Correcting Measurement Error in Satellite Aerosol Optical Depth with Machine Learning for Modeling $PM_{2.5}$ in the Northeastern USA

Allan C. Just, Margherita M. De Carli, Alexandra Shtein, Michael Dorman, Alexei Lyapustin, and Itai Kloog

This supplementary material supports the main text as follows:

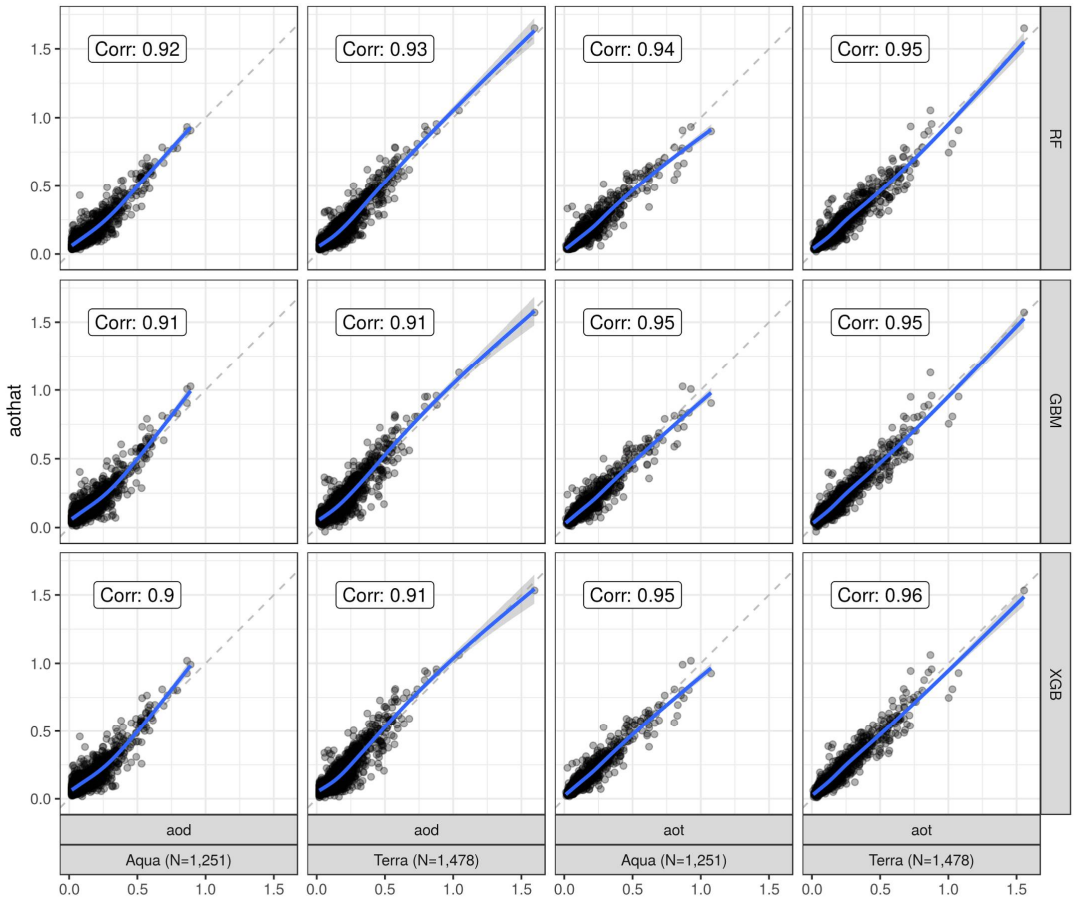

Figure S1. Predictions in testing data from three modeling approaches that correct measurement error versus either MAIAC AOD or AERONET AOT for Aqua and Terra.

