Improving the predictions of black carbon (BC) optical properties at various aging stages using a machine-learning-based approach

Baseerat Romshoo[1,2,*], Jaikrishna Patil[3,a,*], Tobias Michels[3], Thomas Müller[1], Marius Kloft[3], and Mira Pöhlker[1,2,4]

[1]Atmospheric Microphysics Department, Leibniz Institute for Tropospheric Research, 04318 Leipzig, Germany
[2]Multiphase Chemistry Department, Max Planck Institute for Chemistry, 55128 Mainz, Germany
[3]Department of Computer Science, RPTU Kaiserslautern-Landau, 67653 Kaiserslautern, Germany
[4]Faculty of Physics and Earth Sciences, Leipzig Institute for Meteorology, Leipzig University, 04103 Leipzig, Germany
[a]now at: Arizona State University, 699 S Mill Ave, Tempe, AZ, 85281, USA
[*]These authors contributed equally to this work.

Correspondence: Baseerat Romshoo (baseerat@tropos.de) and Marius Kloft (marius.kloft@cs.rptu.de)

Abstract. It is necessary to accurately determine the optical properties of highly absorbing black carbon (BC) aerosols to estimate their climate impact. In the past, there has been hesitation about using realistic fractal morphologies when simulating BC optical properties due to the complexity involved in the simulations and the cost of the computations. In this work, we demonstrate that the predictions of optical properties like single scattering albedo (ω) and mass absorption cross-section (*MAC*)

5   can be improved compared to the conventional Mie-based predictions using a highly accurate benchmark machine learning algorithm. Unlike the computationally intensive simulations of complex scattering models, the ML-based approach accurately predicts optical properties in a fraction of a second. There has been an extensive evaluation procedure carried out in this study. Based on comparisons with laboratory measurements, it was demonstrated that incorporating realistic morphologies of BC significantly improved their optical properties. The results indicate that it is possible to generate optical properties in the

10   visible spectrum using BC fractal aggregates with any desired physicochemical properties, such as size, morphology, or organic coating. Based on these findings, climate models can improve their radiative forcing estimates using such comprehensive parameterizations for the optical properties of BC based on their aging stages.

1   Introduction

Black carbon (BC) aerosols are strong absorbers of solar radiation formed from incomplete combustion of fossil fuels, biofuels,

15   and biomass (Ramanathan and Carmichael, 2008; Bond et al., 2013). In the atmosphere, BC is usually found together with other types of aerosols, which form a coating around it (Sun et al., 2022). To understand the impact of BC on the environment, global climate models require information about its light scattering and absorption properties (Jacobson, 2001). The most common morphology assumed for such BC-containing aerosols in light scattering codes is a spherical core-shell shape (Bond et al., 2013). The Lorentz-Mie theory (Mie, 1908) is often used to calculate the optical properties of such spherical BC particles

20   (Bohren and Huffman, 2008). However, studies have shown significant discrepancies in the results of Lorentz-Mie theory when compared with ambient measurements (Adachi et al., 2010; Wu et al., 2018).

High-resolution Transmission Electron Microscopy (TEM) images showed that the BC particles have a fractal structure composed of numerous spherules known as primary particles (Chakrabarty et al., 2006). The shortcomings of the simplified spherical assumption of BC have caused the scientific community to develop towards the use of such realistic fractal aggregate morphology for computing the optical properties of BC (e.g., Kahnert and Kanngießer, 2020; Romshoo et al., 2021; Kahnert, 2010a; Wu et al., 2018; Liu and Mishchenko, 2018). Romshoo et al. (2022) showed that the discrepancy between modeled and measured optical properties could be reduced to 10% when an aggregate morphology is used. To simulate the optical properties of BC as fractal aggregates, the most commonly used methods are the Rayleigh-Debye-Gans (RDG) approximation (Sorensen, 2001), the discrete dipole approximation DDA (Purcell and Pennypacker, 1973), the Generalized Multi-particle Mie (GMM) method (Xu and Gustafson, 2001) and the T-matrix method (Mishchenko et al., 1996). The Multi Sphere T-Matrix (MSTM) method has found widespread applications in the research field because of its lower computational cost than other methods like the DDA while still providing good accuracy (Liu et al., 2019; Romshoo et al., 2021; Smith and Grainger, 2014).

Although MSTM has lower computational costs when compared to other numerical methods, a single simulation can still take more than 24 hours, depending on the properties of the aggregate. Generally, larger particles with open-chain structures take longer than smaller, compact aggregates. Consequently, pre-calculated databases have been developed for aggregate properties to save time for both the construction of detailed aggregates and the time-consuming optical simulations (Liu et al., 2019; Romshoo et al., 2021). Using these databases as look-up tables mitigates high computational overhead in large-scale applications. Still, this approach is limited by the range and step size of parameters chosen during the database creation. Previous work has trained machine learning (ML) models on such databases (Luo et al., 2018; Lamb and Gentine, 2021) to overcome those limitations. Once trained, those ML models provide predictions for BC optical properties in a fraction of a second. Luo et al. (2018) train a support vector regressor on a database generated using MSTM simulations. However, they do not consider any coating and use pure BC aggregates in their experiments. Their results also suggest that their model has considerable difficulties when attempting to predict optical properties for physicochemical properties not in the range of the training data. Lamb and Gentine (2021) predict a specific BC fractal aggregate instance's optical properties using a graph neural network. The input graph contains one node for each primary particle and an edge between two nodes if the distance between the corresponding primary particles is less than some threshold. The authors generate their ground truth database using the MSTM algorithm, but, like Luo et al. (2018), they do not consider any coating in their experiments.

This study demonstrates the use of a machine-learning-based approach to predict the optical properties of BC aggregates at various aging stages, including coating, which is highly relevant for atmospheric aerosols. Combining this ML-based approach with a laboratory dataset showed optical properties like single scattering albedo (ω) and mass absorption cross-section (*MAC*) can be predicted more accurately than with conventional Mie-based methods. A database of optical and physicochemical properties of BC has been built for this study, which is an extension to the previous work by Romshoo et al. (2021). We trained two ML methods on this database: kernel ridge regression (KRR) and artificial neural networks (ANN). Experiments show that these models predict the optical properties of BC aggregates regardless of their size, morphology, or composition at low computational costs and with high accuracy. The dataset used to train our ML models is freely available at Zenodo

(Romshoo et al., 2023a). Furthermore, we publish our ML models at GitHub[1] together with an easy-to-use wrapper script to allow for integration into higher-level applications. Our approach contributes to improving global climate model radiative forcing estimates by parametrizing BC optical properties using realistic fractal aggregate morphology.

The manuscript is structured as follows: Section 2 provides an overview of the physical, chemical, and optical properties of BC used in this study. Section 3 describes the machine learning techniques, including the data processing, machine learning algorithms, and evaluation procedures. In Section 4, the results demonstrate that realistic morphologies of BC can be used to accurately predict optical properties at various stages of aging. Section 5 discusses how the results compare to laboratory measurements of BC, discussing the atmosphere implications in detail, alongside potential limitations and challenges. The conclusion of the study is presented in Section 6.

2 Database of physicochemical and optical properties of black carbon fractal aggregates

The database for the physicochemical and optical properties of BC fractal aggregates has been designed to consider all the possible aging stages of BC. The database was created using 6192 particles of varying sizes, morphologies, and coating fractions. There are 35 features in the database, which are categorized into 15 physicochemical features, 13 optical features, and seven constants. In Fig. 1, you can see an overview of all the features of the database. In Table A1, the upper and lower bounds of the main features are provided.

2.1 Black carbon fractal aggregates

Black carbon is typically assumed to have a spherical morphology in atmospheric science (Bond et al., 2013). While this assumption is computationally simple and widely applicable, it does not reflect the entire life cycle of BC (Romshoo et al., 2021). Several stages are involved in the formation of BC, which result in changes in their morphology (Dong et al., 2018). TEM images showed that the particles of BC are composed of small primary particles with a radius of 5-25 nm depending on the combustion source (Chakrabarty et al., 2006). As a result of nucleation and coagulation, the BC primary particles come together to form chain-like aggregates, which then keep growing with aggregation (Sorensen, 2001). This concept led to an advanced mathematical description of BC as fractal aggregates, known as fractal law (Mishchenko et al., 2002):

$$N_{pp} = k_f \left(\frac{R_g}{a} \right)^{D_f}, \tag{1}$$

where a is the radius of the primary particle, N_{pp} is the number of primary particles, k_f is the fractal prefactor, and D_f is the fractal dimension. R_g is the radius of gyration, which characterizes the spatial size of the aggregate.

