Supplement of Atmos. Meas. Tech., 18, 1471–1484, 2025
https://doi.org/10.5194/amt-18-1471-2025-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

Estimating hourly ground-level aerosols using Geostationary Environment Monitoring Spectrometer aerosol optical depth: a machine learning approach

Sungmin O et al.

Correspondence to: Seon Ki Park (spark@ewha.ac.kr)

Supplementary figures

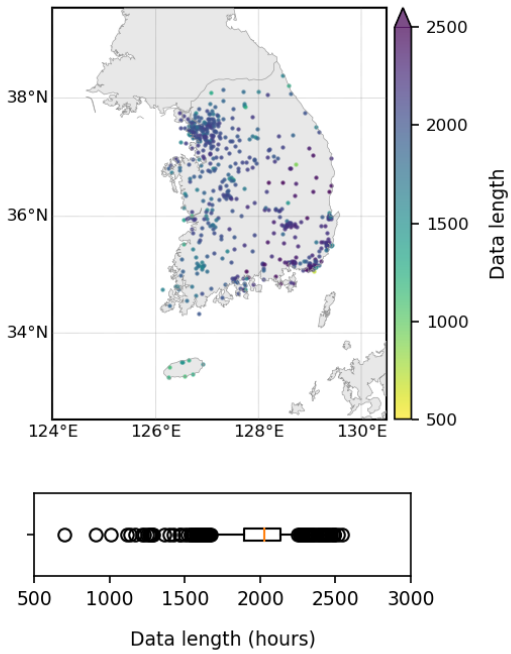

Figure S1. Length of GEMS AOD-PM10 data pairs at each station. Box plot whiskers show the 0.1 to 0.9 quantiles of the data length across all stations.

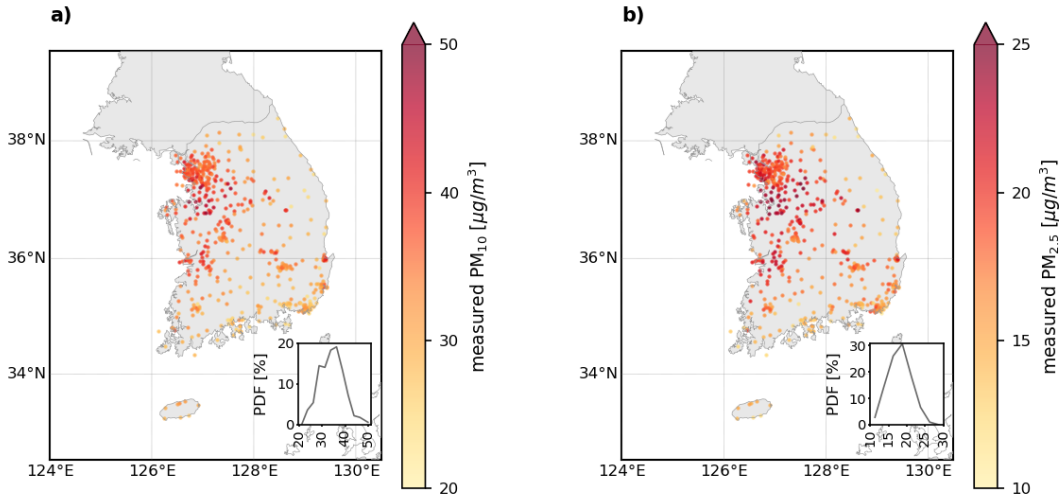

Figure S2. Average concentrations of (a) PM10 and (b) PM2.5 at urban air quality monitoring stations for the period from January 2022 to December 2023 in South Korea. All available PM measurements are used regardless of the GEMS AOD data availability.

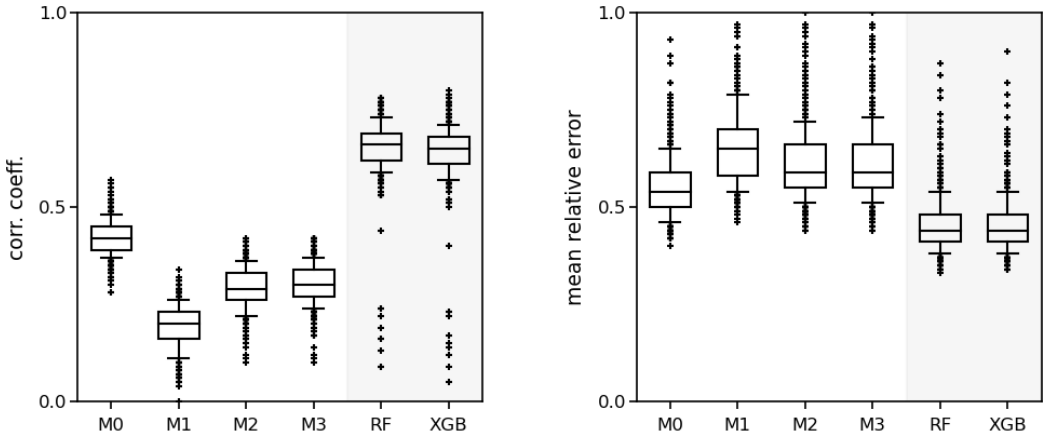

Figure S3. Comparison of model performance at each station based on (a) correlation and (b) mean relative error. For M0, the estimated PM10 for a given month is calculated using the corresponding monthly average. M1, M2, and M3 show the performance of empirical linear models with different sets of input variables; M1 with AOD data only, M2 with AOD, BLH, RH, and TEMP, and M3 with all meteorological variables. RF shows performance of random forest models that are used in the main analysis, while XGB represents the performance of XGBoost models.

