論文 / 著書情報
Article / Book Information

Title	Modeling of near-surface PM2.5 using machine learning regression informed by a chemistry transport model: A case study in Kanto Region, Japan
Author	Roseanne Ramos, Alvin Christopher Varquez, Maria Deandra Crisostomo Andal, Zeyu Chen, Atsushi Inagaki, Manabu Kanda
Citation	JpGU-AGU Joint Meeting 2026
Pub. date	2026, 5

Modeling of near-surface PM2.5 using machine learning regression informed by a chemistry transport model: A case study in Kanto Region, Japan

Roseanne RAMOS[1,2], Alvin Christopher VARQUEZ[1], Maria Deandra ANDAL[1], Zeyu CHEN [1], Atsushi INAGAKI[1], Manabu KANDA[1]

[1] Department of Transdisciplinary Science and Engineering, Institute of Science Tokyo, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
[2] Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City 1101, Philippines

Corresponding author: Roseanne RAMOS
E-mail: ramos.r.4bbb@m.isct.ac.jp; rvramos@up.edu.ph

Abstract

Fine particulate matter ($PM_{2.5}$) levels are regulated globally due to their significant impact on human health. In urban environments, sensors deployed on the ground measure $PM_{2.5}$ for ambient and roadside monitoring. These sensor networks often do not give sufficient measurements to report the spatial variability of ground $PM_{2.5}$ concentrations in highly polluted cities for air quality management. Previous studies utilized statistical and dispersion models to obtain high-resolution maps of $PM_{2.5}$ concentrations and quantify pollution levels in unsampled locations. However, these approaches were limited to ground observations and existing prediction models cannot capture the complex non-linear relationships of $PM_{2.5}$ with environmental variables affecting their transport in the atmosphere. To address these gaps, we propose a machine learning (ML)-based modeling framework for estimating near-surface $PM_{2.5}$ informed by an online-coupled chemical transport and meteorology model. The modeling framework considers physical and chemical processes through numerical simulations using Weather Research and Forecasting model coupled with Chemistry (WRF-Chem v4.6) and applies ML algorithms in developing a prediction model for near-surface $PM_{2.5}$. WRF-Chem model utilized the combined scheme of Model for Ozone and Related Chemical Tracers (MOZART) and Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), with urban grids represented by a Single Layer Urban Canopy model (SLUCM) combined with distributed urban morphological parameters, to simulate the chemical processes for the aerosol species in a 120x120 nested grid with spatial resolutions of 15km, 3km and 1.5km. Inputs for the emission sources were derived from the databases of the Japan's Study for Reference Air Quality Modeling (J-STREAM). The prediction model was designed to be operational for regional applications, with our study site in Kanto Region of Japan. Hence, 17 features that are globally-measured operationally – aerosol optical depth (AOD), rainfall, solar radiation (solarrad), surface pressure, relative humidity (RH), temperature at 2m (T2), rolling averages of T2 (T2_prev6hrs and T2_next6hrs), wind speed, boundary layer height (BLH), surface albedo, surface emissivity, leaf area index (LAI), vegetation fraction, terrain height, anthropogenic heat emissions (AHE) and distance from the coastline – were extracted from WRF-Chem simulation outputs and were used as predictors in the ML to evaluate $PM_{2.5}$. ML-based

regression tasks were conducted using the Extreme Gradient Boosting (XGBoost) algorithm. The trained XGBoost model (with cross-validation test metrics RMSE=1.516 μg/m^3, MAE=0.934 μg/m^3, R^2=0.989) was then applied for inference on observation variables extracted from climate reanalysis and satellite observations from the Moderate Resolution Imaging Spectrometer (MODIS) for a full year. The prediction model for near-surface $PM_{2.5}$ shows that AOD, T2 and rolling averages, RH, surface pressure, LAI, BLH and wind speed were the most important predictors. In future work, the improvements on the model will comprise of additional feature engineering work on the input datasets, fine-tuning of model hyperparameters and re-evaluation of model accuracy, generalizability and transferability for application in other areas.

Acknowledgements: This work is supported by the Doctoral program scholarship provided by the Department of Science and Technology - Engineering Research and Development for Technology (DOST-ERDT) of the Philippines, the JSPS KAKENHI Grant Numbers JP21K14249, JP25K00040, and the Environment Research and Technology Development Fund (JPMEERF2024RA1) of the Environmental Restoration and Conservation Agency provided by Ministry of the Environment of Japan. The authors would also like to acknowledge Dr. Satoru Chatani of the National Institute for Environmental Studies (NIES) for sharing the raw emissions database from J-STREAM.