

# Predictions of the Optical Properties of Brown Carbon Aerosol by Machine Learning with Typical Chromophores

Ying Wang, Ru-Jin Huang,\* Haobin Zhong, Ting Wang, Lu Yang, Wei Yuan, Wei Xu,\* and Zhisheng An



Cite This: *Environ. Sci. Technol.* 2024, 58, 20588–20597



Read Online

ACCESS |

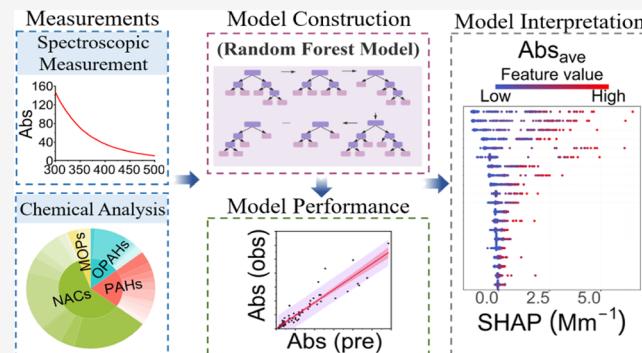
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** The linkages between BrC optical properties and chemical composition remain inadequately understood, with quantified chromophores explaining less than 25% of ambient aerosol light absorption. This study characterized 38 typical chromophores in aerosols collected in Xi'an, with light absorption contributions to BrC ranging from  $1.6 \pm 0.3$  to  $5.8 \pm 2.6\%$  at 365 nm. Based on these quantified chromophores, an interpretable machine learning model and the Shapley Additive Explanation (SHAP) method were employed to explore the relationships between BrC optical properties and chemical composition. The model attained high accuracy with Pearson correlation coefficients ( $r$ ) exceeding 0.93 for the absorption coefficient ( $\text{Abs}_{\lambda}$ ) and surpassing 0.57 for mass absorption efficiency ( $\text{MAE}_{\lambda}$ ) of BrC. It explains more than 80% of the variance in  $\text{Abs}$  and over 50% in  $\text{MAE}$ , significantly improving the understanding of BrC light absorption. Polycyclic aromatic hydrocarbons (PAHs) and oxygenated PAHs (OPAHs) with four and five rings exhibit significant positive effects on  $\text{Abs}_{\lambda}$ , suggesting that similar unidentified chromophores may also notably impact BrC optical characteristics. The model based on chromophore mass concentrations further simplifies studying BrC optical characteristics. This study advances understanding of the relationship between BrC composition and optical properties and guides the investigation of unrecognized chromophores.

**KEYWORDS:** brown carbon, chromophores, machine learning, SHAP, prediction, chemical composition, optical properties



## INTRODUCTION

BrC is an important component of organic aerosol, exerting profound effects on atmospheric radiation and global climate.<sup>1–3</sup> BrC significantly impacts climate by directly absorbing solar radiation. Additionally, it alters the atmospheric photochemistry by reducing the photolysis rates of ozone and volatile organic compounds, which influences aerosol formation pathways and further affects air quality.<sup>1,4–6</sup> BrC can not only be emitted from the incomplete combustion of biomass and fossil fuels,<sup>7–11</sup> but also produced secondarily through multiple atmospheric chemical processes.<sup>12–16</sup> Consequently, BrC contains a diverse chemical composition, and its light absorption undergoes enhancement or bleaching through the dynamic transformations of atmospheric processes.<sup>17,18</sup> Given these characteristics exhibited by BrC, there still exist great challenges in comprehensively understanding the chemical composition and optical properties of BrC. More studies are needed on the chemical characteristics, optical properties and their interrelations within BrC.

Molecular identification of BrC is the foundation for exploring the relationships between the chemical composition and optical properties of BrC. A multitude of studies have conducted molecular characterization and associated optical

measurements of BrC.<sup>4,7,19–32</sup> So far, PAHs, OPAHs, nitrated aromatic compounds (NACs) and methoxyphenols (MOPs) are the widely identified BrC groups in studies of biomass burning, coal combustion and field observations across various regions.<sup>4,9,22,24,33–36</sup> Currently, the chromophores identified in atmospheric samples contribute to OC mass by less than 2%,<sup>4,24</sup> and their light absorption contributions to BrC are generally less than 25%.<sup>4,24,28,34,37</sup> This indicates that the identified BrC chromophores can only explain a relatively small fraction of BrC variations, making it challenging to accurately capture changes in BrC optical properties solely through these chromophores. Therefore, to uncover more potential information about unidentified BrC chromophores and obtain a more comprehensive explanation of the connections between BrC chemical composition and optical properties, it is necessary to not only utilize advanced spectroscopic techniques and mass

Received: August 28, 2024

Revised: October 29, 2024

Accepted: October 30, 2024

Published: November 7, 2024



spectrometry for identification and quantification but also to employ additional tools, such as machine learning.

Machine learning based on the random forest (RF) algorithm presents high prediction accuracy by reducing variance and error in high dimensional data sets and is capable of deciphering complex relationships among multiple features.<sup>38–41</sup> The learning process of RF can be interpreted where the importance of input variables and their interactions are visualized.<sup>42,43</sup> Therefore, RF has recently been applied in studies related to BrC.<sup>43–45</sup> For example, Hong et al.<sup>44</sup> employed RF algorithm to quantify the contribution of each source of water-soluble humic-like substances to BrC optical properties such as MAE and absorption Ångström exponent. Jiang et al.<sup>45</sup> used RF regression model to identify the molecular markers of BrC in the atmosphere from the FT-ICR-MS data and found that 17 nitrogen-containing species could well characterize the variation of BrC absorption. Given the intricate, nonlinear relationships within the multidimensional data set of BrC chromophores and optical properties, the interpretable RF regression model can be employed to facilitate the analysis of the linkages between chemical composition and optical properties of BrC.

In this study, we employed a gas chromatograph–mass spectrometer (GC-MS) and a UV–vis spectrophotometer to perform chemical analysis and spectroscopic measurement of PM<sub>2.5</sub> filter samples collected in Xi'an from 1 December 2018 to 29 January 2020. The mass concentrations and the light absorption of four typical groups of BrC chromophores, including 16 PAHs, 9 OPAHs, 11 NACs, and 2 MOPs, were characterized. Additionally, the optical properties (Abs, MAE) of methanol-soluble BrC were analyzed using an explainable RF model based on these quantified BrC chromophores. The data set collected in Xi'an in 2016 were also considered to ensure the performance of the RF model.<sup>4</sup> To further elucidate the relationships between the identified chromophores and the optical properties of BrC, the SHAP algorithm was employed to quantitatively assess the contributions of these chromophores to the prediction of BrC optical properties.

## MATERIAL AND METHODS

**Sample Collection.** Sampling was conducted on the rooftop of a three-story building (approximately 10 m above ground level) at urban site in Xi'an (34.23° N, 108.93° E). A total of 141 PM<sub>2.5</sub> samples (24 h integrated) were collected every 3 days on prebaked quartz-fiber filter (20.3 × 25.4 cm; Whatman, QM-A, Clifton, NJ, USA) from 1 December 2018 to 29 January 2020, using a high-volume sampler (Tisch, Cleveland, OH) with a flow rate of 1.05 m<sup>3</sup> min<sup>-1</sup>. The filter samples were sealed in aluminum foil bags, which had been baked in a muffle furnace at 450 °C for 3 h, and subsequently stored at -20 °C until analysis.

**Light Absorption Measurement.** A 0.526 cm<sup>2</sup> punch filter taken from each sample was sonicated in 10 mL methanol for 30 min (HPLC grade, J.T. Baker, Phillipsburg, NJ, USA). The methanol-soluble extracts were then filtered with polytetrafluoroethylene (PTFE) syringe filters (Whatman, 0.45 μm) to eliminate undissolved suspended substances. Following this, the filtered extracts were used to measure the light absorption spectra of methanol-soluble BrC, using an UV–vis spectrophotometer (300–700 nm) equipped with a liquid waveguide capillary cell (LWCC-3100, World Precision Instruments, Sarasota, FL, USA). This method, established by Hecobian et al.,<sup>46</sup> was also employed to measure the light

absorption of BrC chromophore standards at a specific concentration. The absorption coefficient at a given wavelength (Abs<sub>λ</sub>, Mm<sup>-1</sup> (10<sup>-6</sup> m<sup>-1</sup>)) was calculated from the following equation:

$$\text{Abs}_{\lambda} = (A_{\lambda} - A_{700}) \frac{V_1}{V_a \times L} \ln 10 \quad (1)$$

where  $A_{\lambda}$  is the measured absorption at a given wavelength  $\lambda$ .  $A_{700}$ , used as a reference for baseline drift, represents the absorption at 700 nm.  $V_1$  is the volume of solvent (methanol) used to extract the loaded sample.  $V_a$  corresponds to volume of the air sampled through the filter punch.  $L$  represents the optical path length (0.94 m), and  $\ln(10)$  is used to convert the Abs from log base 10 to log base  $e$ .