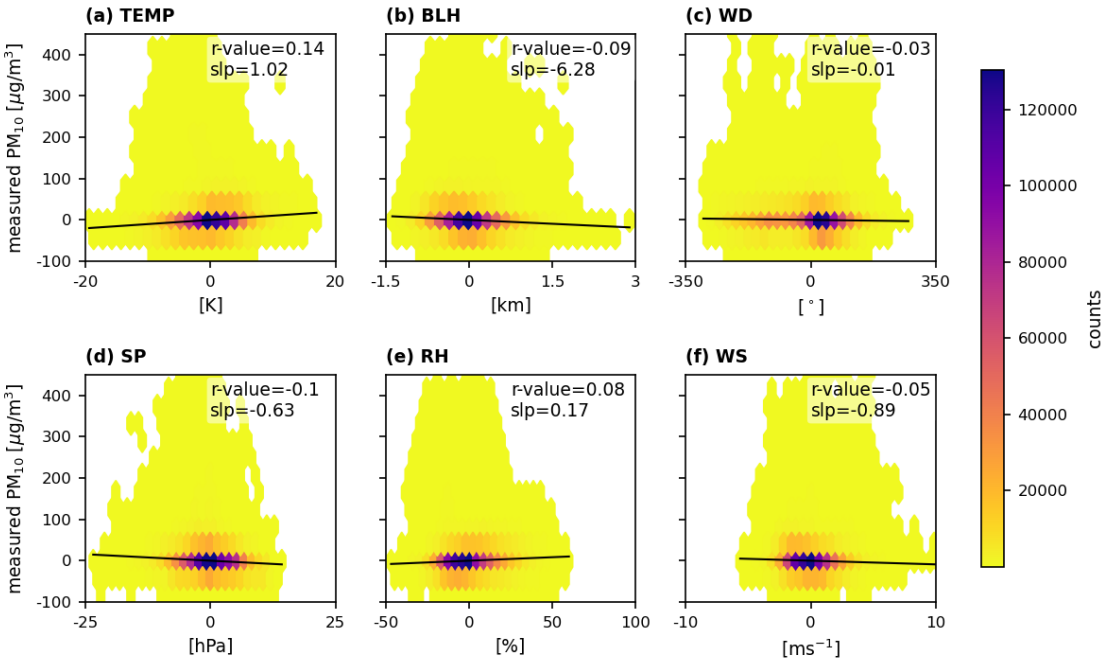

Figure S4. Comparison between anomalies of PM10 measurements and meteorological variables at all stations. TEMP, BLH, WD, SP, RH, and WS represent air temperature, boundary layer height, wind direction, surface pressure, relative humidity, and wind speed, respectively.

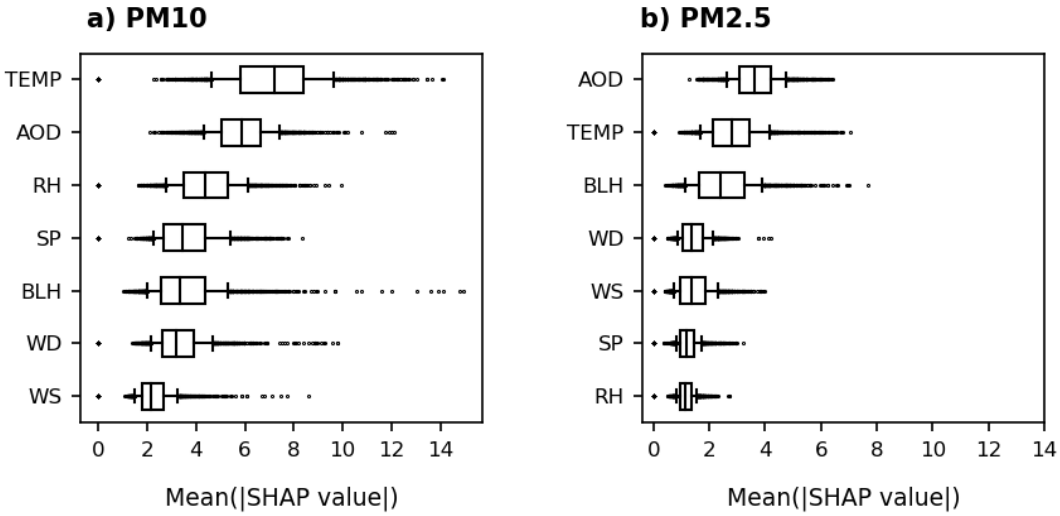

Figure S5. Same as in Fig. 4 in the main text, but for XGB model simulations.