Along with BC, a complex mixture of gas-phase organic compounds is co-emitted during incomplete combustion, forming a coating around the BC aggregates (Gentner et al., 2017). As the BC aggregates stay in the atmosphere, they transform from being hydrophobic to hydrophilic due to water deposition attracting other foreign coatings (Bhandari et al., 2019). The result is that BC particles undergo complex changes in their morphology throughout atmospheric aging, transforming from bare to

---

[1]https://github.com/jaikrishnap/Machine-learning-for-prediction-of-BCFAs
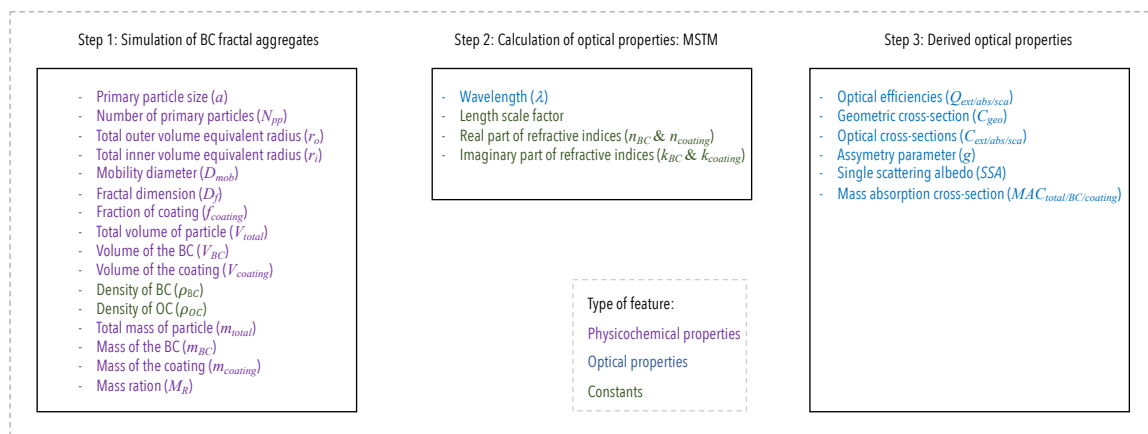
Figure 1. Overview of the various features of the database for physicochemical and optical properties of black carbon fractal aggregates. The features are arranged based on the three steps of constructing this database. As the legend at the bottom indicates, the features are further divided into physicochemical properties, optical properties, and constants.

partially coated aggregates and finally forming spherical structures embedded within external coatings (Coz and Leck, 2011). Therefore, considering BC as fractal aggregates is necessary to represent all the different stages during their atmospheric aging process.

2.2 Physicochemical features of the database

90 The BC fractal aggregate's physicochemical features include size, mass, volume, morphology, and composition. Figure 2 gives some examples of the various BC aggregate particles generated in this study.

2.2.1 Size

Primary particle size (a). The primary particle size of a BC fractal aggregate is sensitive to the emission source or flame condition. Biomass burning produces black carbon aggregates with comparatively large primary particles, ranging from 15 to
95 25 nm in radius (Chakrabarty et al., 2006). Diesel engines produce aggregates whose primary particle radius ranges between 10 and 12 nm (Guarieiro et al., 2017). On the other hand, emissions from aircraft engines consist of particles with a radius as small as 5 nm (Liati et al., 2014). There has also been research indicating that the size distribution of primary particles is largely polydisperse (Bescond et al., 2014). Liu et al. (2015) pointed out that when considering a monodisperse and polydisperse distribution of the radius of the primary particle, their resultant radiative properties differ. However, Kahnert (2010b) has
100 shown that particle light absorption is insensitive to the radius of primary particles when they are between 10 nm and 25 nm.
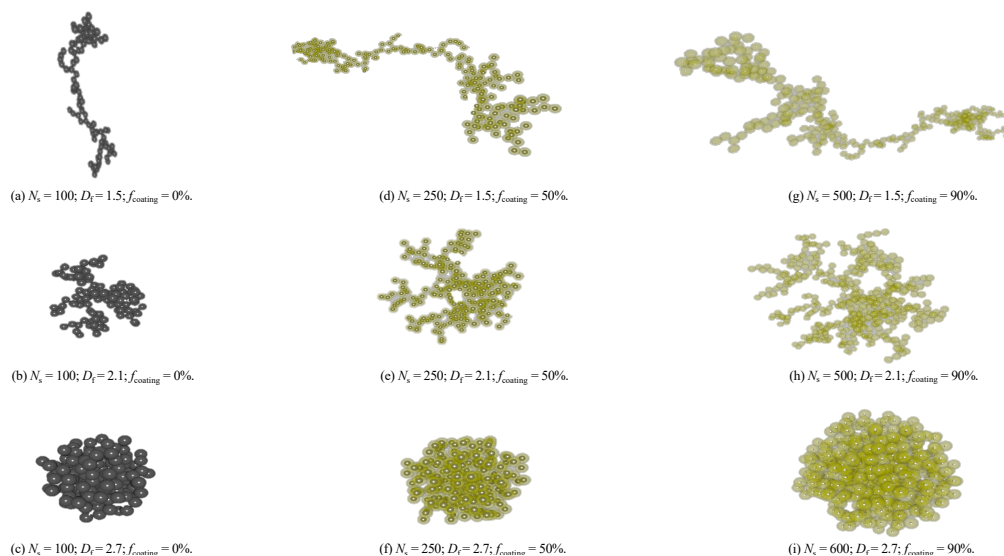
Figure 2. Visualization of the various BC aggregate particles generated in this study. Fresh BC aggregates with no external coating are shown in (a) to (c). Semi-aged BC aggregates with 50% coating are shown in (d) to (f). Aged BC aggregates with 90% coating are shown in (g) to (i)

The black carbon fractal aggregates in this study have a monodisperse distribution of the radius of the primary particle. BC aggregates were simulated with the inner diameter of the primary particle (a_i) fixed at 15 nm. In contrast, the outer radius of the primary particle (a_o), consisting of the organics, varied between 15.1 nm to 30 nm with the coating fraction following Eq. (3) in Romshoo et al. (2021).

Number of primary particles (N_{pp}). The number of primary particles determines the overall size of the particle. The BC fractal aggregates were simulated by varying N_{pp} by 5%, starting from 1 up to 1000.

Volume equivalent radius (r). The volume equivalent radius is defined as the radius of a sphere having the same volume as the BC fractal aggregate, described in Eq. (A1) in the Appendix section. The outer volume equivalent radius (r_o) was calculated for the whole BC aggregate and the coating using a_o. The inner volume equivalent radius (r_i) was calculated using a_i for the BC aggregate without the coating, i.e., pure BC.

Mobility diameter (D_m). The mobility diameter is the diameter of a sphere with the same migration velocity in a constant electric field as that of the BC fractal aggregate (Flagan, 2001). Size spectrometers can measure D_m, which is interesting for ambient and laboratory studies. We derived D_m for the entire range of N_{pp} using the conversion given by Sorensen (2001)—see Eq. (A2) in the Appendix section.

2.2.2 Mixing state: morphology and chemical composition

Fractal dimension (D_f). The fractal dimension is a parameter for morphology that quantifies the folding of BC fractal aggregates into spherical structures with increasing residence time. The value of D_f increases as an aggregate grows into a more

spherical frame. A D_f of 3 is the maximum value describing a complete sphere, whereas D_f of 1 represents an early stage open-chain-like aggregate. In the early stages of the BC aging cycle, D_f is usually between 1.5 and 1.9 (Wentzel et al., 2003).

120 When aggregates age in the atmosphere, they become more compact with a fractal dimension of up to 2.2 (Wang et al., 2017). A humid environment or foreign coatings may further reshape the BC fractal aggregates into more compact structures with a fractal dimension of up to 2.6 (Bambha et al., 2013). In this study, the range of fractal dimensions was taken from 1.5 to 2.9 with a step size of 0.2.

Fraction of coating (f_{coating}). The fraction of coating is the percentage of coating volume compared to the total volume of

125 the BC fractal aggregate. To cover all aging stages, the coating fraction was taken from 1% to 90% in increments of 5%. Note that the coating composition was constrained to non-absorbing organics in this study.

2.2.3 Mass and volume

Volume. Three features in our database describe the volume of a BC aggregate: 1. Total volume of the particle (V_{total}), 2. the volume of the BC (V_{BC}), and 3. the volume of the organic coating (V_{coating}).

130 *Mass*. Similarly, we include five features related to the mass of the BC aggregate: 1. Total mass of the particle (m_{total}), 2. the mass of the BC (m_{BC}), 3. the mass of the coating (m_{coating}), 4. the mass ratio of total mass to BC mass $\left(\frac{m_{\text{total}}}{m_{\text{BC}}}\right)$, and 5. the mass ratio of coating mass to BC mass $\left(\frac{m_{\text{coating}}}{m_{\text{BC}}}\right)$. We computed those values fixing the density of BC as $\rho_{\text{BC}} = 1.8\text{gcm}^{-3}$ (Park et al., 2004), and the density of the organic coating as $\rho_{\text{OC}} = 1.1\text{gcm}^{-3}$ (Schkolnik et al., 2007).