The mass absorption efficiency (MAE, m<sup>2</sup> gC<sup>-1</sup>) of BrC dissolved in methanol can be described as

$$\text{MAE}_{\lambda} = \frac{\text{Abs}_{\lambda}}{M} \quad (2)$$

where  $M$  is the mass concentration of methanol-soluble organic carbon (MSOC, μgC m<sup>-3</sup>). Due to the technical difficulty of direct measurement of MSOC concentration, OC was used to substitute MSOC because most of OC (~85%) can be extracted by methanol based on previous studies.<sup>47–49</sup> This process introduces uncertainty, but due to the difficulty in quantifying its impact, it was not considered in the subsequent evaluation of the impact of data uncertainty on the RF model.

The mass absorption efficiency (MAE, m<sup>2</sup> g<sup>-1</sup>) of BrC chromophore standards at a wavelength of  $\lambda$  can be calculated following the method described in Laskin et al.<sup>1</sup>

$$\text{MAE}_{\text{standards},\lambda} = \frac{A_{\lambda} - A_{700}}{L \times C} \ln(10) \quad (3)$$

where  $C$  (μg mL<sup>-1</sup>) is the concentration of the BrC chromophore standards.

The light absorption coefficient of BrC chromophore (Abs<sub>chromophore,λ</sub>, Mm<sup>-1</sup>) at a wavelength of  $\lambda$  can be obtained using eq 4:

$$\text{Abs}_{\text{chromophore},\lambda} = \text{MAE}_{\text{standards},\lambda} \times C_{\text{chromophore}} \quad (4)$$

where  $C_{\text{chromophore}}$  (μg m<sup>-3</sup>) is the atmospheric concentration of BrC chromophore. Additionally, detailed calculations of light absorption contributions from different groups of chromophores were provided in Text S1.

**Chemical Analysis.** A gas chromatograph–mass spectrometer (GC-MS; Agilent Technologies, Santa Clara, CA, USA) was employed to quantify the mass concentrations of BrC chromophores, including 16 PAHs, 9 OPAHs, 11 NACs, and 2 MOPs (See the full names and abbreviations of these compounds in Table S1). Details of the measurement are described in previous studies.<sup>4,50</sup> Briefly, a half of a 47 mm punch filter was subjected to ultrasonic extraction three times with 3 mL of a dichloromethane/methanol mixture (v/v = 2:1) for 15 min. The extracts were then purified by quartz wool packed in a Pasteur pipet and evaporated to approximately 1 mL using a rotary evaporator, before being dried with a gentle nitrogen stream. After the addition of 50 μL of *N,N*-bis(trimethylsilyl) trifluoroacetamide (BSTFA-TMCS; Fluka Analytical, 99%) and 10 μL of pyridine, the mixture were derivatized at 70 °C for 3 h for silylation. After the silylation derivatization, the mixture was adjusted to 200 μL with *n*-hexane. In our study, Chrysene-d12, Phenanthrene-d10, 4-

nitrophenol-2,3,5,6-d4 were used as internal standards to account for potential losses, thus enabling accurate quantification of the organic compounds. A procedural blank and a mixed standards with a concentration of 2 ppm were measured every 10 samples to check the interferences. Recoveries were calculated based on the signal ratios between the spiked blank filter and the standards, with average recoveries of all compounds ranging from 72 to 114%. The repeatability and reproducibility of the analysis were evaluated by repeating the measurements on a mixture of standards (2000 ng mL<sup>-1</sup>) spiked on prebaked quartz fiber filters and the relative standard deviations (RSDs) were <10% for measured organic compounds. The method detection limits were 0.04–0.1 ng m<sup>-3</sup> for nonpolar compounds and 0.06–0.2 ng m<sup>-3</sup> for polar compounds.

A punch (0.526 cm<sup>2</sup>) from each filter sample was used to measure organic carbon (OC) and elemental carbon (EC) using a thermal-optical analyzer (DRI, model 2001). The IMPROVE-A temperature protocol was employed for this purpose. Further details of this method can be found in the study by Chow et al.<sup>51</sup>

**Random Forest (RF) Model.** A random forest regression model was applied to investigate the relationships between the identified chromophores and the optical properties of BrC, while determining the importance of each chromophore measured in this study to the BrC optical properties. Here, an RF model was built for a BrC optical parameter (Abs or MAE) at a certain wavelength using the light absorption coefficients of the measured BrC chromophores. Nineteen wavelengths from 300 to 390 nm with a step size of 5 nm and 3 additional wavelengths (400, 425, and 450 nm) were selected to investigate variations of BrC optical properties from near-ultraviolet to visible region. To ensure the model learns from sufficient samples and address the imbalance between the large number of features and the relatively small sample size when using only one year of data, the data set collected in Xi'an in 2016 was included to ensure a more robust and reliable model.<sup>52</sup> With this addition, the total number of samples in the data set reached 253. To align with the observational results from 2016, only 33 identified chromophores were used as predictor features for model construction. The RF model was built using the “tidymodels” R package<sup>53</sup> with the “randomForest” set engine.<sup>38,54</sup> First, we adopted the stratified random sampling to ensure the balance of the data distribution of the training set and the test set, and to improve the stability and generalization ability of the model. Specifically, the entire data set was divided into five categories based on the four quantiles (0.2, 0.4, 0.6, 0.8) of the target variable (Abs or MAE). A random sample of 70% of each category was combined into a training set (175 data points) to construct the model. The remaining 30% was combined into a test set (78 data points), a separate set of values which was kept solely for evaluating model performance. To reduce the risk of model overfitting, make more efficient use of data, and select appropriate hyperparameters, 10-fold-cross-validation was conducted during the model construction process. The training set was further divided into ten subsets, in each iteration, nine subsets were used for training and one for testing. This process was repeated ten times, ensuring each subset served as the test set once. Then the overall process was repeated 10 times. Model performance in this process was evaluated using the coefficient of determination ( $R^2$ ) and the root-mean-square error (RMSE). Appropriate hyperparameters were selected based

on model performance in cross-validation, using grid search with varying numbers of decision trees (ntree) and varying numbers of predictor variables sampled at each split (mtry) (See Text S2 for details on grid Settings). The constructed model was validated on the test set using the correlation coefficient ( $r$ ), coefficient of determination ( $R^2$ ), variance explained ratios, root-mean-square error (RMSE), fraction of predictions within a factor of 2 (FAC2), mean gross error (MGE), normalized mean gross error (NMGE), coefficient of efficiency (COE) and index of agreement (IOA) (see Table S2). The calculation of these parameters can be found in the Text S2. In addition, the impact of potential sources of uncertainty and model sensitivity are detailed in the Text S3 and Text S4 and illustrated in Figures S1–S4 and Table S3.

