Supplement of Atmos. Meas. Tech., 18, 1149–1162, 2025
https://doi.org/10.5194/amt-18-1149-2025-supplement

Atmospheric Measurement Techniques
Open Access

Supplement of

Inversion algorithm of black carbon mixing state based on machine learning

Zeyuan Tian et al.

Correspondence to: Jiandong Wang (jiandong.wang@nuist.edu.cn)

S1 Construction of feature dataset

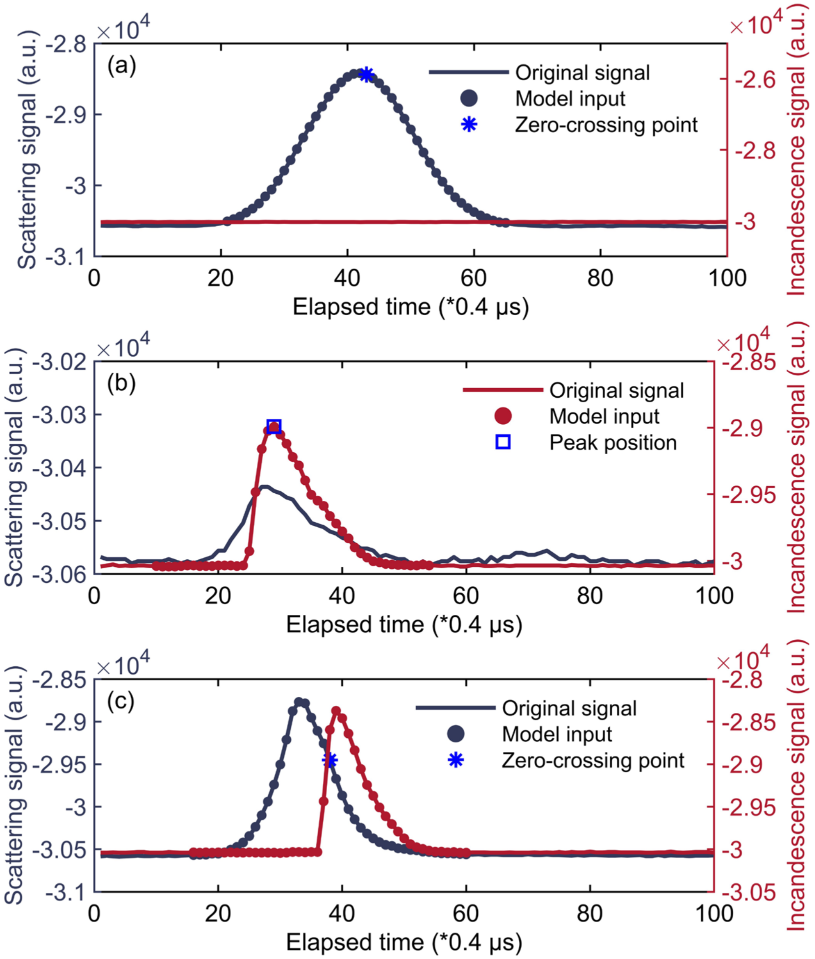

Figure S1. Relationship between the original SP2 signals (line plots) and the feature signals used in machine learning model construction (scatter plots) for different particle types: (a) purely scattering particles; (b) externally mixed BC; (c) internally mixed BC. The method for selecting feature signals used in retrieving the core diameter (D_c) of BC-containing particles is identical to that used for externally mixed BC.

S2 LightGBM model

For each type of particle, the number of samples used in machine learning reaches an order of 10^5. The dataset is randomly partitioned into training and testing sets with a ratio of 7:3, and this unbiased selection method helps improve the reliability and generalizability of the model. To demonstrate the effectiveness of this data division, we analyse the normalized number size distributions of particle diameter (D_p) in both the training and testing sets for the internally mixed BC inversion model. As shown in Fig. S2, the consistent distributions between these two sets validate the rationality of our data partitioning approach.