2.3 Optical model and the optical features of the database

135 The MSTM calculates the electromagnetic properties of a system that consists of a set of spheres (Mishchenko et al., 2004; Mackowski and Mishchenko, 2011). The tunable Diffusion Limited Aggregation (DLA) software (Wozniak et al., 2012) was used to simulate BC fractal aggregates of various physicochemical properties. In this study, we use MSTM version 3.0 (Mackowski, 2013) written in FORTRAN to compute the electromagnetic properties for fixed and random orientations. For every BC fractal aggregate, the MSTM algorithm presents an orientational average of the combined spherical expansions of each primary

140 particle. The MSTM code is best suited to calculate the optical properties of coated BC fractal aggregates since it consists of nested spheres. However, a limiting condition in MSTM is that primary particles cannot overlap. The optical features of the database are given below:

Wavelength (λ). The optical properties were calculated in the visible spectrum, i.e., for $\lambda \in \{467\text{nm}, 530\text{nm}, 660\text{nm}\}$. The real ($n$) and imaginary ($k$) part of the refractive indices for BC and coating (non-absorbing organics) at different wavelengths

145 (Kim et al., 2015) used in this study are summarized in Table A2.

Optical efficiencies ($Q_{ext/abs/sca}$). The MSTM directly calculates the extinction efficiency (Q_{ext}), absorption efficiency (Q_{abs}), and scattering efficiency (Q_{sca}) of the BC aggregate.

Geometric cross-section (C_{geo}). The geometric cross-section is the area of the cross-section of a volume-equivalent sphere given as Eq. (A3) in the Appendix section.

150   *Optical cross-sections ($C_{ext/abs/sca}$).* The optical cross-section is the product of efficiency and geometric cross-section—see Eq. (A4) in the Appendix section.

    Asymmetry parameter (g). The asymmetry parameter is directly obtained from the MSTM, defined as the intensity-weighted average of the cosine of the scattering angle.

    Single scattering albedo (ω). The single scattering albedo is the ratio of scattering efficiency (Q_{sca}) and extinction efficiency

155   (Q_{ext}), given as Eq. (A5) in the Appendix section.

    Mass absorption cross-section (MAC). The mass absorption cross section is calculated from the ratio of absorption cross section (C_{abs}) and mass (m) as detailed in Eq. (A6) in the Appendix section. The three kinds of MAC calculated in this study are total mass absorption cross-section (MAC_{total}), BC mass absorption cross-section (MAC_{BC}), and coating mass absorption cross-section ($\text{MAC}_{coating}$).

160   **3   Machine learning method for predicting optical properties of BC fractal aggregates**

As mentioned in Section 1, several high-impact applications, such as climate modeling (Jacobson, 2001), depend on accurate optical properties for specific BC particles. Hence, we propose to train an ML model on a pre-computed database containing physicochemical and corresponding optical properties of BC fractal aggregates at several life cycle stages. This model will learn patterns and structures within the data and should generalize to unseen data values when used in applications, as evidenced by

165   the success of ML in several domains (Radford et al., 2021; Ramesh et al., 2022). In this work, we train kernel ridge regression and a multi-layer perceptron on the database introduced in Section 2. The following sections detail our data processing routines, models, and evaluation procedures.

3.1   Data pre-processing

As mentioned in Section 2.2, not all physical properties in the database are independent of each other, as some can be derived

170   from others using simple formulae. Including all properties as inputs for the ML model will thus present it with redundant information, increasing its computational overhead and possibly even harming its performance. Therefore, we use only a subset of the database's properties as input for our ML models: Wavelength (λ), fractal dimension (D_f), fraction of coating ($f_{coating}$), primary particle size (a), number of primary particles (N_{pp}), volume equivalent radii (r_i and r_o), and mobility diameter (D_m).

    Similarly, a BC fractal aggregate's optical properties are also not independent. Thus, we make the ML model predict only

175   the following three properties and compute the rest using the formulae in Appendix A1: absorption efficiency (Q_{abs}), scattering efficiency (Q_{sca}), and asymmetry parameter (g).

    After feature selection, we transform input features using the Box-Cox transformation (Box and Cox, 1964), where we choose the parameter λ by maximum-likelihood estimation. We also tried to apply the Box-Cox transformation to the target features, but since this did not improve results, we decided not to use any transformation on the target features for the

180   experiments that we report in Section 4.

3.2 Kernel ridge regression

Given a labeled dataset of $N \in \mathbb{N}$ points $\{(\boldsymbol{x}^{(1)}, \boldsymbol{y}^{(1)}), (\boldsymbol{x}^{(2)}, \boldsymbol{y}^{(2)}), \ldots, (\boldsymbol{x}^{(N)}, \boldsymbol{y}^{(N)})\} \subset \mathbb{R}^D \times \mathbb{R}^{D'}$, the *regression* problem consists of finding a function $f \colon \mathbb{R}^D \to \mathbb{R}^{D'}$ such that $f(\boldsymbol{x}^{(n)}) \approx \boldsymbol{y}^{(n)}$ for all $n \in \{1, \ldots, N\}$. *Kernel ridge regression* (KRR) (Shawe-Taylor and Cristianini, 2004) learns a function of the form $f(\boldsymbol{x})_d = \sum_{n=1}^{N} \alpha_{nd}^* k(\boldsymbol{x}^{(n)}, \boldsymbol{x})$, where $k \colon \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is a positive semidefinite *kernel function* (Cortes and Vapnik, 1995) and $\boldsymbol{\alpha}^* \in \mathbb{R}^{N \times D'}$ is a solution of the following convex optimization problem:

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^{N \times D'}} \lambda \operatorname{Tr}(\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}) + \|\mathbf{Y}^T - \mathbf{K}\boldsymbol{\alpha}\|_{\text{Fro}}^2, \tag{2}$$

where $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the so-called *kernel matrix* defined by $K_{ij} = k(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$, $\lambda \in \mathbb{R}^+$ is a trade-off parameter that controls the influence of the regularization term, and $\mathbf{Y} = (\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(N)})^T \in \mathbb{R}^{N \times D'}$. Note that Eq. (2) has a closed-form solution:

$$\boldsymbol{\alpha}^* := (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{Y}. \tag{3}$$

Popular choices for the kernel function include the polynomial kernel $k(\boldsymbol{x}, \boldsymbol{x}') = (a\boldsymbol{x}^T \boldsymbol{x}' + b)^c$, where $a, b \in \mathbb{R}^*$ and $c \in \mathbb{N}$ are parameters, and the Gaussian or radial basis function (RBF) kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\gamma \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2\right), \tag{4}$$

where $\gamma \in \mathbb{R}^+$ is a parameter called *bandwidth*.

We use *scikit-learn*'s KRR implementation[2] with the RBF kernel for our experiments. This method has two hyperparameters that need tuning: the RBF kernel's $\gamma \in \mathbb{R}^+$ and $\lambda \in \mathbb{R}^+$ (see Eq. (2)). We optimize hyperparameters using grid search—please see Table B1 for the grid and Section 3.4 for more detailed information on our evaluation procedure.

3.3 Artificial neural networks

Artificial neural networks (ANN) constitute one of the founding pillars of ML's success during the last ten years. Originally, their design was inspired by the structure of neurons inside the nervous system of several organisms (Rosenblatt, 1958). Most designs used in practice nowadays abandoned that idea, but the name remains.

In our experiments, we use a *feed-forward* ANN, sometimes also called *multi-layer perceptron* (MLP). It consists of an arbitrary number ($L \geq 2$) of *layers*, of which the first is called *input* layer, the last is called *output* layer, and all layers in between are called *hidden* layers. Each layer consists of a certain number of *neurons*, which are connected to the neurons in the previous and the following layer

Formally, we can define an MLP as a function $f \colon \mathbb{R}^D \to \mathbb{R}^{D'}$ that is composed of $L - 1$ layer functions, i.e., $f(\boldsymbol{x}) := f^{(L-1)}(f^{(L-2)}(\ldots f^{(1)}(\boldsymbol{x}) \ldots))$, where each $f^{(l)} \colon \mathbb{R}^{D^{(l)}} \to \mathbb{R}^{D^{(l+1)}}$ represents a connection between two layers. They are defined as $f^{(l)}(\boldsymbol{x}) := \sigma^{(l)}(\mathbf{W}^{(l)} \boldsymbol{x} + \boldsymbol{b}^{(l)})$, where $\mathbf{W}^{(l)} \in \mathbb{R}^{D^{(l+1)} \times D^{(l)}}, \boldsymbol{b}^{(l)} \in \mathbb{R}^{D^{(l+1)}}$ are learnable parameters, and $\sigma^{(l)}$ is a so-called *activation function* that is applied separately to each element of its input vector. Common choices for $\sigma^{(l)}$ include,

---

[2]https://scikit-learn.org/stable/modules/generated/sklearn.kernel_ridge.KernelRidge.html

210 for example, the rectified linear unit (ReLU) $\sigma^{(l)}(x) = \max(x, 0)$ or the $\tanh$ function. We use the same activation function for each layer except the last, where we always use the identity function, i.e., $\sigma^{(L-1)}(x) := x$. Finally, $D^{(l)} \in \mathbb{N}$ denotes the number of neurons in layer l, with $D^{(1)} = D$ and $D^{(L)} = D'$.