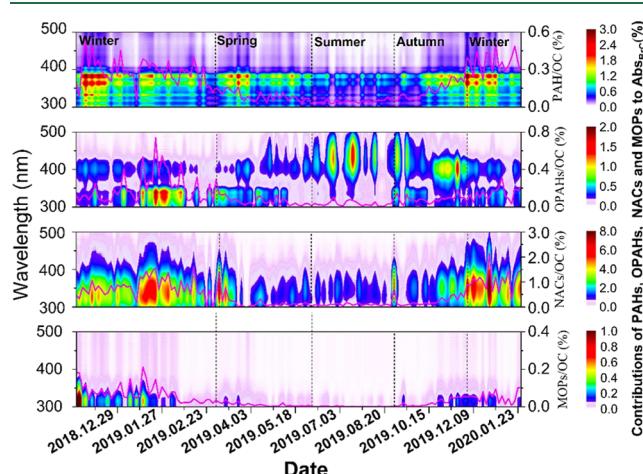
**Model Interpretation.** To quantify the contribution of the input BrC chromophores to a single prediction, the SHAP approach was applied to interpret the RF models using the “treeshap” R package.<sup>55</sup> SHAP is a game theoretic approach that enables the equitable allocation of “profits” (BrC optical properties:  $\text{Abs}_\lambda$  and  $\text{MAE}_\lambda$ ) among input features.<sup>56</sup> The functional expression of SHAP is as follows:

$$f(x_i) = \phi_0(f, x) + \sum_{i=1}^M \phi_j(f, x_i) \quad (5)$$

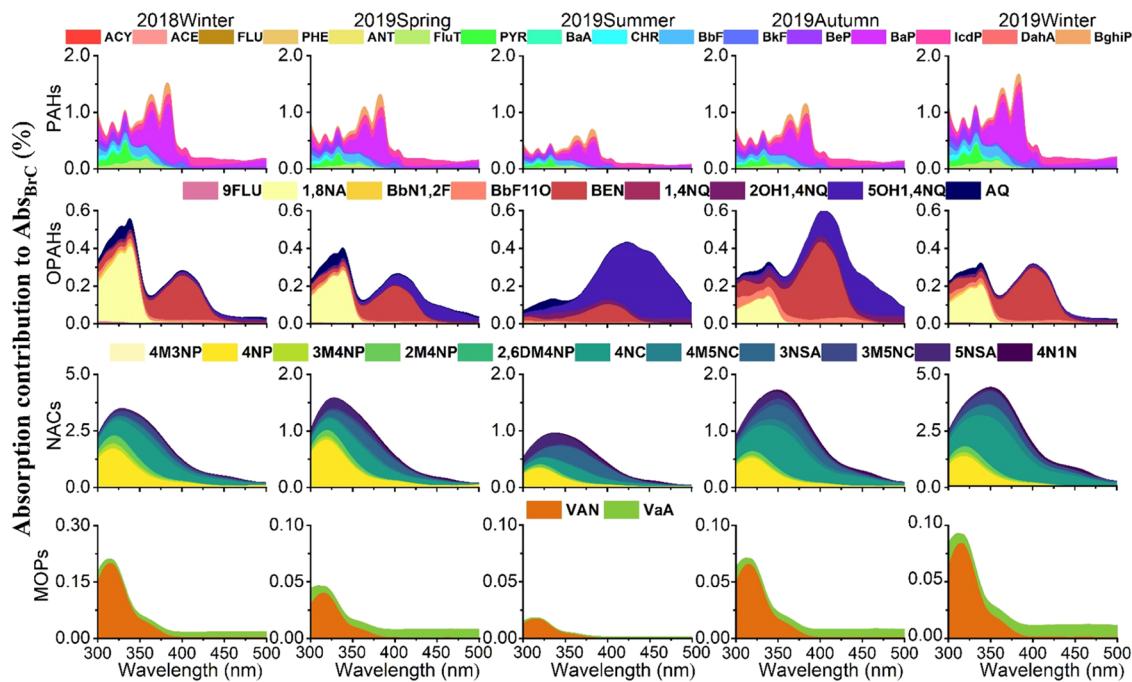
where  $f(x_i)$  is the predicted value generated for each sample ( $x_i$ ) with M features.  $\phi_0(f, x)$  is the base value representing the expected value of RF model output over the data set.  $\phi_j(f, x_i)$  is the SHAP value elucidating the impact of the feature  $j$  in the sample ( $x_i$ ) on the prediction of the sample. Details calculation of  $\phi_j(f, x_i)$  can be found in Lundberg et al.<sup>56</sup>

## RESULTS AND DISCUSSION

**Characterization of the Typical Chromophores.** The light absorption contributions of PAHs, OPAHs, NACs, MOPs to BrC show distinct wavelength dependence and seasonal variations as shown in Figures 1 and S5 (Details of the calculation of the light absorption contributions were shown in Text S1). PAHs present multiple absorption peaks at wavelengths <400 nm (see Figure 1), with absorption



**Figure 1.** Time series of the light absorption contributions of the total measured PAHs, OPAHs, NACs, MOPs to BrC over wavelengths from 300 to 500 nm (color scale) and the mass ratios of the total measured PAHs, OPAHs, NACs, and MOPs to OC are shown by the purple line on the right axis.

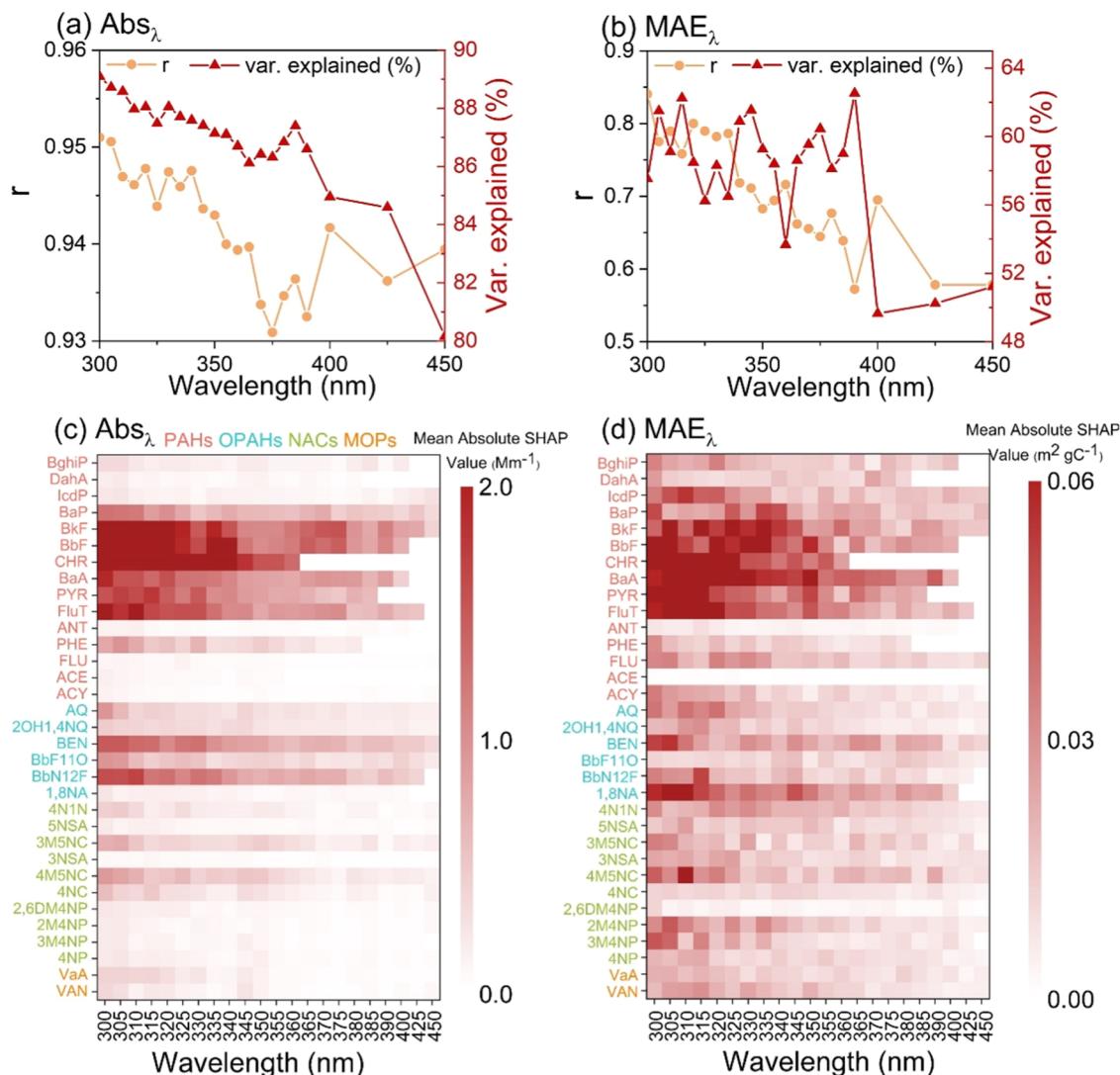


**Figure 2.** Light absorption contributions of distinct chromophores within the groups of PAHs, OPAHs, NACs, and MOPs to BrC over the wavelength range of 300–500 nm in winter 2018, spring 2019, summer 2019, autumn 2019, and winter 2019.

