Contribution ID: **16**   Type: **Oral presentation**

Predicting PM2.5 concentration across India using multisource earth observation data and machine learning

Thursday, 9 June 2022 12:20 (20 minutes)

According to the World Health Organisation, 3.7 million people around the world died in 2012 as a result of outdoor air pollution. In India, according to the Central Pollution Control Board (CPCB) in India, 43.5% of children have reduced lung function and breathing problems. Air pollution is a major environmental health problem that affects people in developed and developing countries alike. With millions of people dying prematurely every year as a direct result of poor air quality, it has never been more important to monitor the air we breathe.

Various approaches have been proposed to model PM 2.5 (particulate matter with an aerodynamic diameter less than 2.5 microns) in the recent decade, with satellite-derived aerosol optical depth, land-use variables, and several meteorological variables as major predictor variables. The goal of this study is to develop and compare an ensemble-based regression machine learning model and predictor variables to estimate daily PM 2.5 across contiguous India at a resolution of 1km * 1km. By combining data from multiple sources including MODIS (NASA), Sentinel5P (ESA), meteorology, and landcover information from ECMWF Copernicus services, the prediction model has been trained to predict the PM 2.5 concentrations.

Findings from our models show that it does a significantly great job in predicting PM 2.5 daily, monthly, and annual concentrations at an unprecedented spatial resolution and accuracy across pan India. A Cross-validation (CV) using more than 300 ground monitors both spatially and temporally shows high and stable accuracy with a coefficient of determination of 0.85, a root-mean-square error of 15.57 ug m-3, and a mean prediction error of 9.77 ug m-3. This shows good agreement between CV predictions and observations. The model implements variable importance to examine the effects of each predictor on the target PM 2.5 concentration estimation. In general, our model is robust, and the estimates are in line with regular ground validation. It may thus also be useful for applications in related air pollution studies, especially those focused on urban areas.

Primary authors: Mr KOCHHAR, Ishaan (Blue Sky Analytics); AHMED, Saheel (Blue Sky Analytics); Mr PURWAR, Kshitij (Chief Technology Officer)

Presenters: Mr KOCHHAR, Ishaan (Blue Sky Analytics); AHMED, Saheel (Blue Sky Analytics)

Session Classification: Using Copernicus and climate data

Track Classification: UEF2022