The number of hidden layers, the number of neurons in those hidden layers, and the activation function are usually chosen by a human before training a neural network. Together, they define the *architecture* of the MLP. We can *learn* values for the

215 parameters $\mathbf{W} := (\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(L-1)})$ and $\boldsymbol{b} := (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(L-1)})$ by minimizing a so-called *loss function* $\mathcal{L} : \mathbb{R}^{D'} \times \mathbb{R}^{D'} \to \mathbb{R}$ over a dataset:

$$\min_{\mathbf{W}, \boldsymbol{b}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}(f(\boldsymbol{x}^{(n)}), \boldsymbol{y}^{(n)}). \tag{5}$$

When solving a regression problem, the most common choice for $\mathcal{L}$ is the *squared loss* $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y}) := \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2$, but practitioners sometimes use other loss functions as well, for example, the *Huber loss* (Huber, 1964):

$$220 \quad \mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \sum_{d=1}^{D'} \begin{cases} \frac{1}{2}(y_d - \hat{y}_d)^2 & \text{if } |y_d - \hat{y}_d| \leq \delta \\ \delta\left(|y_d - \hat{y}_d| - \frac{1}{2}\delta\right), & \text{otherwise,} \end{cases} \tag{6}$$

where $\delta \in \mathbb{R}^+$ determines the cut-off point between squared and absolute loss and is usually chosen as $\delta = 1$. The entire procedure of adapting the ANN's parameters using a given dataset is called *training* in the ANN literature.

Note that, in general, Eq. (5) is not convex and does not have a closed-form solution. Hence, practitioners use gradient-based optimization methods, i.e., variants of mini-batch stochastic gradient descent (SGD) (Bottou et al., 2018), to find a local

225 minimum of Eq. (5).

For our experiments, we implemented an MLP using *keras*[3]. Table B2 contains the hyperparameter grid for the MLP's architecture and training procedure.

3.4 Evaluation procedure

After preprocessing, we split the database into a training and a test set. Models perform their training procedures and hyperpa-

230 rameter tuning on the training set only, and we then evaluate the model's performance exclusively on the test set. We consider three different methods of performing this split—each one intends to measure another aspect of the model's performance:

1. *Random split*: We randomly assign each point in the database to either the training or test set. Note that we use 30% of the data for the test set and the rest for the training set. Using this split, the training and test set's feature distribution should be similar. Thus, measuring the performance on the test set produces a general measure of the model's capability

235   to learn the underlying patterns in the data.

---

[3]https://keras.io/

[1]https://keras.io/api/layers/activations/#sigmoid-function

[2]http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

[3]https://keras.io/api/losses/regression_losses/#logcosh-class

2. *Interpolation split*: Here, we choose a feature and a certain range in the middle of that feature's range and choose all data points within that range as the test set. To achieve high test scores, the model must, therefore be capable to *interpolate* predictions for data points it has not seen during training. Table B3 shows the features and ranges used for the interpolation split.

240   3. *Extrapolation split*: Similar to the interpolation split, we also consider choosing a test set at the boundaries of certain features. This measures the model's extrapolation capabilities. Table B4 shows the features and ranges used for the extrapolation split.

We use the *mean absolute error* (MAE) as our primary performance metric: given a dataset $\mathcal{D} \subset \mathbb{R}^D \times \mathbb{R}^{D'}$, and our prediction model $f\colon \mathbb{R}^D \to \mathbb{R}^{D'}$ we can compute the MAE as follows:

245
$$\mathrm{MAE}(f, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}} \|\boldsymbol{y} - f(\boldsymbol{x})\|_1, \tag{7}$$

where $\|\boldsymbol{z}\|_1 := \sum_{d=1}^{D'} |z_d|$ is the L_1-norm.

We split the training set once more into a train and a validation set using the random split method during the training phase. Our models then train on the train set for all possible hyperparameter configurations defined in the grid, and we record the MAE on the validation set for each combination. Finally, we choose the combination with the lowest MAE
250   and evaluate the corresponding model's MAE on the test set.

4   Performance of the machine learning models

The MAEs for our experiments are reported in Table 1. In the case of the random split, both ML models are quite accurate, with the percentage of MAEs ranging from 0.1 to 0.4 % when compared to the feature range. The comparison of the two ML methods for random split showed that KRR generally results in a lower MAE for predictions of Q_{abs} and
255   Q_{sca}. Contrary to this, the ANN was able to predict g with a lower MAE. In line with expectations, the MAE for the splits based on interpolation and extrapolation is somewhat higher. The errors, however, are still considered relatively minor when compared to the features' range. The extrapolation and interpolation experiments were used to test the performance of the ML algorithm under various scenarios of data available for training. The ML models we publish for use in applications were trained on the entire dataset using the best parameters from the random split experiments. As a
260   result, the errors should be similar to those we report for the random split here.

In Fig. 3, the error distributions for the ML methods are presented for different experimental scenarios when the data is split according to the parameter fractal dimension. The median error is close to zero for the random and interpolation splits, meaning that our models do not generally over- or underestimate any optical value. The distribution of errors (excluding outliers) for the random and interpolation splits is relatively narrow, indicating that most test points have
265   minor errors. In the extrapolation case, both ML models exhibit bias, such as overestimation of Q_{sca} by the ANN and

Table 1. Mean absolute errors of the predicted optical properties for different experiments. The training data for the interpolation split consists of fractal dimensions in $D_f = [1.5, 2.1) \cup (2.5, 2.9]$, whereas the extrapolation split uses $D_f = [1.5, 2.5]$.