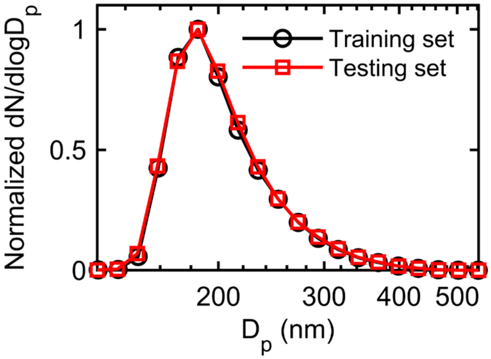

Figure S2. The normalized number size distribution of the training set (black marks and line) and testing set (red marks and line) used in the D_p inversion model for internally mixed BC.

S3 Comparison of D_p inversion results between training and testing sets for internally mixed BC

As illustrated in Fig. S3, the coefficients of determination R^2 for the training and testing sets are 0.99 and 0.98, respectively. These high R^2 values indicate excellent model performance, with the close R^2 values demonstrating the model's strong predictive capability and good generalization performance.

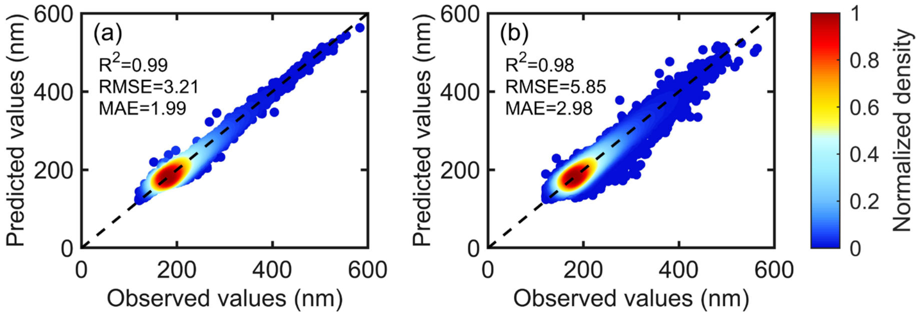

Figure S3. The D_p inversion results of internally mixed BC for both training set (a) and testing set (b).

S4 SHAP interpretations

S4.1 SHAP analysis for the D_p inversion model of internally mixed BC

Figure S4 shows the scattering and incandescence signals of 6 typical particles used as input for the machine learning (ML) model. Figure. S4 also indicates the specific distribution of seven important features, identified by the SHapley Additive exPlanation (SHAP) method, within the incandescence signals of different particles. These 6 typical particles have similar D_c, but the coating thickness (CT) varies, decreasing sequentially from panel (a) to (f). Since the evaporation time of the coating of BC-containing particles is positively correlated with the CT, the time delay (Δt) decreases sequentially from panel (a) to (f). When the coating is thicker and Δt is longer, the seven important incandescence signal features are mostly distributed at the baseline position where the incandescence signal has not yet begun to rise, and their corresponding values are smaller. As the coating becomes thinner and Δt decreases, the position where the incandescence signal starts to rise from the baseline earlier. Consequently, the relative positions of the seven important incandescence signal features gradually approach the peak of the incandescence signal, and their corresponding values increase. Therefore, these seven important incandescence signal features exhibit the same characteristic: as their values increase, the corresponding SHAP values decrease, leading to a reduction in the D_p of the BC-containing particles predicted by the model.