contributions reaching maximum at 365 and 380 nm. The light absorption contribution of PAHs to BrC at 365 nm peaks in winter 2019 ( $1.4 \pm 0.2\%$ ) and winter 2018 ( $1.3 \pm 0.2\%$ ), while reaching lowest contribution in summer 2019 ( $0.6 \pm 0.1\%$ ) (Figure 2). This is mainly attributed to variations in emission sources such as coal combustion and biomass burning and influences of meteorological conditions.<sup>24,57,58</sup> Thereinto, BaP, IcdP, BbF, and BghiP are the most prominent PAHs contributors in all seasons. Compared with results in Xi'an in 2016 where seasonal contributions of PAHs ranged from  $1 \pm 0.7\%$  to  $2.9 \pm 1.7\%$  at 365 nm,<sup>4</sup> the light absorption contributions of PAHs to BrC in this study are much lower. This decrease could be primarily related to the significant reduction in emissions from residential coal combustion and biomass burning, as a consequence of air pollution control measures.<sup>59</sup> Different from PAHs, OPAHs have multiple absorption peaks at wavelengths  $<350$  nm and wavelengths  $>400$  nm and the principal contributing factors of OPAHs vary across different seasons (see Figure 1). In winter 2018 and winter 2019, 1,8NA and BEN emerge as the main contributors to light absorption in OPAHs, whereas during summer 2019, SOH1, 4NQ stands out as the chromophore with the highest light absorption contribution (see Figure 2). This discrepancy suggests a seasonal difference in OPAH sources, with solid fuel combustion dominating during winter and secondary formation prevailing in summer.<sup>58,60,61</sup> NACs exhibit a single absorption peak and have the highest absorption contributions among the four groups of chromophores (see Figure 1). Similar to PAHs, NACs show the largest light contribution in winter 2019 ( $4.2 \pm 0.6\%$  at 365 nm), while obtaining lowest light contribution in summer 2019 ( $0.9 \pm 0.1\%$  at 365 nm). 4NC and 4NP are two predominant chromophores within the NACs across all seasons. However, during the spring and summer of 2019, the light absorption contributions of 3NSA and 5NSA present markedly higher levels compared to the other three seasons, likely due to the enhanced formation of

more oxidized components in spring and summer.<sup>22,34,62</sup> It should be noted that the light absorption contributions of NACs are much higher than the results in 2016 in Xi'an (ranging from  $0.1 \pm 0.1\%-0.9 \pm 0.3\%$  at 365 nm), likely due to the enhancement of water-soluble BrC following the implementation of air pollution control measures (Yuan et al., 2023). MOPs have one absorption peak at about 310 nm and present stronger absorption contributions in winter. As the chromophores with the lowest light absorption contribution among the four groups of chromophores, the light absorption contributions of MOPs are generally less than 0.2%. The variations in spectral characteristics and light absorption contributions illustrate the complex interplay among BrC molecular features, absorption properties, emission sources and secondary formation processes.

The mass contributions of PAHs, OPAHs, NACs, MOPs, and the total identified chromophores to OC exhibit noticeable fluctuations as the seasons changed (see Figures S6 and S7). PAHs, OPAHs, NACs, and MOPs all display highest contributions during the winter, while reaching lowest levels in summer. As shown in Tables S4 and S5, the average seasonal light absorption contributions of the chromophores are several times greater than their mass contributions to OC. For example, the average seasonal light absorption contributions of NACs at 365 nm are 3.5–9.4 times that of their mass contributions to OC. This suggests that even small amounts of chromophores could exert a substantial impact on the light absorption characteristics of BrC.<sup>4,24,63</sup> In general, the light absorption peaks of the four groups of BrC chromophores, influenced by emission sources and atmospheric formation processes, demonstrate distinct characteristics and substantial seasonal variations. The highest average mass contribution of the total identified BrC chromophores to OC (in winter 2019) is  $1.7 \pm 0.5\%$ , while the highest average light absorption contribution of the total identified BrC chromophores to BrC is approximately  $5.8 \pm 2.6\%$  (in winter 2019 at 365 nm),



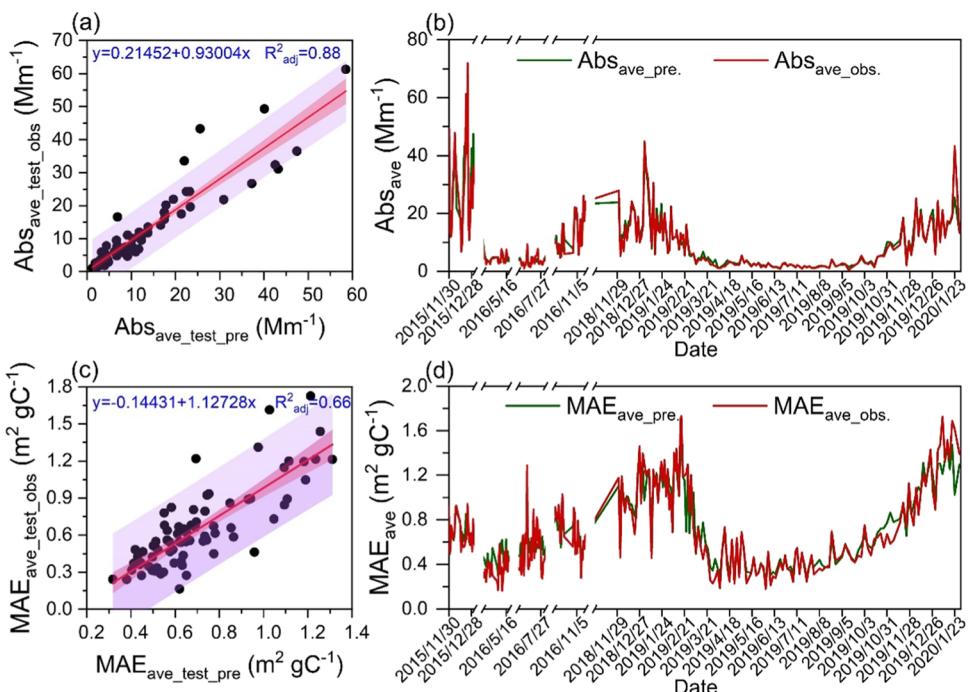
**Figure 3.** Correlation coefficients ( $r$ ) and variance explained ratios of the random forest model for (a)  $\text{Abs}_\lambda$  and (b)  $\text{MAE}_\lambda$  of BrC. The mean absolute SHAP values of BrC chromophores for the random forest built for (c)  $\text{Abs}_\lambda$ , and (d)  $\text{MAE}_\lambda$  of BrC over wavelengths from 300 to 500 nm.

indicating that the identified BrC chromophores are still limited for explaining the overall BrC optical properties. To address this limitation, data-driven interpretable machine learning can be employed to explore variations in BrC optical properties based on the limited information on the identified chromophores and gain a deeper understanding of the relationships between the chemical composition and optical properties of BrC.

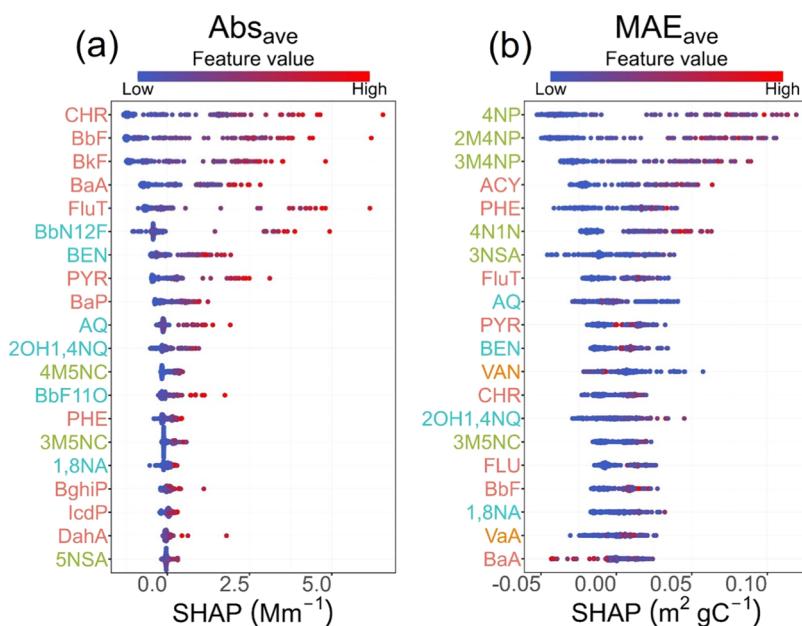
**Predictions of BrC Optical Properties Using the RF Model.** To investigate the relationships between the chemical composition and optical properties of BrC, we built a decision tree model-based RF using limited chromophore information to reproduce temporal variations of  $\text{Abs}_\lambda$  and  $\text{MAE}_\lambda$  of BrC across wavelengths ranging from 300 to 500 nm. The RF model achieved high accuracy on the test data with  $r > 0.93$  for  $\text{Abs}_\lambda$  and  $r > 0.57$  for  $\text{MAE}_\lambda$ . The variance explained ratios exceed 80% for  $\text{Abs}_\lambda$  and exceed 49% for  $\text{MAE}_\lambda$  (see Figure 3a,b). Additionally, the model effectively captures the daily variations in observed  $\text{Abs}_\lambda$  and  $\text{MAE}_\lambda$ , indicating the feasibility of understanding BrC optical properties with limited chromophore information (see Figures S8 and S9).