Optical property	Random split		Interpolation split		Extrapolation split		Feature range
	KRR	ANN	KRR	ANN	KRR	ANN	
Q_{abs}	0.0022	0.0039	0.0122	0.0287	0.0329	0.0354	$0-2$
Q_{sca}	0.0019	0.0031	0.0224	0.0466	0.0393	0.0939	$0-2$
g	0.0044	0.0038	0.0429	0.0289	0.0879	0.0485	$0-1$
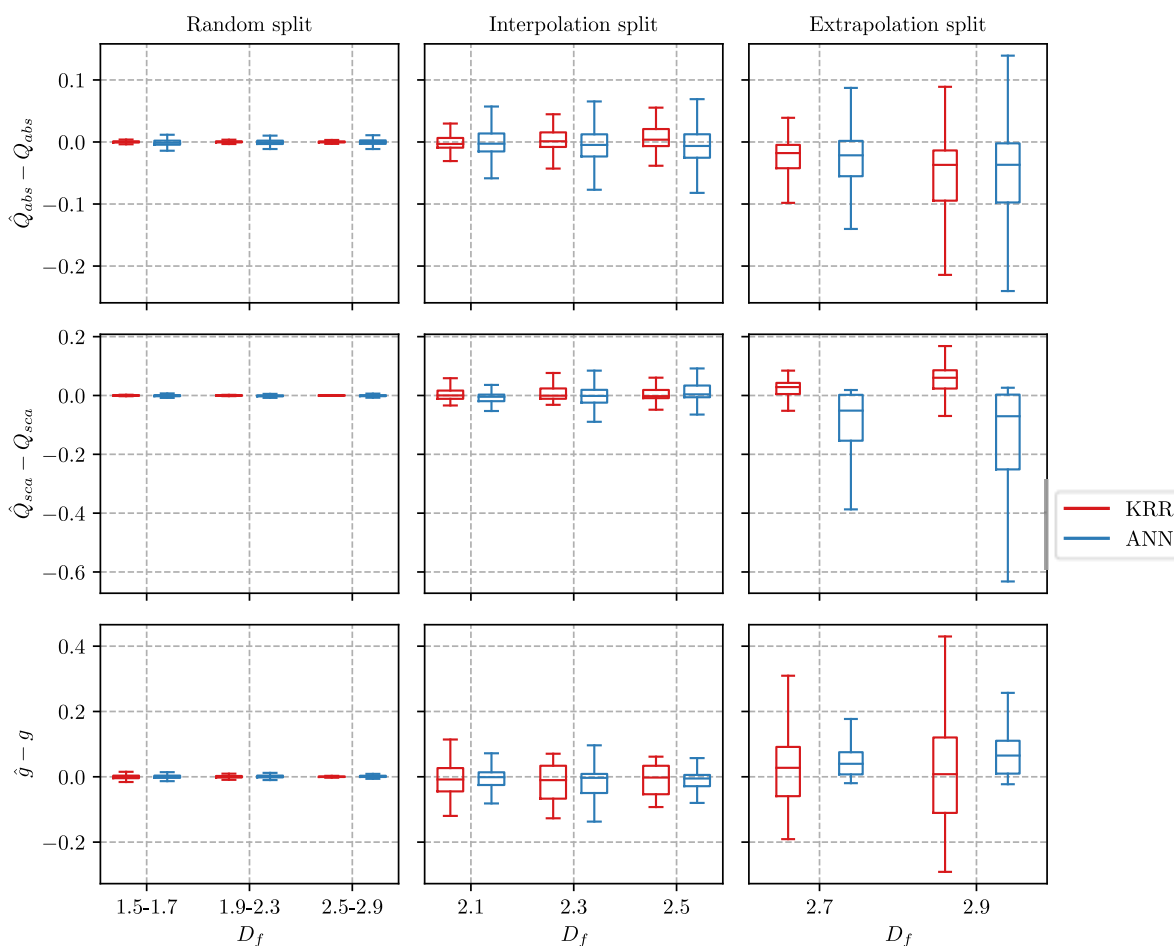
Figure 3. Boxplots summarizing the error between the predicted value $(\hat{Q}_{\mathrm{abs}}, \hat{Q}_{\mathrm{sca}}, \hat{g})$ and the true value for three optical properties. The training data for the interpolation split consists of fractal dimensions in $D_f = [1.5, 2.1) \cup (2.5, 2.9]$, whereas the extrapolation split uses $D_f = [1.5, 2.5]$. Note that we do not show outliers to prevent unnecessary clutter.
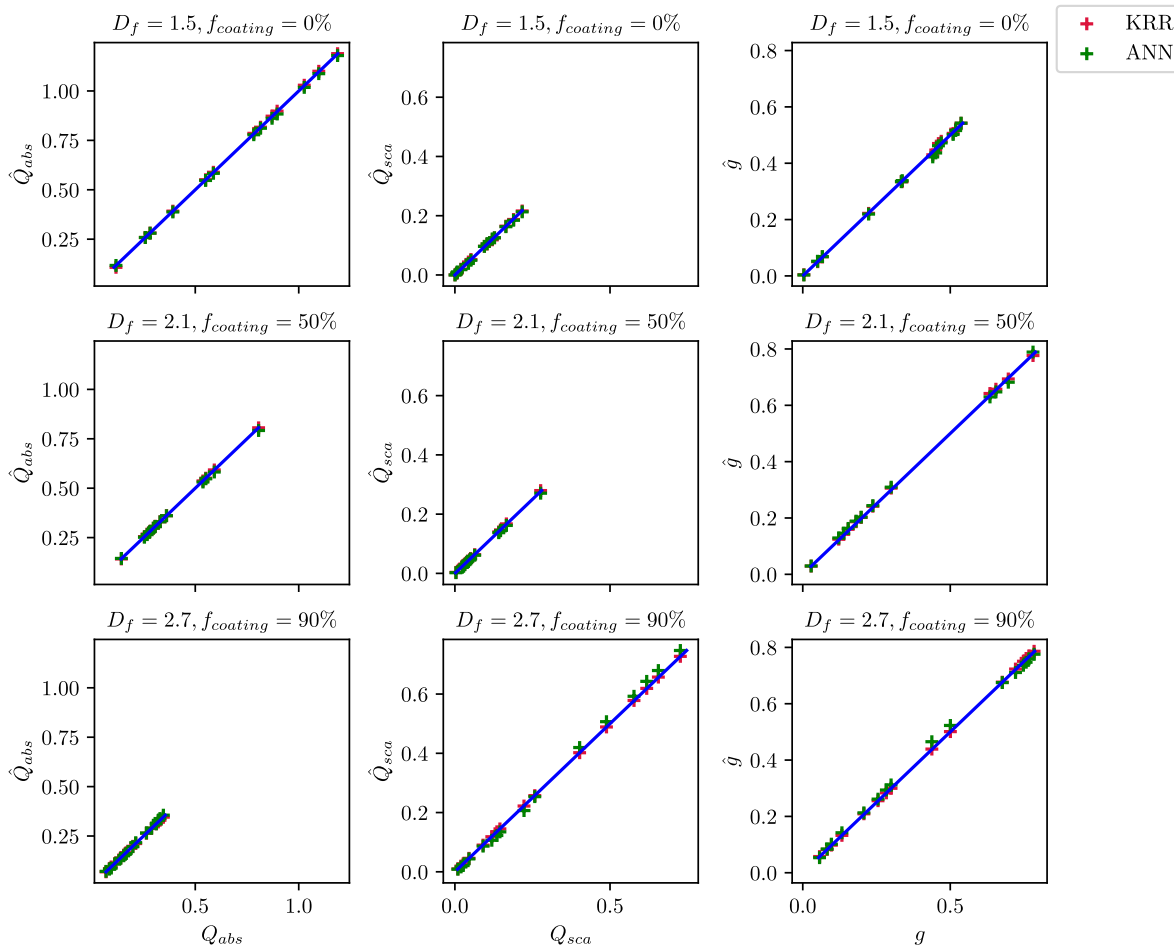
Figure 4. Comparison of the predicted optical properties with their true values when the ML models are trained on a random subset of data.

overestimation of g by the KKR. However, the mean absolute error, even for the extrapolation split, is 1.5 to 8 %, which is still within reasonable limits. The interpolation and extrapolation results are similar if training and test data are split according to the parameter of the fraction of coating. A more detailed discussion can be found in Fig. C1 in the appendix. Overall, the narrow box plots of the errors in the random split demonstrate the effectiveness of the ML algorithms for

270   predicting the optical properties of coated BC fractal aggregates.

A one-to-one comparison was performed between the estimates and true values to better understand how the ML methods perform in predicting optical properties. Fig. 4 compares the estimated and true values for the wavelength of 660 nm when the training and test data are randomly split. The values of $\hat{Q}_{abs}, \hat{Q}_{sca}, \hat{g}$ obtained from the KRR and ANN methods are compared to the true values derived from the MSTM method. The performance of both ML methods was studied for

275   BC fractal aggregates with three representative morphologies, and coating fractions ($D_\text{f} = 1.5$ & $f_\text{coating} = 0\%$; $D_\text{f} = 2.1$
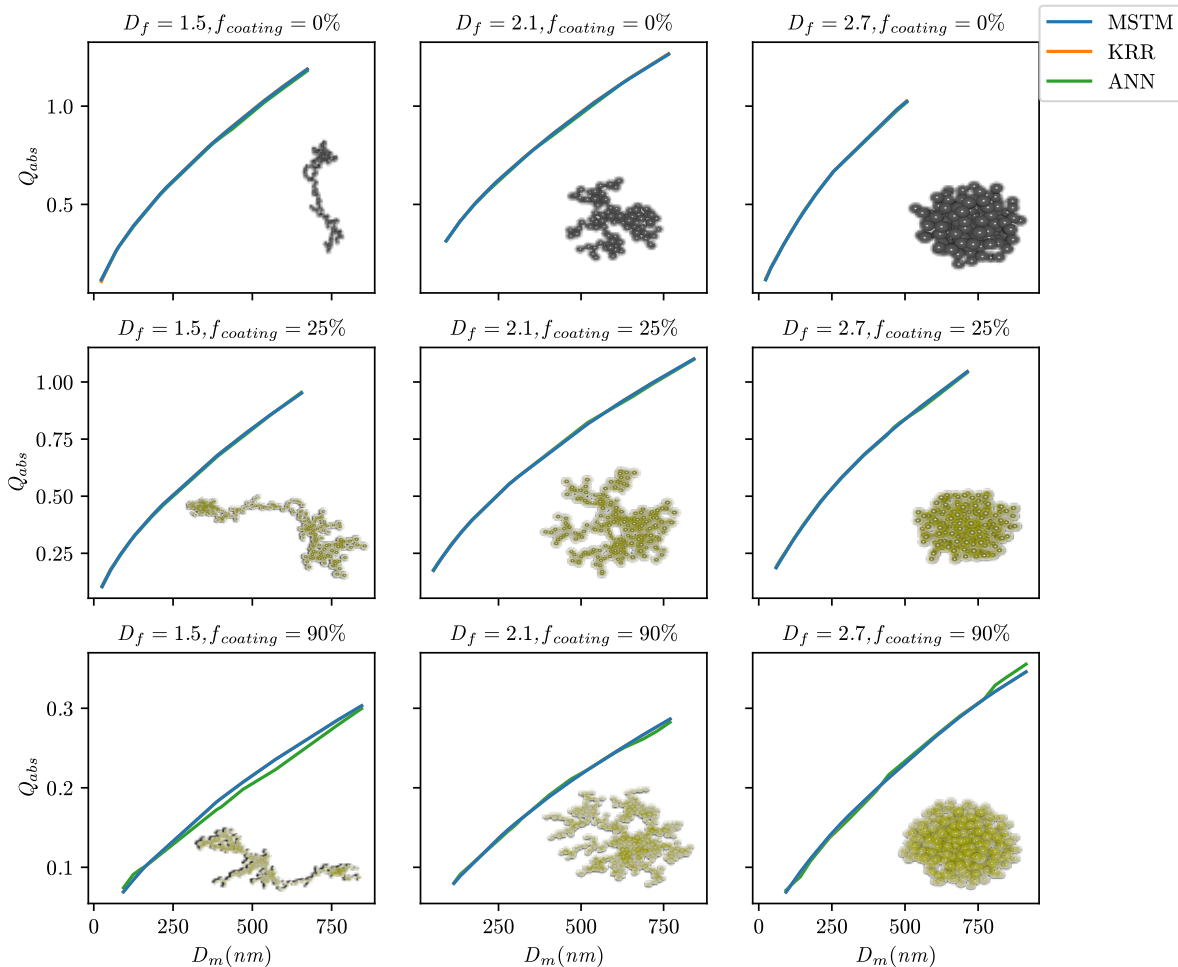
Figure 5. Absorption efficiency (Q_{abs}) at a wavelength of 660 nm predicted using KRR and ANN for nine representative BC aggregates with a variety of morphologies (represented by D_f) and coatings (represented by $f_{coating}$). Both models were trained on a random split of training data.

