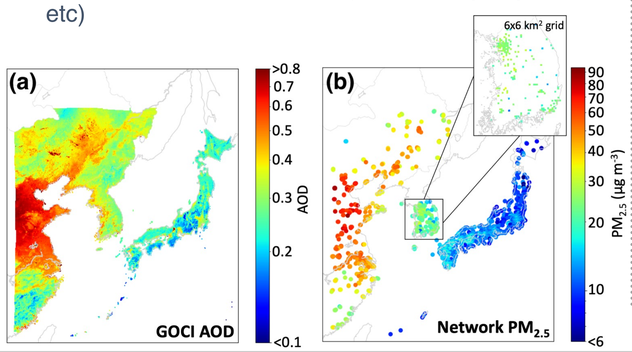


Training data

We train our random forest (RF) machine learning algorithm to predict 24-hr surface $PM_{2.5}$ observed at sites in eastern China, South Korea, and Japan.

The RF algorithm predicts surface PM2.5 using:

- Gap-filled GOCI AOD data (using fusion of satellite, GEOS-Chem, and statistical methods)
- ERA5 reanalysis meteorological fields (boundary layer height, 2m temperature, relative humidity, 10m u/v winds, sea level pressure)
- Location and time metadata (latitude, day of year, etc)



We **trained a machine learning algorithm** on 2011-2019 aerosol optical depth and air quality network data.

We then **produced a continuous 24-h $PM_{2.5}$ dataset** for eastern China, South Korea, and Japan at 6x6 km^2 resolution.

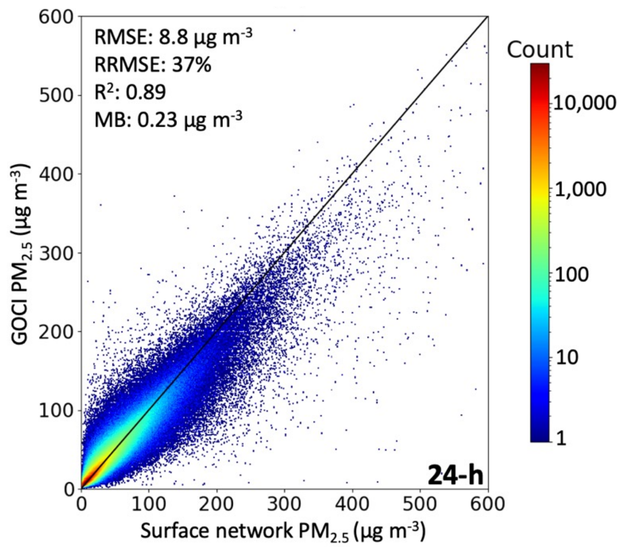
Use the GOCI $PM_{2.5}$ data

Read the AMT paper

Random forest model accuracy

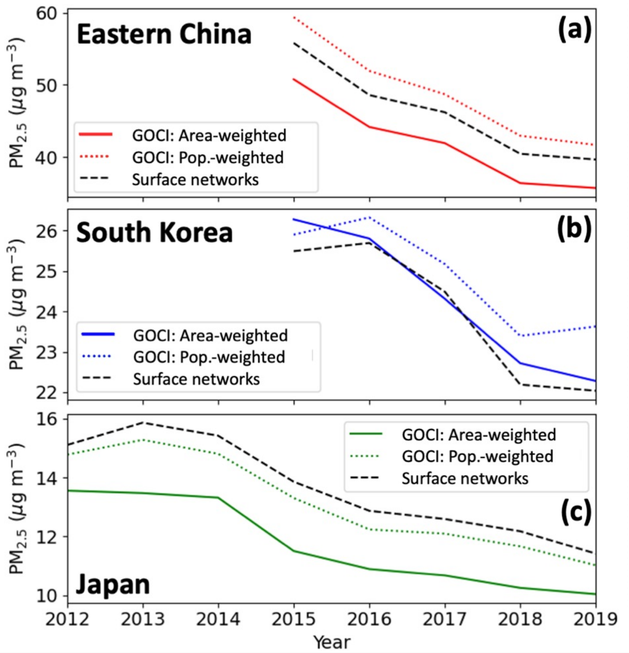
Below is a scatterplot, color-coded by count, comparing surface observations of 24-h $PM_{2.5}$ to the predicted GOCI $PM_{2.5}$ values in grid cells whose records are entirely withheld from training in a crossvalidation procedure.

When aggregated to annual resolution, RMSE between observed and predicted $PM_{2.5}$ is 3.3 μg m^{-3} corresponding to a relative RMSE of 14%. The prediction captures 96% of observed annual variability. Equivalent 24-h statistics in figure.



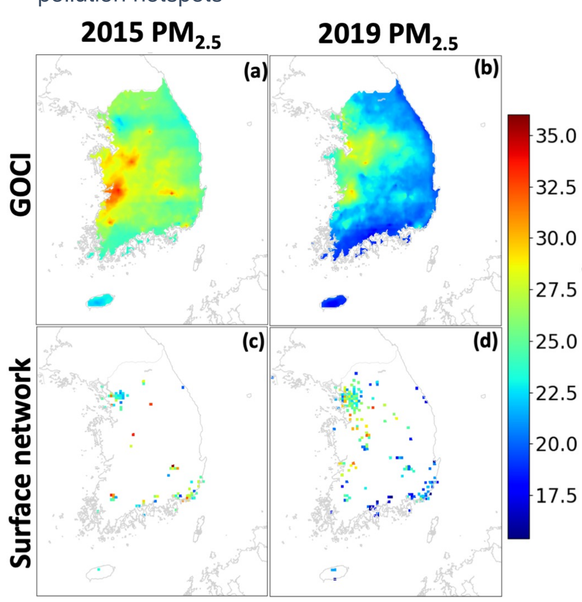
National trends in $PM_{2.5}$

The $PM_{2.5}$ networks show decreasing trends in all three countries and these trends are consistent with the GOCI $PM_{2.5}$ for both areal and population-weighted means. However, the $PM_{2.5}$ networks in eastern China and South Korea underestimate the population-weighted means.



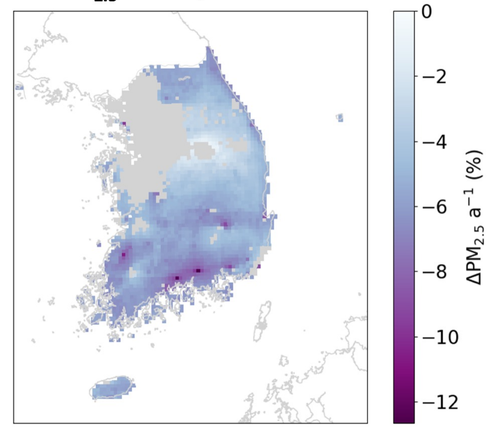
Results for Korea

Coverage is much improved and reveals new pollution hotspots



Fine particulate matter decreases throughout South Korea, but no trend in Seoul despite emissions controls

$PM_{2.5}$ trends, 2015-2019



Acknowledgements

This work was funded by the Samsung $PM_{2.5}$ Strategic Research Program and the Harvard-NUIST Joint Laboratory for Air Quality and Climate (JLAQC). GOCI data was provided by Korea Institute of Ocean Science and Technology (KIOST). DCP was funded by a US National Science Foundation Graduate Fellowship.

Contact information and links

Contact Drew Pendergrass at pendergrass@g.harvard.edu

Read the AMT paper: https://doi.org/10.5194/amt-15-1075-2022

Download the GOCI $PM_{2.5}$ data: https://doi.org/10.7910/DVN/0L3IP7

Please use this data!