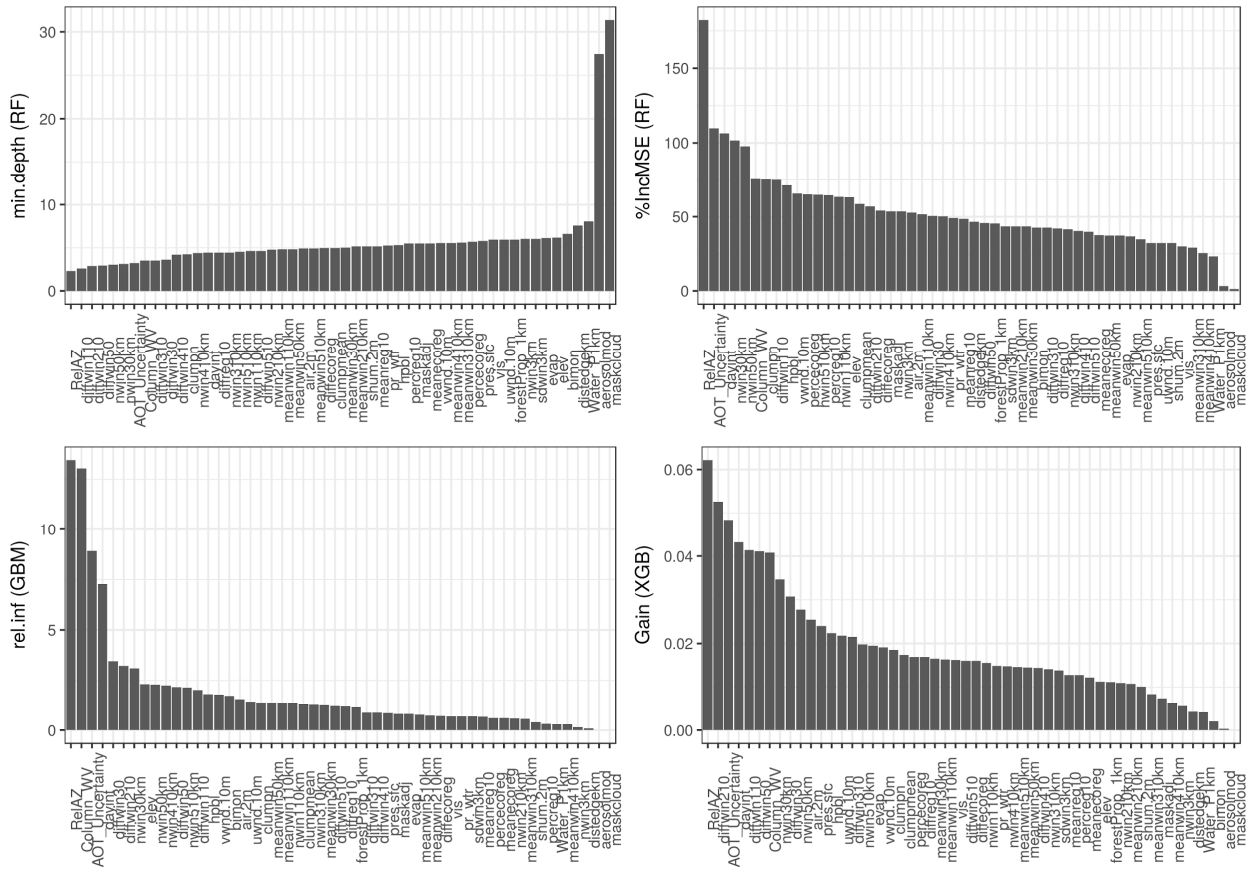

Figure S2. Variable importance measurements for three ensemble approaches to predicting measurement error in the MAIAC AOD training dataset (n=7,280 for Aqua). For RF model, variable importance is shown using both minimum depth and % increase in MSE, while the GBM model is summarized with rel.inf measure and XGBoost uses the gain in node purity.