& $f_{coating} = 50\%$; $D_f = 2.7$ & $f_{coating} = 90\%$). There was reasonable agreement between the KRR and the ANN for all the sub-cases. Therefore, the machine learning models appear to be applicable in a broader context. The model does not overfit with different coating fractions and complex morphologies.

During their lifetime, BC fractal aggregates undergo complex changes in size, composition, and morphology due to atmospheric processing. Figure 5 shows a visualization of how the ML predictions compare to the MSTM reference for different aging scenarios for BC fractal aggregates. It compares the estimated and true values of the optical properties for the random split. The models trained using a random split of training data generally show a good agreement with the ground-truth data over the entire range of D_m. Only in the case of $f_{coating} = 90\%$, the ANN seems to underestimate the

280

Table 2. Training time for 18526 samples in the dataset and prediction time per sample in seconds. Values were recorded on a machine with Intel(R) Core(TM) i7-9750H CPU, 8 GB RAM, and NVIDIA GeForce GTX 1650 GPU.

ML model	Training time	Prediction time
KRR	33.3s	0.0006s
ANN	1770s	0.0005s

true value of Q_{abs} for intermediate values of D_m. Appendix C3 contains similar plots for the interpolation and extrapolation split. In general, errors tend to increase with increasing aggregate sizes for the interpolation and extrapolation splits. The ML models we publish are based upon random split experiments, and Fig. 5 shows how well both the ML methods provide accurate estimates of the optical properties of BC fractal aggregates at each aging stage.

Apart from making accurate predictions, our ML models should also be fast to provide a benefit over time-consuming simulations. Hence, we recorded the time needed to train on the entire training dataset and the time for making a single prediction in Table 2. As a result, the prediction time of both algorithms is less than one millisecond, which is a drastic improvement compared to the MSTM method, which can take up to 24 hours, depending on the particle. It should be noted that the prediction time for the ANN does not depend on the input data. Training the models takes comparatively longer, but it is usually done offline. Therefore, it is irrelevant for users using the pre-trained models we provide for their applications (see Section 7).

5 Atmospheric implications

It is important to incorporate the fractal morphology of BC global model calculations as the BC radiative forcing can decrease up to 61% when a particle becomes more compact and aged (Romshoo et al., 2021). In the past, there has been hesitation about using realistic fractal morphologies when simulating BC optical properties due to the complexity involved in the simulations and the cost of the computations. Our ML algorithms address both problems, encouraging the use of realistic fractal morphologies in large-scale applications. When using optical models, the computation time could be as long as one day for certain complexly shaped BC fractal aggregates. Using our advanced algorithms, we can predict the same optical properties in just a matter of seconds.

The advantage of using such ML algorithms is the possibility of simulating the optical properties of BC fractal aggregates at any desired size, morphology, and coating fraction or organics. Previous studies have trained ML models on similar datasets; however, they considered the fractal aggregates to be purely BC in their experiments (Luo et al., 2018; Lamb and Gentine, 2021). In the atmosphere, BC fractal aggregates are primarily found in conjunction with other materials, such as organic carbon. It is, therefore, more relevant to predict the optical properties of BC fractal aggregates with organic coatings for atmospheric applications. To give an example of applying the ML algorithm to real-world atmospheric
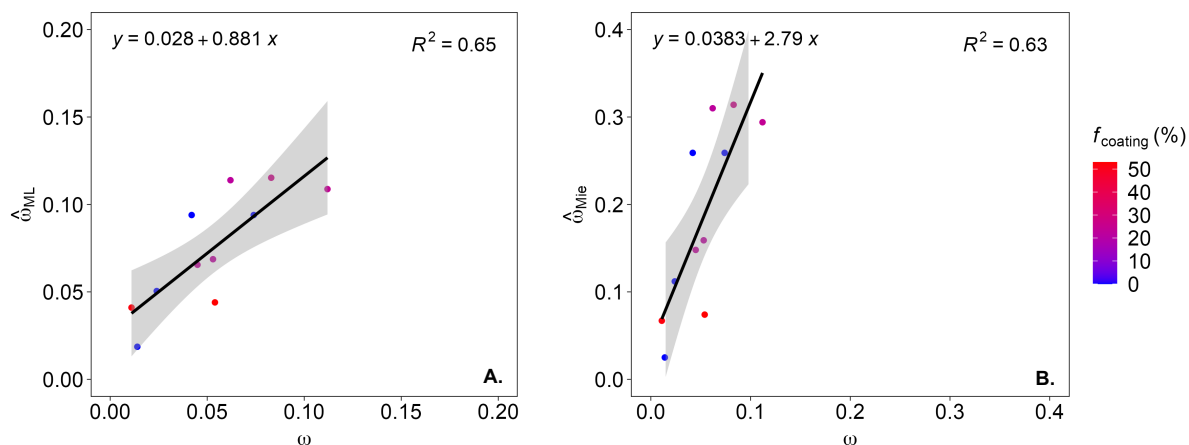
Figure 6. Single scattering albedo ω for coated BC particles generated in a laboratory study at different f_{organics} (Romshoo et al., 2022). **(A)** compares the $\hat{\omega}_{\text{ML}}$ with the measured ω from the laboratory experiment. **(B)** compares the $\hat{\omega}_{\text{Mie}}$ with the measured ω.

research, we predicted the optical properties of laboratory-generated soot for experiments described in the Table. 1 of our previous study (Romshoo et al., 2022). Figure 6A compares the ML algorithm predicted single scattering albedo ($\hat{\omega}_{\text{ML}}$) with the measured ω from the laboratory experiment. The $\hat{\omega}_{\text{ML}}$ predictions are in good agreement with the measured results for a range of f_{organics} going up to 55%. On the contrary, Figure 6B demonstrates that if the conventionally used Mie-core-shell theory is used, the predictions are overestimated by a large margin. The ML predictions of *MAC* are also compared to the measured *MAC* and Mie-based predictions, whose results are given in Figure C7 of the Appendix. The predictions $\hat{MAC}_{\text{ML}}$ were found to be less sensitive to the change in D_{mob}. Due to a lack of monodisperse mass measurements, comparing the predictions and measured values is not so straightforward. However, one can see that the discrepancies in the ML-based predictions of *MAC* are comparatively lower than the Mie-derived *MAC* values.

Based on the success of the ML-based approach in predicting the optical properties of coated BC particles, it has great potential for future development to predict the optical properties of mixtures of BC and other aerosols. Because such a study would be exhaustive, we initially tested this approach on BC fractal aggregates and organic coatings to determine its effectiveness. Further research is necessary to develop an ML algorithm with features representing different morphological shapes and other chemical compositions, such as inorganics. In the long run, the goal should be to develop an ML algorithm that can be used to integrate all atmospheric aerosols into global climate models. To develop such a universal algorithm for all atmospheric aerosols, we must incorporate the conventional spherically shaped particles into the current prediction algorithm to represent the fraction of aged aerosols.

5.1 Limitations and future challenges

The experiments conducted for this study show that our ML methods predict the optical properties of BC fractal aggregates with high accuracy as long as they are trained on sufficient data. However, the interpolation and extrapolation experiments show that the performance of both KRR and ANN significantly deteriorates when entirely removing certain ranges from the training data. This suggests that our models possess only limited generalization capabilities. Still, it should be noted that we train the models for practical use on the entire physically feasible range of D_f and f_{coating}. Hence, those models will not have to extrapolate for any reasonable inputs.

Our models treat the wavelength λ as a continuous variable, meaning they should support computing optical properties at wavelengths that are not part of the training data. However, we did not test the models' generalization capabilities about the wavelength since omitting just one wavelength from the training data would reduce the dataset size by one-third. Generating more ground truth data for other wavelengths requires refractive indices of BC and organics for that specific wavelength, which are unavailable in the literature. Even if they were available, it would be time-consuming as MSTM simulations can take a long time to compute. Nevertheless, examining the models' generalization capabilities on other wavelengths in the future would be interesting.