The average importance of chromophores is elucidated through the mean absolute SHAP values (see Figure 3c,d). PAHs with four or five rings, such as CHR and BbF, as well as OPAHs with four rings, such as BEN and Bbn12F, present strong influences on  $\text{Abs}_\lambda$  of BrC. NACs, except for 3MSNC and 4MSNC, exhibit relatively minor impacts on  $\text{Abs}_\lambda$ . As the least contributing factors, MOPs have a slightly greater impact in the low wavelength range (<365 nm) compared to the high wavelength range. Different from the results of  $\text{Abs}_\lambda$ , these four groups of chromophores all have certain contributions to  $\text{MAE}_\lambda$  and the dominant contributors vary at different wavelengths. At wavelengths less than 400 nm, significant contributions are made by CHR and BaA in PAHs, and by 1,8-NA and BEN in OPAHs. However, in the visible region, IcdP and BkF from PAHs, along with BEN from OPAHs, are prominent contributors.

It is worth noting that, in the discussion of the observed results, NACs contribute the most to BrC optical absorption among the identified chromophores, while four and five-ring PAHs and OPAHs contribute prominently in predictions of BrC optical properties according to the SHAP values. This difference arises from the distinct meanings they represent:



**Figure 4.** (a) Random forest performance for test data set of the average Abs of BrC, (b) time series of the predicted average Abs ( $\text{Abs}_{\text{ave\_pre.}}$ ) and the observed average Abs ( $\text{Abs}_{\text{ave\_obs.}}$ ), (c) Random forest performance for test data set of the average MAE of BrC, (d) time series of the predicted average MAE ( $\text{MAE}_{\text{ave\_pre.}}$ ) and the observed average MAE ( $\text{MAE}_{\text{ave\_obs.}}$ ).



**Figure 5.** SHAP values of BrC chromophores toward individual prediction in the random forest built for (a)  $\text{Abs}_{\text{ave}}$  and (b)  $\text{MAE}_{\text{ave}}$ . (Feature value represents the mass concentrations of chromophores. Red represents high chromophore concentration and blue represents low chromophore concentration. The color of the Y-axis label represents the group to which the chromophore belongs).

The contribution of NACs in the observed results refers to the actual optical impact of the identified chromophores, reflecting the direct role of these specific compounds in BrC. In contrast, SHAP enables the equitable allocation of “profits” (BrC optical properties:  $\text{Abs}_\lambda$  and  $\text{MAE}_\lambda$ ) among input features,<sup>56</sup> reflecting the impacts of input features on individual predictions relative to the base value (the mean of the model’s predicted values). The significant contributions of these four and five-ring PAHs and OPAHs, according to SHAP, not only underscore their

notable roles in elucidating BrC optical properties but also imply contributions from potential chromophores with similar structures or sources that are not identified here. For example, Kuang et al.<sup>29</sup> identified 11 new BrC species (mainly OPAHs and PAHs) in atmosphere of Beijing. Thereinto, two of these chromophores (naphtho[1,2-*c*]furan-1,3-dione and benzo[*a*]-fluoren-11-one) are isomers of the chromophores identified in our study (i.e., 1,8-NA and BbF11O). Although it is a challenge to identify and quantify all BrC chromophores,

machine learning provides a promising approach for more precise identification of unknown chromophores by analyzing the key chemical features or structural characteristics that influence BrC optical properties.

As depicted in Figure S10, the chromophores positively influence  $\text{Abs}_{\lambda}$  levels and they all present increasing contributions with higher predicted values. But for  $\text{MAE}_{\lambda}$ , the contributions of the chromophores exhibit larger variations (see Figure S11). This is mainly because the MAE is connected with chemical composition of organic aerosol which is complex and influenced by a multitude of factors, including the sources of organic aerosol, atmospheric aging processes and aerosol acidity.<sup>4,8,46,64,65</sup>

**Simplifying Predictions of BrC Optical Properties Utilizing Mass Concentrations of Chromophores.** To streamline estimation of BrC optical properties, we utilized the mass concentrations of BrC chromophores as predictors to elucidate the variations of BrC optical properties. Considering the difficulty in explaining the wavelength dependence of BrC optical properties by the mass concentrations of BrC chromophores, the average BrC light absorption coefficient ( $\text{Abs}_{\text{ave}}$ ) and the average mass absorption efficiency ( $\text{MAE}_{\text{ave}}$ ) were calculated as observed results to build the model. The RF model achieved high accuracy on the test data, with  $R^2$  and RMSE values of 0.88 and 4.28 for  $\text{Abs}_{\text{ave}}$  and 0.7 and 0.19 for  $\text{MAE}_{\text{ave}}$  respectively (see Figure 4 and Table S2). Temporal variations of  $\text{Abs}_{\text{ave}}$  and  $\text{MAE}_{\text{ave}}$  are well reproduced by the data-driven model. Furthermore, the model explains over 85% of the variance of  $\text{Abs}_{\text{ave}}$  and over 65% of the variance of  $\text{MAE}_{\text{ave}}$ , confirming its robust performance (see Table S2).

As shown in Figure 5a, all measured chromophores positively contribute to  $\text{Abs}_{\text{ave}}$  levels. Notably, CHR, BkF, and BbF stand out as the most influential predictors of  $\text{Abs}_{\text{ave}}$  variation, aligning consistently with the distribution of SHAP values obtained using  $\text{Abs}_{\text{chromophore},\lambda}$  as predictors. For  $\text{MAE}_{\text{ave}}$ , however, the influences of these predictors show differences (see Figure 5b). The prominent chromophores are NACs (4NP, 2M4NP, and 3M4NP), which exert positive effects on  $\text{MAE}_{\text{ave}}$ . This further indicates that, in contrast to the Abs of BrC, the MAE of BrC is more intricate and influenced by a greater number of factors. Therefore, additional research is essential to explore the driving factors of MAE. Nevertheless, these results still demonstrate that the optical properties of BrC can be well predicted by utilizing the mass concentrations of the identified BrC chromophores. This, in turn, aids in quantifying the contribution of BrC to radiative forcing and helps develop more accurate climate models.

## ■ IMPLICATIONS AND LIMITATIONS

BrC plays a crucial role in atmospheric radiation and global climate owing to its notable light absorption properties in the near-ultraviolet and visible regions.<sup>1,3</sup> Molecular characterization of BrC serves as a vital approach for probing its chemical composition and optical properties. However, the explanation of the BrC optical properties remains greatly constrained by the current knowledge of identified chromophores.<sup>4,24,28,37</sup> In this study, using an interpretable machine learning model, we achieved good predictions of  $\text{Abs}_{\lambda}$  and  $\text{MAE}_{\lambda}$  for BrC based on the identified 33 BrC chromophores belonging to four groups. The model explains more than 80% of the variance in Abs and over 50% of the variance in MAE within the 300–450 nm range, significantly improving our understanding of BrC light absorption characteristics com-

pared to traditional methods. In addition, by using the SHAP algorithm, we quantitatively evaluated the influences of these chromophores on the optical properties of BrC. It is noted that all chromophores contribute positively to  $\text{Abs}_{\lambda}$  of BrC, with the four-ring and five-ring PAHs and OPAHs exhibiting particularly prominent contributions. This indicates the possible existence of unidentified chromophores with structural characteristics resembling PAHs and OPAHs, providing guidance for inferring unknown species of BrC. Furthermore, estimating  $\text{Abs}_{\text{ave}}$  and  $\text{MAE}_{\text{ave}}$  based on the mass concentrations of the identified chromophores enables us to investigate BrC optical properties without requiring detailed optical information for each individual chromophore. This approach further simplifies the study of BrC optical properties and enhances the assessment of the radiative forcing attributed to BrC. However, in this study, the BrC chromophores used as feature variables are limited, preventing a full exploration of the impact of chromophore structure on BrC optical properties. Additionally, with the current information on identified BrC chromophores, directly determining the molecular structures of unknown BrC chromophores remains a challenge. Future studies could employ high resolution mass spectrometry combined with spectral analysis, such as high-performance liquid chromatography coupled with a photodiode array and high-resolution orbitrap mass spectrometry, along with non-targeted analysis. This approach would help identify a broader range of chromophore types for machine learning model development, providing a more robust foundation for understanding the optical properties of BrC and offering more comprehensive guidance for identifying unknown chromophores.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.4c09031>.

