*URSI-RCRS 2024, ARIES/GEHU, Bhimtal, India, 22 - 25 October, 2024*

# A machine learning approach to estimate $PM_{2.5}$ in Western India: combining satellite remote sensing and ground-based measurements

Yash Dahima[(1)], Narendra Ojha[(2)], Imran Girach[(3)], Amit Sharma[(4)], and Aditya Vaishya[(*1)]
(1) School of Arts and Sciences, Ahmedabad University, Ahmedabad, 380009, India; e-mail: yash.d2@ahduni.edu.in; e-mail: aditya.vaishya@ahduni.edu.in
(2) Physical Research Laboratory, Ahmedabad, 380009, India; e-mail: ojha@prl.res.in
(3) Space Applications Centre, Indian Space Research Organisation, Ahmedabad, 380015, India; email: imran.girach@gmail.com
(4) Department of Civil & Infrastructure Engineering, Indian Institute of Technology Jodhpur, Jodhpur, 342030, India; email: amit.sharma@iitj.ac.in

Particulate air pollution is a major environmental and public health concern. Particulate matter of size <= 2.5µm ($PM_{2.5}$) can penetrate deep into the lungs and cause severe respiratory and cardiovascular diseases. However, due to remarkably diverse emissions and strong impact of atmospheric dynamics, it is very challenging to model spatio-temporal distribution of $PM_{2.5}$ from conventional techniques. Therefore, continuous monitoring of $PM_{2.5}$ is essential not only in the urban agglomerates with high population density but also in the rural areas for quantifying the impacts and inform mitigation strategies. However, making dense measurements of $PM_{2.5}$ over large geographical regions is not viable economically and the data driven approaches especially machine learning (ML) can be applied to infer $PM_{2.5}$ concentrations. In this regard, we present a robust approach for estimating $PM_{2.5}$ in different cities located in Western India affected by natural dust emissions and strong anthropogenic emissions. ML model has been trained with satellite-based observations of aerosol properties (aerosol optical depth, fine-mode and coarse-mode fractions, and single scattering albedo), and meteorological reanalysis (ERA5) for estimating $PM_{2.5}$ variability. We employed various ML techniques viz. random forest model, artificial neural network, etc. with a focus on evaluating its performance against ground data. Results indicate that machine learning techniques, particularly Random Forest, demonstrated superior performance in estimating $PM_{2.5}$ levels. The root mean square error ranged from 20 – 40 $\mu g/m^3$, mean absolute error 13 – 30 $\mu g/m^3$, and $R^2$ from 0.1 – 0.5 for different stations. The study underscores the significance of advanced modeling approaches in reliably predicting $PM_{2.5}$ concentrations by exploiting remote sensing data. The study has implications to better understand the relative importance of different parameters in governing air quality of the region and thereby offering insights for air quality monitoring and management.