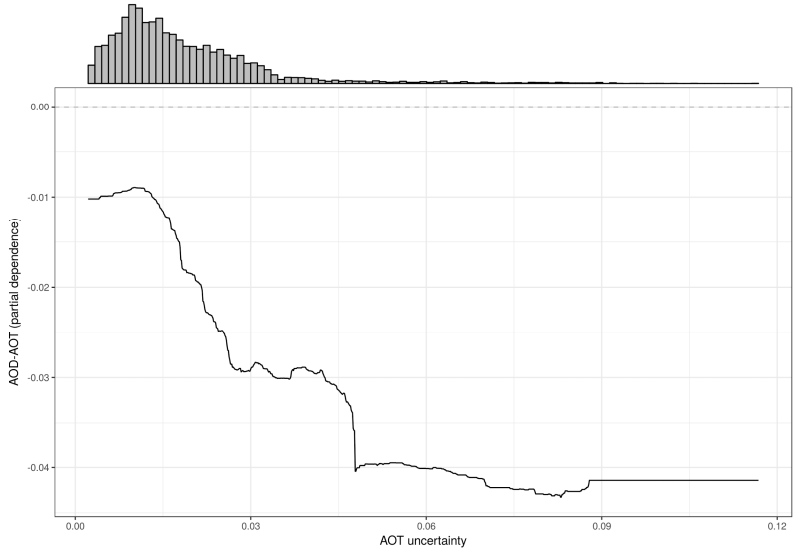

Figure S3. Partial dependence plot for the MAIAC AOT uncertainty variable in the Aqua training dataset (n=7,280) from the GBM approach. Zero on the y-axis is no measurement error. The histogram shows the distribution of (unitless) AOT uncertainty.

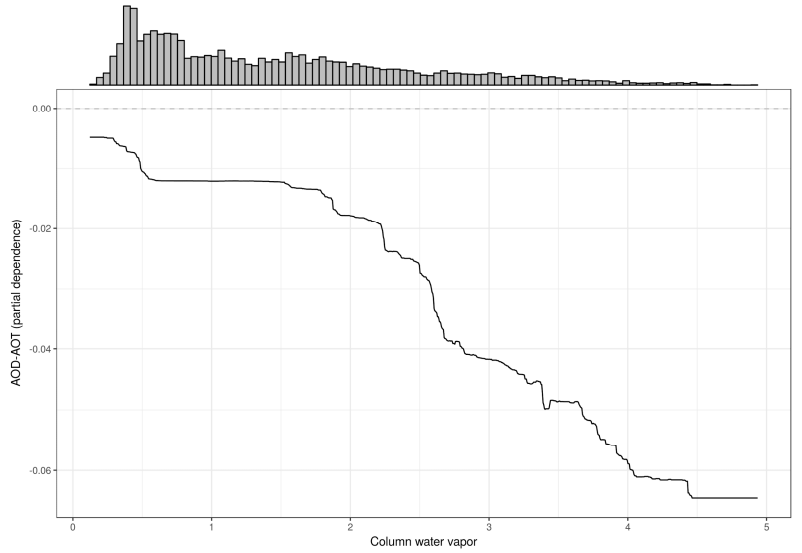

Figure S4. Partial dependence plot for column water vapor in the Aqua training dataset (n=7,280) from the GBM approach. Zero on the y-axis is no measurement error. The histogram shows the distribution of column water vapor (cm).

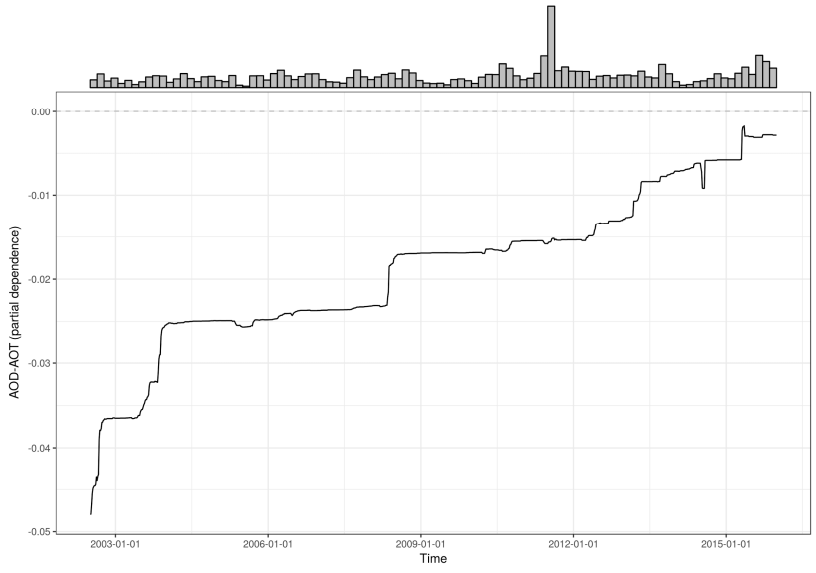

Figure S5. Partial dependence plot for the long-term time trend in the Aqua training dataset (n=7,280) from the GBM approach. Zero on the y-axis is no measurement error. The histogram shows the distribution of training data in time with a peak during the 2011 DRAGON campaign in the DC-area.