Additional texts, figures and tables about model hyperparameters optimization, model uncertainty, model sensibility, abbreviations of organics, mass contributions of chromophores, light absorption contributions of chromophores, model performance, and SHAP values for the analysis of the importance of BrC chromophores (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Ru-Jin Huang – State Key Laboratory of Loess Science, Center for Excellence in Quaternary Science and Global Change, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China; Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an 710049, China; University of Chinese Academy of Sciences, Beijing 100049, China; [orcid.org/0000-0002-4907-9616](https://orcid.org/0000-0002-4907-9616); Email: [ruijin.huang@ieecas.cn](mailto:ruijin.huang@ieecas.cn)

Wei Xu – Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China; [orcid.org/0000-0002-9590-1906](https://orcid.org/0000-0002-9590-1906); Email: [wxu@iue.ac.cn](mailto:wxu@iue.ac.cn)

### Authors

Ying Wang – Interdisciplinary Research Center of Earth Science Frontier, State Key Laboratory of Earth Surface

Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; State Key Laboratory of Loess Science, Center for Excellence in Quaternary Science and Global Change, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

**Haobin Zhong** – School of Advanced Materials Engineering, Jiaxing Nanhua University, Jiaxing 314001, China

**Ting Wang** – State Key Laboratory of Loess Science, Center for Excellence in Quaternary Science and Global Change, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

**Lu Yang** – State Key Laboratory of Loess Science, Center for Excellence in Quaternary Science and Global Change, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China; University of Chinese Academy of Sciences, Beijing 100049, China

**Wei Yuan** – State Key Laboratory of Loess Science, Center for Excellence in Quaternary Science and Global Change, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

**Zhisheng An** – Interdisciplinary Research Center of Earth Science Frontier, State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China; State Key Laboratory of Loess Science, Center for Excellence in Quaternary Science and Global Change, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.est.4c09031>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 41925015, 42430708, the Strategic Priority Research Program of Chinese Academy of Sciences (XDB40000000), the Key Research Program of Frontier Sciences from the Chinese Academy of Sciences (ZDBS-LY-DQC001), the State Key Laboratory of Loess Science (SKLLQGPY2411), the New Cornerstone Science Foundation through the XPLORER PRIZE, and Natural Science Basic Research Program of Shaanxi (No. S2022-JC-QN-2049).

## REFERENCES

- (1) Laskin, A.; Laskin, J.; Nizkorodov, S. A. Chemistry of atmospheric brown carbon. *Chem. Rev.* **2015**, *115* (10), 4335–4382.
- (2) Zhang, Y.; Forrister, H.; Liu, J.; Dibb, J.; Anderson, B.; Schwarz, J. P.; Perring, A. E.; Jimenez, J. L.; Campuzano-Jost, P.; Wang, Y.; Nenes, A.; Weber, R. J. Top-of-atmosphere radiative forcing affected by brown carbon in the upper troposphere. *Nat. Geosci.* **2017**, *10* (7), 486–489.
- (3) Zeng, L.; Zhang, A.; Wang, Y.; Wagner, N. L.; Katich, J. M.; Schwarz, J. P.; Schill, G. P.; Brock, C.; Froyd, K. D.; Murphy, D. M.; Williamson, C. J.; Kucic, A.; Scheuer, E.; Dibb, J.; Weber, R. J. Global measurements of brown carbon and estimated direct radiative effects. *Geophys. Res. Lett.* **2020**, *47* (13), No. e2020GL088747.
- (4) Yuan, W.; Huang, R.-J.; Yang, L.; Guo, J.; Chen, Z.; Duan, J.; Wang, T.; Ni, H.; Han, Y.; Li, Y.; Chen, Q.; Chen, Y.; Hoffmann, T.; O'Dowd, C. Characterization of the light-absorbing properties, chromophore composition and sources of brown carbon aerosol in

Xi'an, northwestern China. *Atmos. Chem. Phys.* **2020**, *20* (8), 5129–5144.

(5) Mok, J.; Krotkov, N. A.; Arola, A.; Torres, O.; Jethva, H.; Andrade, M.; Labow, G.; Eck, T. F.; Li, Z.; Dickerson, R. R.; Stenchikov, G. L.; Osipov, S.; Ren, X. Impacts of brown carbon from biomass burning on surface UV and ozone photochemistry in the Amazon Basin. *Sci. Rep.* **2016**, *6*, No. 36940.

(6) Yang, L.; Huang, R.; Yuan, W.; Huang, D.; Huang, C. pH-Dependent aqueous-phase brown carbon formation: Rate constants and implications for solar absorption and atmospheric photochemistry. *Environ. Sci. Technol.* **2024**, *58*, 1236–1243.

(7) Lin, P.; Bluvshtein, N.; Rudich, Y.; Nizkorodov, S. A.; Laskin, J.; Laskin, A. Molecular chemistry of atmospheric brown carbon inferred from a nationwide biomass burning event. *Environ. Sci. Technol.* **2017**, *51* (20), 11561–11570.

(8) Li, D.; Wu, C.; Zhang, S.; Lei, Y.; Lv, S.; Du, W.; Liu, S.; Zhang, F.; Liu, X.; Liu, L.; Meng, J.; Wang, Y.; Gao, J.; Wang, G. Significant coal combustion contribution to water-soluble brown carbon during winter in Xingtai, China: Optical properties and sources. *J. Environ. Sci.* **2023**, *124*, 892–900.

(9) Huang, R.-J.; Yang, L.; Shen, J.; Yuan, W.; Gong, Y.; Ni, H.; Duan, J.; Yan, J.; Huang, H.; You, Q.; Li, Y. J. Chromophoric fingerprinting of brown carbon from residential biomass burning. *Environ. Sci. Technol. Lett.* **2022**, *9* (2), 102–111.

(10) Huang, R.-J.; Yuan, W.; Yang, L.; Yang, H.; Cao, W.; Guo, J.; Zhang, N.; Zhu, C.; Wu, Y.; Zhang, R. Concentration, optical characteristics, and emission factors of brown carbon emitted by on-road vehicles. *Sci. Total Environ.* **2022**, *810*, No. 151307.

(11) Yuan, W.; Huang, R.-J.; Shen, J.; Wang, K.; Yang, L.; Wang, T.; Gong, Y.; Cao, W.; Guo, J.; Ni, H.; Duan, J.; Hoffmann, T. More water-soluble brown carbon after the residential “coal-to-gas” conversion measure in urban Beijing. *npj Clim. Atmos. Sci.* **2023**, *6* (1), No. 20.

(12) He, Q.; Li, C.; Siemens, K.; Morales, A. C.; Hettiyadura, A. P. S.; Laskin, A.; Rudich, Y. Optical properties of secondary organic aerosol produced by photooxidation of naphthalene under NO<sub>x</sub> condition. *Environ. Sci. Technol.* **2022**, *56* (8), 4816–4827.

(13) Klodt, A. L.; Aiona, P. K.; MacMillan, A. C.; Ji Lee, H.; Zhang, X.; Helgestad, T.; Novak, G. A.; Lin, P.; Laskin, J.; Laskin, A.; Bertram, T. H.; Cappa, C. D.; Nizkorodov, S. A. Effect of relative humidity, NO<sub>x</sub>, and ammonia on the physical properties of naphthalene secondary organic aerosols. *Environ. Sci. Atmos.* **2023**, *3* (6), 991–1007.

(14) Zhao, R.; Lee, A. K. Y.; Huang, L.; Li, X.; Yang, F.; Abbatt, J. P. D. Photochemical processing of aqueous atmospheric brown carbon. *Atmos. Chem. Phys.* **2015**, *15* (11), 6087–6100.

(15) Yang, L.; Huang, R.-J.; Shen, J.; Wang, T.; Gong, Y.; Yuan, W.; Liu, Y.; Huang, H.; You, Q.; Huang, D. D.; Huang, C. New insights into the brown carbon chromophores and formation pathways for aqueous reactions of  $\alpha$ -Dicarbonyls with amines and ammonium. *Environ. Sci. Technol.* **2023**, *57* (33), 12351–12361.