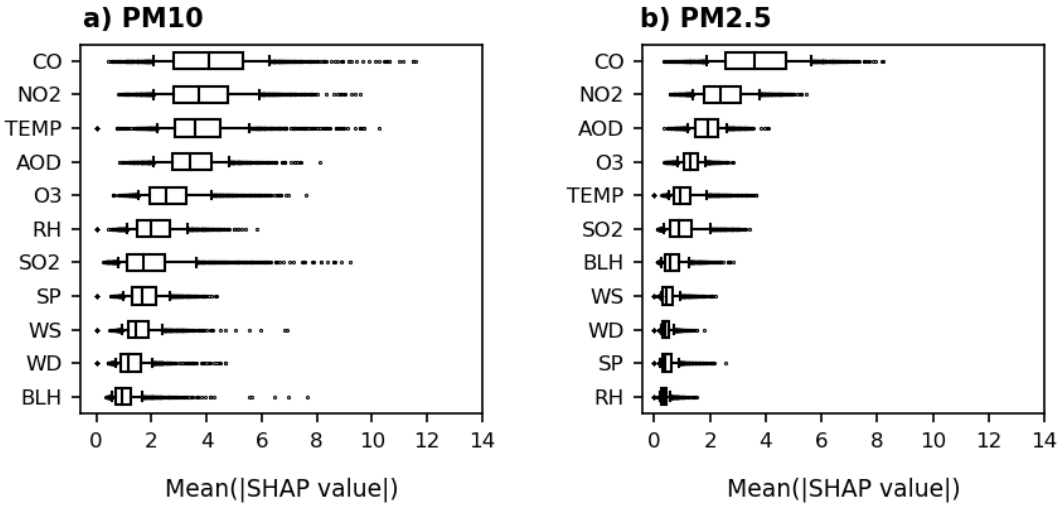

Figure S6. Same as in Fig. 4 in the main text, but for RF model simulations with additional input features (chemical pollutants).

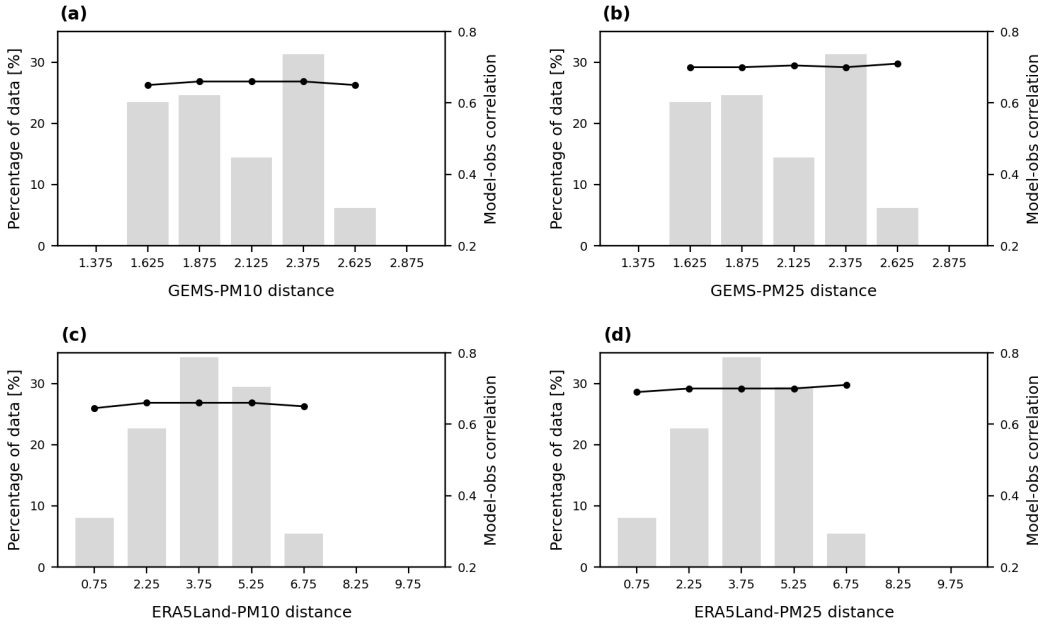

Figure S7. Model performance and data distribution as a function of the distance between input features and target variables. Panels (a) and (b) show results for GEMS AOD paired with PM10 and PM2.5, respectively, while panels (c) and (d) present results for ERA5 reanalysis meteorological variables paired with PM10 and PM2.5, respectively. Gray bars indicate the percentage of data within each distance range, and black dots connected by lines represent the average model performance (correlation) for each distance range.