Both KRR and the ANN provide only a single-point prediction for each input. In particular, their estimate does not quantify any uncertainty in the prediction. Bayesian ML methods such as Gaussian Process Regression (Rasmussen and Williams, 2005) can provide information about the uncertainty of a prediction via credible intervals as they return an entire probability distribution instead of a single point estimate. Thus, it would be interesting to examine Bayesian ML for the prediction of BC fractal aggregates' optical properties. This method could be further developed for reporting the predictions for an ensemble of BC-containing aerosols with various physicochemical properties. However, applying them directly to our problem is not trivial since the assumptions made by their statistical model (e.g., target variables follow a multivariate Gaussian distribution) are often violated in practice. Therefore, we leave the application of Bayesian ML to the BC aerosol problem to future work.

6 Conclusions

The present study demonstrated that the predictions of BC optical properties can be improved by incorporating their realistic morphologies. Unlike the computationally intensive simulations of complex scattering models, the ML-based approach accurately predicts optical properties in fractions of a second. In conjunction with a laboratory dataset, it was shown that optical properties like single scattering albedo ω and mass absorption cross-section (*MAC*) can be predicted with greater accuracy than with a Mie-based approach. Using an extensive database for the physicochemical and optical properties of BC fractal aggregates, we trained two ML models—KRR and ANN—that can be used to predict the optical properties of coated BC aggregates at all aging stages. In particular, we could accurately predict the optical properties

in the visible spectrum for BC fractal aggregates of any desired size, shape, and fraction of organic coating. Thus, this work illustrates the use of this realistic approach in real-world atmospheric research applications.

We summarize the key conclusions of the study as:

– *Active investigation area*: BC is a highly relevant and active field of research, as it affects the climate system and human health. Global climate models require information about the optical properties of BC to simulate their radiative forcing. The BC research will benefit from using this ML algorithm to generate the optical properties of BC based on more realistic fractal aggregates.

– *Broader application*: The ML algorithm can predict the optical properties absorption efficiency, scattering efficiency, and asymmetry parameter for a wide range of BC fractal aggregates with physiochemical properties specified by particle size, morphology, and coating fraction. Previous studies did not consider the critical parameter of coating fraction in their ML models. Therefore, even though we discuss the results in terms of *Number of primary particles* (N_{pp}), the user is additionally able to specify the particle size in terms of *Volume equivalent diameter* (R_v) or *Mobility diameter* (D_m) depending on numerical or in-situ based nature of the study. We tested the use of the ML algorithm for predicting the scattering properties of laboratory-generated soot particles and found that it was well in agreement with the measured values.

– *User-friendly*: We published a simple Python script that allows users to predict optical properties for BC fractal aggregates using our pre-trained models at GitHub[4]. The user must specify the physicochemical properties of a BC fractal aggregate as a .csv file, from which the prediction script generates the corresponding optical properties using either KRR or ANN.

– *Low computational and energy costs*: Our ML models have a low computational cost, taking fractions of a second to provide the predictions on a run-of-the-mill desktop PC. The same optical properties could take more than 24 hours to be generated when using a T-matrix optical model. Using such ML algorithms will thus reduce the energy expenditures associated with running optical models on supercomputers.

– *Citability and reproducibility*: The dataset used for developing the ML algorithm is available for download at Zenodo (Romshoo et al., 2023a). Furthermore, the baseline experiments can be reproduced with the code that is openly available on GitHub[5].

In summary, we demonstrated the feasibility of incorporating the realistic morphology of BC to improve their predictions of optical properties using a first-of-its-kind machine-learning approach. This ML-based approach constitutes a significant step forward in BC aerosol research in two ways: first, it is the first attempt to provide optical properties of coated BC fractal aggregates at different stages of atmospheric aging using realistic representations. Second, this approach significantly reduces the heavy computational costs of using previous complex scattering models. Previous studies of BC

---

[4]https://github.com/jaikrishnap/Machine-learning-for-prediction-of-BCFAs
[5]https://github.com/jaikrishnap/Optical-properties-of-black-carbon-aggregates

avoid using complex scattering theories because of the high computational costs and prefer the more simplistic Mie theory. This research will be further developed in the future with the final goal of accurately predicting the optical properties of any mixture of atmospheric aerosols.

7 Code and data availability

The dataset of simulated physio-chemical and optical properties that we describe in Section 2 is available at https://doi.org/10.5281/zenodo.7523058 (Romshoo et al., 2023a). In case they want to reproduce any of the results in this work, readers may find the entire source code that we used to perform the ML-based experiments and generate figures included in this work on https://github.com/jaikrishnap/Optical-properties-of-black-carbon-aggregates (Romshoo et al., 2023b).

Furthermore, we provide a Python script that predicts the optical properties of BC fractal aggregates using the ML-based models we trained in a GitHub repository https://github.com/jaikrishnap/Machine-learning-for-prediction-of-BCFAs (Romshoo et al., 2023c). In order to run the prediction script, the physio-chemical properties need to be provided as a .csv file that contains the fractal dimension D_f, the fraction of coating f_{coating}, and the wavelength (λ) at which the optical properties should be calculated. Users may specify the particle size either by specifying the number of primary particles (N_{pp}) or the mobility diameter (D_m). The prediction script will then generate a .csv file with the corresponding optical properties for the provided physio-chemical properties. Please check the README file inside the repository for more detailed information on using the script.

Appendix A: Details about the physiochemical and optical properties of BC fractal aggregates

A1 Formulae

The volume equivalent radius (r) is defined as the radius of a sphere having the same volume as the BC fractal aggregate, given as:

$$r = a \sqrt[3]{N_{pp}} \tag{A1}$$

where N_{pp} is the number of primary particles, and a is the radius of a single primary particle. The outer volume equivalent radius (r_o) was calculated for the whole BC aggregate and the coating using a_o. The inner volume equivalent radius (r_i) was calculated using a_i for the BC aggregate without the coating, i.e., pure BC.

The mobility diameter of a sphere (D_m) was defined by Sorensen (2001) as:

$$D_m = 2a_o \left(10^{-2x+0.92}\right) N_{pp}^x, \tag{A2}$$

where N_{pp} is the number of primary particles, and a_o is the radius of a primary particle with coating, x is the mobility mass scaling exponent given by $x = 0.51 Kn^{0.043}, 0.46 < x < 0.56$. Kn is the Knudsen number, which is the ratio of the molecular free path to the agglomerate mobility radius. The error estimated in the mobility mass scaling exponent (x) is ± 0.02.

The geometric cross-section (C_{geo}) is the area of the cross-section of the volume-equivalent sphere, given as:

$$C_{\text{geo}} = \pi r_o{}^2 \tag{A3}$$

The optical cross-sections ($C_{\text{ext/abs/sca}}$) are defined as the product of efficiency ($Q_{\text{ext/abs/sca}}$) and geometric cross-section (C_{geo}) as:

$$C_{\text{ext/abs/sca}} = Q_{\text{ext/abs/sca}} C_{\text{geo}} \tag{A4}$$

The single-scattering albedo (ω) is derived from the ratio of the scattering efficiency (Q_{sca}) to the extinction efficiency (Q_{ext}) as:

$$\omega = \frac{Q_{\text{sca}}}{Q_{\text{ext}}} \tag{A5}$$

The total mass absorption cross-section ($\text{MAC}_{\text{Total}}$), BC mass absorption cross-section (MAC_{BC}), and coating mass absorption cross-section ($\text{MAC}_{\text{Coating}}$) were calculated from the ratio of (C_{abs}) with total mass (m_{Total}), BC mass (m_{BC}), and coating mass (m_{Coating}), respectively, as:

$$\text{MAC}_{\text{total/BC/coating}} = \frac{C_{\text{abs}}}{m_{\text{total/BC/coating}}} \tag{A6}$$

A2 Range of features and constants

Table A1. Upper and lower bounds of some features from the database of physicochemical and optical properties of black carbon fractal aggregates.

Parameter	Lower limit	Upper limit
Wavelength (λ)	467	660
Fractal dimension (D_f)	1.5	2.9
Fraction of coating (f_{coating})	0	90
Primary particle size (a)	15	29
Number of primary particles (N_{pp})	1	1000
Outer volume equivalent radius (r_o)	12	290
Inner volume equivalent radius (r_i)	12	150
Mobility diameter (D_m)	17	1561
Extinction cross-section (C_{ext})	0.043	3.02
Absorption cross-section (C_{abs})	0.041	1.75
Scattering cross-section (C_{sca})	0.00038	1.82
Asymmetry parameter (g)	0.00036	0.91
Single scattering albedo (SSA)	0.00030	0.776
Mass absorption cross-section (MAC)	3.89	24.5

Table A2. Refractive indices (both real and imaginary parts) of BC and organics at various wavelengths in the visible range (Kim et al., 2015).

Parameter	Wavelength (nm)		
	467	530	660
n_{BC}	1.92	1.96	2.00
k_{BC}	0.67	0.65	0.63
n_{coating}	1.59	1.47	1.47
k_{coating}	0.11	0.04	0.00

Appendix B: Details about the machine learning methods

Table B1. Hyperparameter values for the kernel ridge regression (KRR) experiments.