(16) Ni, H.; Huang, R.-J.; Pieber, S. M.; Corbin, J. C.; Stefanelli, G.; Pospisilova, V.; Klein, F.; Gysel-Beer, M.; Yang, L.; Baltensperger, U.; Haddad, I. E.; Slowik, J. G.; Cao, J.; Prévôt, A. S. H.; Dusek, U. Brown carbon in primary and aged coal combustion emission. *Environ. Sci. Technol.* **2021**, *55* (9), 5701–5710.

(17) Li, C.; He, Q.; Fang, Z.; Brown, S. S.; Laskin, A.; Cohen, S. R.; Rudich, Y. Laboratory insights into the diel cycle of optical and chemical transformations of biomass burning brown carbon aerosols. *Environ. Sci. Technol.* **2020**, *54* (19), 11827–11837.

(18) Hems, R. F.; Schnitzler, E. G.; Liu-Kang, C.; Cappa, C. D.; Abbatt, J. P. D. Aging of atmospheric brown carbon aerosol. *ACS Earth Space Chem.* **2021**, *5* (4), 722–748.

(19) Zhou, Y.; West, C. P.; Hettiyadura, A. P. S.; Pu, W.; Shi, T.; Niu, X.; Wen, H.; Cui, J.; Wang, X.; Laskin, A. Molecular characterization of water-soluble brown carbon chromophores in snowpack from Northern Xinjiang, China. *Environ. Sci. Technol.* **2022**, *56* (7), 4173–4186.

- (20) Xu, J.; Hettiyadura, A. P. S.; Liu, Y.; Zhang, X.; Kang, S.; Laskin, A. Atmospheric brown carbon on the Tibetan Plateau: Regional differences in chemical composition and light absorption properties. *Environ. Sci. Technol. Lett.* **2022**, *9* (3), 219–225.
- (21) Siemens, K.; Morales, A.; He, Q.; Li, C.; Hettiyadura, A. P. S.; Rudich, Y.; Laskin, A. Molecular analysis of secondary brown carbon produced from the photooxidation of naphthalene. *Environ. Sci. Technol.* **2022**, *56* (6), 3340–3353.
- (22) Yuan, W.; Huang, R.-J.; Yang, L.; Wang, T.; Duan, J.; Guo, J.; Ni, H.; Chen, Y.; Chen, Q.; Li, Y.; Dusek, U.; O'Dowd, C.; Hoffmann, T. Measurement report: PM<sub>2.5</sub>-bound nitrated aromatic compounds in Xi'an, Northwest China—seasonal variations and contributions to optical properties of brown carbon. *Atmos. Chem. Phys.* **2021**, *21* (5), 3685–3697.
- (23) Yuan, W.; Huang, R.-J.; Yang, L.; Ni, H.; Wang, T.; Cao, W.; Duan, J.; Guo, J.; Huang, H.; Hoffmann, T. Concentrations, optical properties and sources of humic-like substances (HULIS) in fine particulate matter in Xi'an, Northwest China. *Sci. Total Environ.* **2021**, *789*, No. 147902.
- (24) Huang, R.-J.; Yang, L.; Cao, J.; Chen, Y.; Chen, Q.; Li, Y.; Duan, J.; Zhu, C.; Dai, W.; Wang, K.; Lin, C.; Ni, H.; Corbin, J. C.; Wu, Y.; Zhang, R.; Tie, X.; Hoffmann, T.; O'Dowd, C.; Dusek, U. Brown carbon aerosol in urban Xi'an, Northwest China: The composition and light absorption properties. *Environ. Sci. Technol.* **2018**, *52* (12), 6825–6833.
- (25) Huang, R.-J.; Yang, L.; Shen, J.; Yuan, W.; Gong, Y.; Guo, J.; Cao, W.; Duan, J.; Ni, H.; Zhu, C.; Dai, W.; Li, Y.; Chen, Y.; Chen, Q.; Wu, Y.; Zhang, R.; Dusek, U.; O'Dowd, C.; Hoffmann, T. Water-insoluble organics dominate brown carbon in wintertime urban aerosol of China: Chemical characteristics and optical properties. *Environ. Sci. Technol.* **2020**, *54* (13), 7836–7847.
- (26) Lin, P.; Fleming, L. T.; Nizkorodov, S. A.; Laskin, J.; Laskin, A. Comprehensive molecular characterization of atmospheric brown carbon by high resolution mass spectrometry with electrospray and atmospheric pressure photoionization. *Anal. Chem.* **2018**, *90* (21), 12493–12502.
- (27) West, C. P.; Hettiyadura, A. P. S.; Darmody, A.; Mahamuni, G.; Davis, J.; Novoselov, I.; Laskin, A. Molecular composition and the optical properties of brown carbon generated by the ethane flame. *ACS Earth Space Chem.* **2020**, *4* (7), 1090–1103.
- (28) Yan, C.; Zheng, M.; Desyaterik, Y.; Sullivan, A. P.; Wu, Y.; Collett, J. L. Molecular characterization of water-soluble brown carbon chromophores in Beijing, China. *J. Geophys. Res.: Atmos.* **2020**, *125* (15), No. e2019JD032018.
- (29) Kuang, Y.; Shang, J.; Sheng, M.; Shi, X.; Zhu, J.; Qiu, X. Molecular composition of Beijing PM<sub>2.5</sub> brown carbon revealed by an untargeted approach based on gas chromatography and time-of-flight mass spectrometry. *Environ. Sci. Technol.* **2023**, *57* (2), 909–919.
- (30) Zhang, R.; Li, S.; Fu, X.; Pei, C.; Wang, J.; Wu, Z.; Xiao, S.; Huang, X.; Zeng, J.; Song, W.; Zhang, Y.; Bi, X.; Wang, X. Emissions and light absorption of PM<sub>2.5</sub>-bound nitrated aromatic compounds from on-road vehicle fleets. *Environ. Pollut.* **2022**, *312*, No. 120070.
- (31) Li, X.; Zhao, Q.; Yang, Y.; Zhao, Z.; Liu, Z.; Wen, T.; Hu, B.; Wang, Y.; Wang, L.; Wang, G. Composition and sources of brown carbon aerosols in megacity Beijing during the winter of 2016. *Atmos. Res.* **2021**, *262*, No. 105773.
- (32) Cai, D.; Li, C.; Lin, J.; Sun, W.; Zhang, M.; Wang, T.; Abudumutailifu, M.; Lyu, Y.; Huang, X.; Li, X.; Chen, J. Comparative study of atmospheric brown carbon at Shanghai and the East China Sea: Molecular characterization and optical properties. *Sci. Total Environ.* **2024**, *941*, No. 173782.
- (33) Huang, S.; Yang, X.; Xu, H.; Zeng, Y.; Li, D.; Sun, J.; Ho, S. S. H.; Zhang, Y.; Cao, J.; Shen, Z. Insights into the nitroaromatic compounds, formation, and light absorption contributing emissions from various geological maturity coals. *Sci. Total Environ.* **2023**, *870*, No. 162033.
- (34) Cao, M.; Yu, W.; Chen, M.; Chen, M. Characterization of nitrated aromatic compounds in fine particles from Nanjing, China: Optical properties, source allocation, and secondary processes. *Environ. Pollut.* **2023**, *316*, No. 120650.
- (35) Gong, Y.; Huang, R.-J.; Yang, L.; Wang, T.; Yuan, W.; Xu, W.; Cao, W.; Wang, Y.; Li, Y. Measurement report: Brown carbon aerosol in polluted urban air of the North China Plain-day-night differences in the chromophores and optical properties. *Atmos. Chem. Phys.* **2023**, *23* (24), 15197–15207.
- (36) Lei, Y.; Zhang, K.; Lu, Y.; Qin, Y.; Li, L.; Li, J.; Liu, X.; Wu, C.; Zhang, S.; Chen, Y.; Zhang, J.; Zhang, F.; Wang, G. Characterization of water-soluble brown carbon in atmospheric fine particles over Xi'an, China: Implication of aqueous brown carbon formation from biomass burning. *Sci. Total Environ.* **2023**, *881*, No. 163442.
- (37) Li, X.; Hu, M.; Wang, Y.; Xu, N.; Fan, H.; Zong, T.; Wu, Z.; Guo, S.; Zhu, W.; Chen, S.; Dong, H.; Zeng, L.; Yu, X.; Tang, X. Links between the optical properties and chemical compositions of brown carbon chromophores in different environments: Contributions and formation of functionalized aromatic compounds. *Sci. Total Environ.* **2021**, *786*, No. 147418.
- (38) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
- (39) Grange, S. K.; Carslaw, D. C.; Lewis, A. C.; Boleti, E.; Hueglin, C. Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmos. Chem. Phys.* **2018**, *18* (9), 6223–6239.
- (40) Vu, T. V.; Shi, Z.; Cheng, J.; Zhang, Q.; He, K.; Wang, S.; Harrison, R. M. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. *Atmos. Chem. Phys.* **2019**, *19* (17), 11303–11314.
- (41) Shi, Z.; Song, C.; Liu, b.; Lu, G.; Xu, J.; Vu, T. V.; Elliott, R. J. R.; Li, W.; Bloss, W. J.; Harrison, R. M. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Sci. Adv.* **2021**, *7*, No. eabd6696.
- (42) Men, Y.; Li, Y.; Luo, Z.; Jiang, K.; Yi, F.; Liu, X.; Xing, R.; Cheng, H.; Shen, G.; Tao, S. Interpreting highly variable indoor PM<sub>2.5</sub> in rural North China using machine learning. *Environ. Sci. Technol.* **2023**, *57*, 18183–18192.
- (43) Guo, Z.; Sun, W.; Hu, X.; Lin, J.; Fu, Y.; Peng, X.; Jiang, B.; Liao, Y.; Zhang, G.; Wang, X.; Peng, P.; Bi, X. Molecular characteristics and compositions affecting light absorption features of cloud water revealed by Fourier transform ion cyclotron resonance mass spectrometry. *Atmos. Environ.* **2023**, *295*, No. 119565.
- (44) Hong, Y.; Cao, F.; Fan, M. Y.; Lin, Y. C.; Bao, M. Y.; Xue, Y. W.; Wu, J. Y.; Yu, M. Y.; Wu, X.; Zhang, Y. L. Using machine learning to quantify sources of light-absorbing water-soluble humic-like substances (HULISws) in Northeast China. *Atmos. Environ.* **2022**, *291*, No. 119371.
- (45) Jiang, H.; Li, J.; Sun, R.; Tian, C.; Tang, J.; Jiang, B.; Liao, Y.; Chen, C. E.; Zhang, G. Molecular dynamics and light absorption properties of atmospheric dissolved organic matter. *Environ. Sci. Technol.* **2021**, *55* (15), 10268–10279.
- (46) Hecobian, A.; Zhang, X.; Zheng, M.; Frank, N.; Edgerton, E. S.; Weber, R. J. Water-soluble organic aerosol material and the light-absorption characteristics of aqueous extracts measured over the Southeastern United States. *Atmos. Chem. Phys.* **2010**, *10*, 5965–5977.
- (47) Chen, Y.; Bond, T. C. Light absorption by organic carbon from wood combustion. *Atmos. Chem. Phys.* **2010**, *10*, 1773–1787.
- (48) Cheng, Y.; He, K. B.; Du, Z. Y.; Engling, G.; Liu, J.; Ma, Y. L.; Zheng, M.; Weber, R. J. The characteristics of brown carbon aerosol during winter in Beijing. *Atmos. Environ.* **2016**, *127*, 355–364.
- (49) Xie, M.; Chen, X.; Holder, A. L.; Hays, M. D.; Lewandowski, M.; Offenberg, J. H.; Kleindienst, T. E.; Jaoui, M.; Hannigan, M. P. Light absorption of organic carbon and its sources at a southeastern U.S. location in summer. *Environ. Pollut.* **2019**, *244*, 38–46.
- (50) Wang, T.; Huang, R. J.; Li, Y.; Chen, Q.; Chen, Y.; Yang, L.; Guo, J.; Ni, H.; Hoffmann, T.; Wang, X.; Mai, B. One-year characterization of organic aerosol markers in urban Beijing: Seasonal variation and spatiotemporal comparison. *Sci. Total Environ.* **2020**, *743*, No. 140689.
- (51) Chow, J. C.; Watson, J. G.; Robles, J.; Wang, X.; Chen, L. W.; Trimble, D. L.; Kohl, S. D.; Tropp, R. J.; Fung, K. K. Quality assurance and quality control for thermal/optical analysis of aerosol

