Machine Learning Algorithm for Estimating Surface PM2.5 in Thailand

Pawan Gupta1,2*, Shanshan Zhan3, Vikalp Mishra3,7, Aekkapol Aekakkararungroj4, Amanda Markert3,7, Sarawut Paibong5, Farrukh Chishtie4,6

1Universities Space Research Association (USRA), Huntsville, USA
2Marshall Space Flight Center, Huntsville, AL, USA
3Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, USA,
4Asian Disaster Preparedness Center, Bangkok, Thailand
5Thai Pollution Control Department, Bangkok, Thailand
6Spatial Informatics Group, Pleasanton, CA, USA
7SERVIR Science Coordination Office, NASA Marshall Space Flight Center, Huntsville, AL, USA,

*Corresponding author: Tel: 1-2564687651
Email Address: pawan.gupta@nasa.gov

Supplementary Material
Fig. S1. Inter-comparison between observed (x-axis) and estimated (y-axis) PM2.5 for the data sets used in training ten machine learning models.
Figure S2. Same as Figure S1 but for validation data sets.
Figure S3. Fractional feature importance of input parameter in estimating PM2.5 at hourly time scale using ensembled random forest model.
Figure S4 – The variability of observed, MERRA2, and MERRA2_ML PM2.5 averaged over three months period are presented for March-April-May (MAM), June-July-August (JJA), September-October-November (SON), and December-January-February (DJF).
Figure S5 – The hourly MERRA2_ML PM2.5 estimates are averaged over 24-hour period and compared with observed daily mean PM2.5 for all the stations and for the entire time period. The color bar show density of points in the scatter plot.
Figure S6. The 24-hour averaged PM2.5 values from observed, MERRA2_ML, and MERRA2 were used to evaluate day-to-day variability averaged over all the station.