Abstract

The mathematical solution to estimate surface fine particulate matter ($PM_{2.5}$) from columnar aerosol optical depth (AOD) includes complex variables and involves a bunch of assumptions. Hence, researchers tend to use training-based models to predict $PM_{2.5}$ from AOD. Here, we integrated regulatory composite $PM_{2.5}$ measurements, high-resolution satellite AOD, reanalysis meteorological parameters, and a few other auxiliary parameters to train ten different regression models. The performance of these (seven statistical and three machine learning) models was evaluated and inter-compared to identify the best performing model. The accuracies of the model predicted $PM_{2.5}$ were quantified based on the coefficient of determination (R^2), mean absolute bias (MAB), normalized root mean square error (NRMSE), and other relevant regression coefficients. The model's performance on unseen data was investigated in terms of 10-fold cross-validation (CV) and Leave-one station-out CV (LOOCV). For this exercise, we considered the case of NCT-Delhi due to: (i) the availability of dense regulatory $PM_{2.5}$ measurements, (ii) the possibility of understanding the model performance over a large range of $PM_{2.5}$ (the daily mean $PM_{2.5}$ values ranged between ∼ 4 and 492 μg m^{-3} during the study period), and (iii) the scope of better understanding the influence of extreme meteorological conditions (e.g. the ambient surface temperature varies between ∼5 and 40 °C during a calendar year) on the AOD-$PM_{2.5}$ relationship. All the models were trained using data collected for the year 2019 (a non-COVID year). Among models under investigation, Machine Learning (ML) models performed better with R^2, MAB, and NRMSE values for the CV exercises ranging between 0.88 and 0.93, 14.1 and 18.2 μg m^{-3}, and 0.18 and 0.23, respectively. The generalizability of the results obtained in this study was discussed.