samples for organic and elemental carbon. *Anal. Bioanal. Chem.* **2011**, *401* (10), 3141–3152.

(52) Zhu, J. J.; Yang, M.; Ren, Z. J. Machine learning in environmental research: Common pitfalls and best practices. *Environ. Sci. Technol.* **2023**, *57* (46), 17671–17689.

(53) Kuhn, M.; Wickham, H. Easily Install and Load the 'Tidymodels' Packages. 2023, Available online: <https://github.com/tidymodels/tidymodels> (accessed on 29 October 2024).

(54) Breiman, L.; Cutler, A.; Liaw, A.; Wiener, M. Breiman and Cutler's random forests for classification and regression. 2022, Available online: <https://www.stat.berkeley.edu/~breiman/RandomForests/> (accessed on 29 October 2024).

(55) Komisarczyk, K.; Kozminski, P.; Maksymiuk, S.; Kapsner, L. A.; Spytek, M.; Krzyszynski, M.; Biecek, P. treeshap: Compute SHAP values for your tree-based models using the 'TreeSHAP' algorithm. 2024, Available online: <https://modeloriented.github.io/treeshap/> (accessed on 29 October 2024).

(56) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S. I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2* (1), 56–67.

(57) Zhu, C. S.; Li, L. J.; Huang, H.; Dai, W. T.; Lei, Y. L.; Qu, Y.; Huang, R. J.; Wang, Q. Y.; Shen, Z. X.; Cao, J. J. n-Alkanes and PAHs in the Southeastern Tibetan Plateau: Characteristics and correlations with brown carbon light absorption. *J. Geophys. Res.: Atmos.* **2020**, *125* (19), No. e2020JD032666.

(58) Bandowe, B. A. M.; Meusel, H.; Huang, R.-j.; Ho, K.; Cao, J.; Hoffmann, T.; Wilcke, W. PM2.5-bound oxygenated PAHs, nitro-PAHs and parent-PAHs from the atmosphere of a Chinese megacity: Seasonal variation, sources and cancer risk assessment. *Sci. Total Environ.* **2014**, *473*–474, 77–87.

(59) Xi'an Municipal People's Government. Available online: <https://www.xa.gov.cn/> (accessed on 29 October 2024).

(60) Chen, L.; Liu, W.; Tao, S.; Liu, W. Spatiotemporal variations and source identification of atmospheric nitrated and oxygenated polycyclic aromatic hydrocarbons in the coastal cities of the Bohai and Yellow Seas in northern China. *Chemosphere* **2021**, *279*, No. 130565.

(61) Lin, Y.; Ma, Y.; Qiu, X.; Li, R.; Fang, Y.; Wang, J.; Zhu, Y.; Hu, D. Sources, transformation, and health implications of PAHs and their nitrated, hydroxylated, and oxygenated derivatives in PM2.5 in Beijing. *J. Geophys. Res.: Atmos.* **2015**, *120* (14), 7219–7228.

(62) Wang, L.; Wang, X.; Gu, R.; Wang, H.; Yao, L.; Wen, L.; Zhu, F.; Wang, W.; Xue, L.; Yang, L.; Lu, K.; Chen, J.; Wang, T.; Zhang, Y.; Wang, W. Observations of fine particulate nitrated phenols in four sites in northern China: Concentrations, source apportionment, and secondary formation. *Atmos. Chem. Phys.* **2018**, *18* (6), 4349–4359.

(63) Teich, M.; van Pinxteren, D.; Wang, M.; Kecorius, S.; Wang, Z.; Müller, T.; Moćnik, G.; Herrmann, H. Contributions of nitrated aromatic compounds to the light absorption of water-soluble and particulate brown carbon in different atmospheric environments in Germany and China. *Atmos. Chem. Phys.* **2017**, *17* (3), 1653–1672.

(64) Yan, J.; Wang, X.; Gong, P.; Wang, C.; Cong, Z. Review of brown carbon aerosols: Recent progress and perspectives. *Sci. Total Environ.* **2018**, *634*, 1475–1485.

(65) Lei, Y.; Shen, Z.; Wang, Q.; Zhang, T.; Cao, J.; Sun, J.; Zhang, Q.; Wang, L.; Xu, H.; Tian, J.; Wu, J. Optical characteristics and source apportionment of brown carbon in winter PM2.5 over Yulin in Northern China. *Atmos. Res.* **2018**, *213*, 27–33.