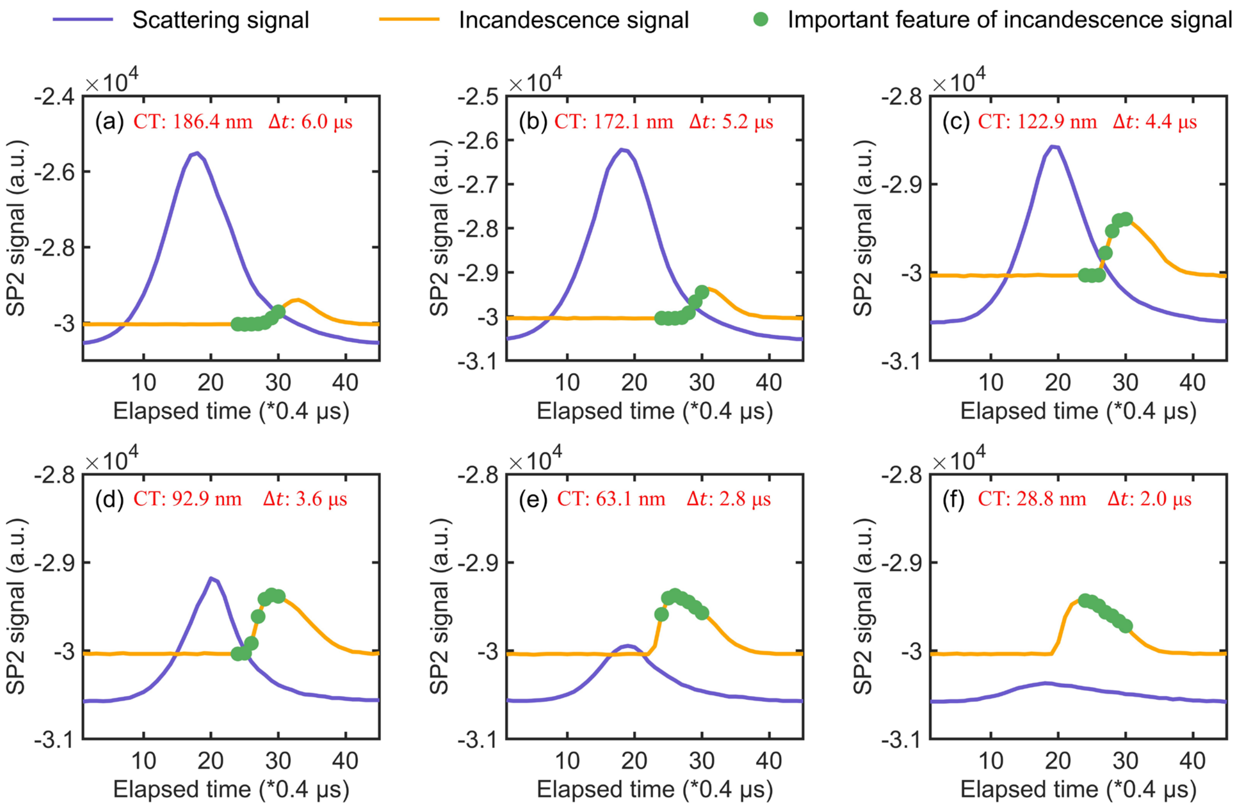

Figure S4. The scattering and incandescence signals of 6 typical particles used as input features for the ML model. The green dots scattered on the incandescence signals indicate the specific positions of seven important features within the incandescence signals, as identified by the SHAP method. From panel (a) to (f), the CT of the particles decreases sequentially, and correspondingly, Δt also gradually decreases.

S4.2 SHAP results for the optical cross-section inversion model of internally mixed BC

Figures S5a and S5b illustrate the SHAP summary plots for the inversion models of absorption and scattering cross-sections of internally mixed BC, respectively. Due to the small magnitude of the optical cross-section, the SHAP values shown here have been amplified for clarity. The SHAP summary plot for the absorption cross-section of internally mixed BC shows that most of the top 15 important features are related to the incandescence signal. Specifically, features BBLG28, BBLG29, and BBLG27, corresponding to the feature dimensions near the peak of the incandescence signal, contribute most significantly to the absorption cross-section. Among the top 15 features, several features related to the scattering signal (SCLG12 to SCLG15) are also present. These features, associated with the peak positions of the scattering signal, demonstrate that larger feature values positively contribute to the inversion of the absorption cross-section. This phenomenon can be attributed to the “lensing effect” of the coating (Cappa et al., 2012; Schwarz et al., 2008), which enhances the absorption of the BC core.

For the inversion model of the scattering cross-section of internally mixed BC, scattering signal features are particularly important, showing a positive correlation with the scattering cross-section. Meanwhile, the incandescence signal, which reflects the characteristics of the BC core, also plays an important role in the inversion process.

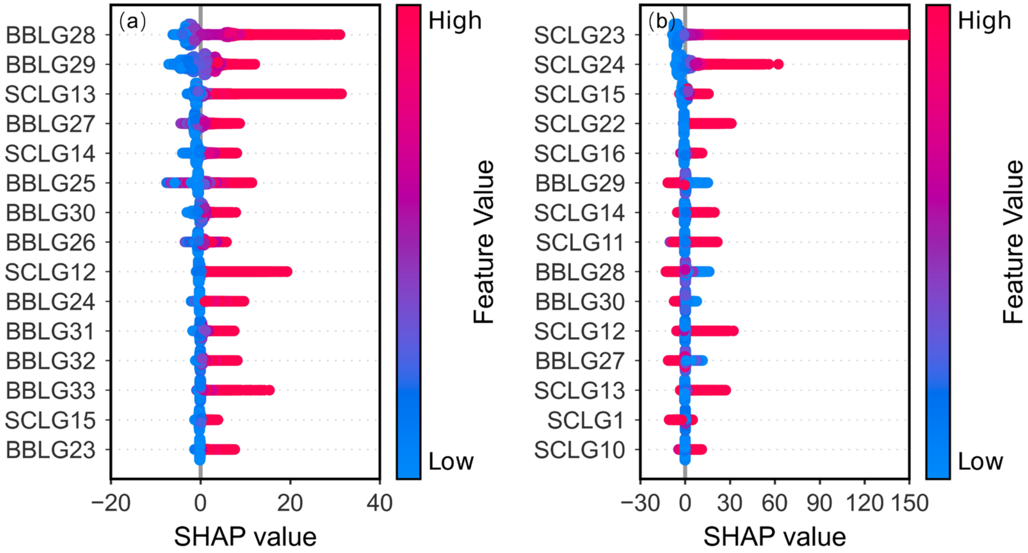

Figure S5. The SHAP summary plot for the optical cross-section inversion model of internally mixed BC: (a) absorption cross-section; (b) scattering cross-section.

S5 Model application

Table S1. The statistical results of D_p and D_c of internally mixed BC obtained by applying the LightGBM inversion model to the single-particle soot photometer (SP2) data in April 2022.

Statistical results	D_c	D_p
R^2	0.99	0.98
RMSE (nm)	0.21	8.28
MAE (nm)	0.14	4.19
Third Quartile (nm)	150.8	236.7
Median (nm)	130.8	200.0
First Quartile (nm)	114.0	177.5

References

Cappa, C. D., Onasch, T. B., Massoli, P., Worsnop, D. R., Bates, T. S., Cross, E. S., Davidovits, P., Hakala, J., Hayden, K. L., Jobson, B. T., Kolesar, K. R., Lack, D. A., Lerner, B. M., Li, S.-M., Mellon, D., Nuaaman, I., Olfert, J. S., Petäjä, T., Quinn, P. K., Song, C., Subramanian, R., Williams, E. J., and Zaveri, R. A.: Radiative Absorption Enhancements Due to the Mixing State of Atmospheric Black Carbon, Science, 337, 1078–1081, https://doi.org/10.1126/science.1223447, 2012.

Schwarz, J. P., Spackman, J. R., Fahey, D. W., Gao, R. S., Lohmann, U., Stier, P., Watts, L. A., Thomson, D. S., Lack, D. A., Pfister, L., Mahoney, M. J., Baumgardner, D., Wilson, J. C., and Reeves, J. M.: Coatings and their enhancement of black carbon light absorption in the tropical atmosphere, J. Geophys. Res., 113, 2007JD009042, https://doi.org/10.1029/2007JD009042, 2008.