Parameter	Values
RBF kernel bandwidth (γ)	$0.0001, 0.001, 0.01, 0.05, 0.1,$ $0.5, 0.75, 1$
Regularization coefficient (λ)	$0.0001, 0.001, 0.01, 0.05, 0.5,$ $0.75, 1$

Table B2. Hyperparameter values for the multi-layer perceptron (MLP) experiments.

Parameter	Values
Number of layers (L)	$3, 4, \ldots, 12$
Number of neurons ($D^{(l)}$)	$1, 8, 16, 32, 64, 128, 256, 512, 1024$
Activation function ($\sigma^{(l)}$)	id, ReLU, Sigmoid[6], tanh, ELU (Clevert et al., 2016), Leaky ReLU (Maas et al., 2013)
Optimizer	SGD, Adam (Kingma and Ba, 2015), RMSProp[7]
Learning rate	$0.001, 0.005, 0.075, 0.01, 0.05, 0.075, 0.1$
Loss function ($\mathcal{L}$)	MSE, MAE, Huber, LogCosh[8]

Table B3. Training range and test range of the features during the interpolation split.

Feature	Range	Test range	Training range
D_f	$[1.5, 2.9]$	$[2.1, 2.5]$	$[1.5, 2.1) \cup (2.5, 2.9]$
f_{coating}	$[0, 90]$	$[35, 50]$	$[0, 35) \cup (50, 90]$

Table B4. Training range and test range of the features during the extrapolation split.

Feature	Range	Test range	Training range
D_f	$[1.5, 2.9]$	$[2.5, 2.9]$	$[1.5, 2.5)$
D_f	$[1.5, 2.9]$	$[1.5, 1.9]$	$(1.9, 2.9]$
f_{coating}	$[0, 90]$	$[75, 90]$	$[0, 75)$
f_{coating}	$[0, 90]$	$[0, 15]$	$(15, 90]$

Table B5. Maximum errors of different splits for their test sets.

Optical property	Random split		Interpolation split		Extrapolation split		Feature range
	KRR	ANN	KRR	ANN	KRR	ANN	
Q_{abs}	0.17	0.34	0.38	0.34	0.23	0.21	$0-2$
Q_{sca}	0.14	0.17	0.32	0.44	0.55	1.42	$0-2$
g	0.14	0.22	0.46	0.44	0.42	0.32	$0-1$

Appendix C: Additional figures

C1   Error Boxplots

Fig. C1 shows the residuals for the machine-learning methods for the three splits related to the feature f_{coating}: random, extrapolation (training data $f_{\text{coating}} = [0, 75)$), and interpolation (training data: $f_{\text{coating}} = [0, 35] \cup (50, 90]$. When the training and testing data is randomly split, we see that residual errors are concentrated near zero for all intervals of f_{coating} similar to Fig. 3. The errors from the KRR and ANN are comparable in the random split. For the case of interpolation split, the errors from both ANN and KRR models are comparatively higher for all the three optical properties, i.e., Q_{abs}, Q_{abs}, and g. It was noted in the errors from the interpolation split that KRR performs better for predicting the Q_{abs}, whereas the ANN performs better for g predictions. The errors in the Q_{abs}, Q_{abs}, and g from the extrapolation split were the highest. The error is largest for the predictions when $f_{\text{coating}} = 90$, which is the case farthest away from the training data during an extrapolation split. The relative performance of the ANN and KRR are comparable to those observed in the interpolation split.

C2   Point-wise comparison of predicted and true values

Fig. C3 and Fig. C4 compare the machine-learning predictions to their true values for the cases where the data was excluded while training the ML model. In the Fig. C3, ML predictions were made after removing the intermediate values of the D_f feature (i.e., 2.1, 2.3, 2.5) from the training data. It was observed that the predictions $\hat{Q}_{abs}$ fitted well with the true values, especially for the KRR method. However, the predictions $\hat{Q}_{sca}$ fluctuate from the true value Q_{sca} as the approach maximum values above 1. For predictions $\hat{g}$, the ML methods ANN and KRR perform slightly differently. In the case of extrapolation split, as shown in Fig. C4, the predictions deviated from their true values for $D_f = 2.7, 2.9$ since the ML models did not see the data. However, we can see that for $D_f = 2.5$ (first row), all the predictions are in better agreement with their true values since it was present in the training data. The predictions $\hat{Q}_{abs}$ and $\hat{Q}_{sca}$ showed reasonable agreement in the case of $D_f = 2.7$. The predictions $\hat{Q}_{sca}$ for the unseen D_f features were observed to be smaller than their true values. The predictions $\hat{Q}_{abs}$, $\hat{Q}_{sca}$, $\hat{g}$ are most inconsistent with their true values when the $D_f = 2.9$ which is the case farthest away from the training data. Therefore, it is demonstrated that there
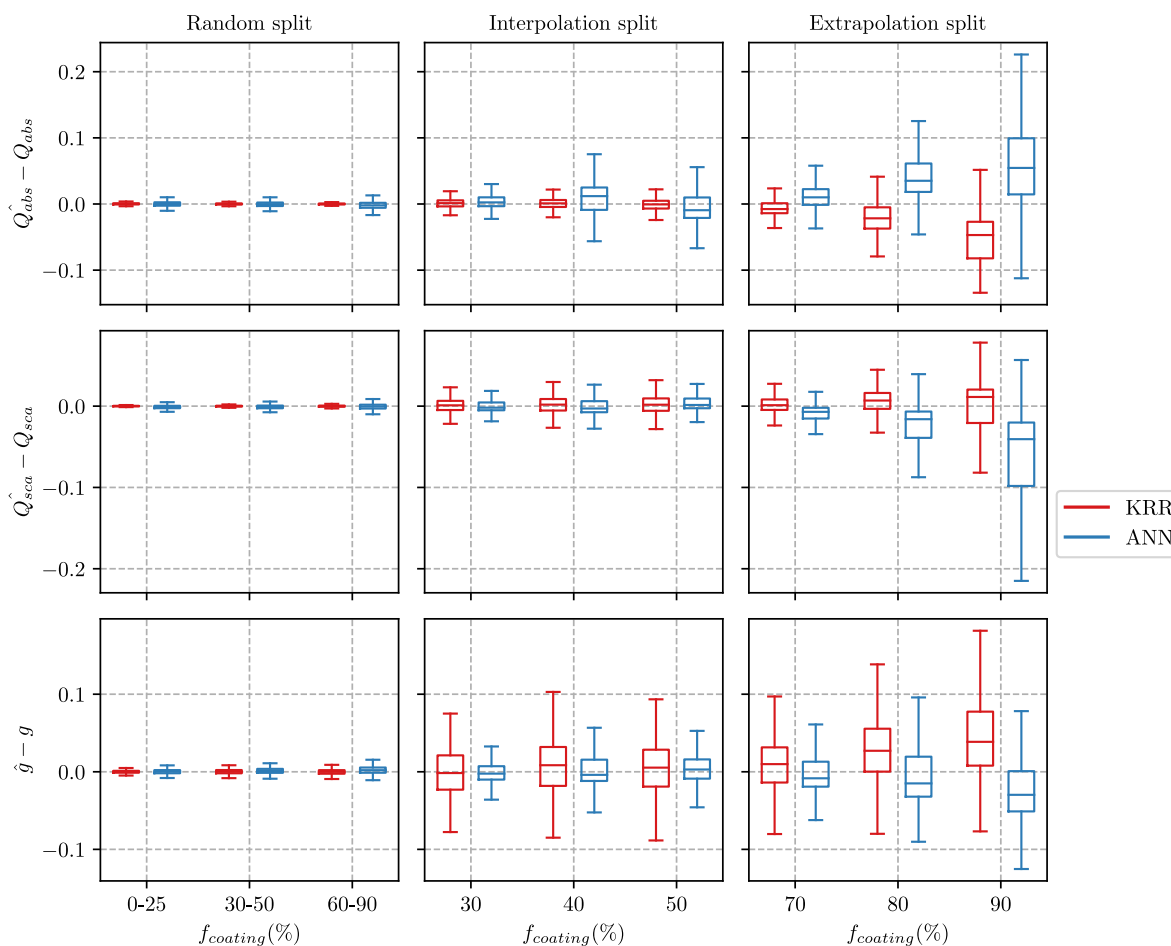
Figure C1. Error between the predicted and true values for three optical properties. The residuals are shown when models are trained on data with different ranges of fractions of coating ($f_{coating}$). The residuals for both KRR and NN predictions are presented in each panel.

is comparatively higher uncertainty for predicting optical properties for features out of the range of the training data. Further, the performance of the KRR and ANN varied for different optical properties in such cases of interpolation and extrapolation split. The interpolation split performed better for predicting the optical properties out of the range of the training data. Therefore, adding more data in the training set for boundary values to let it interpolate would result in better predictions.

C3 Line plots showing performance as aggregate size changes

Fig. C5 compares the machine-learning predictions to their true values for interpolation split. The predictions for the case $D_f = 2.3$ (middle row) showed the highest deviations from the true values since it is the farthest point in the training
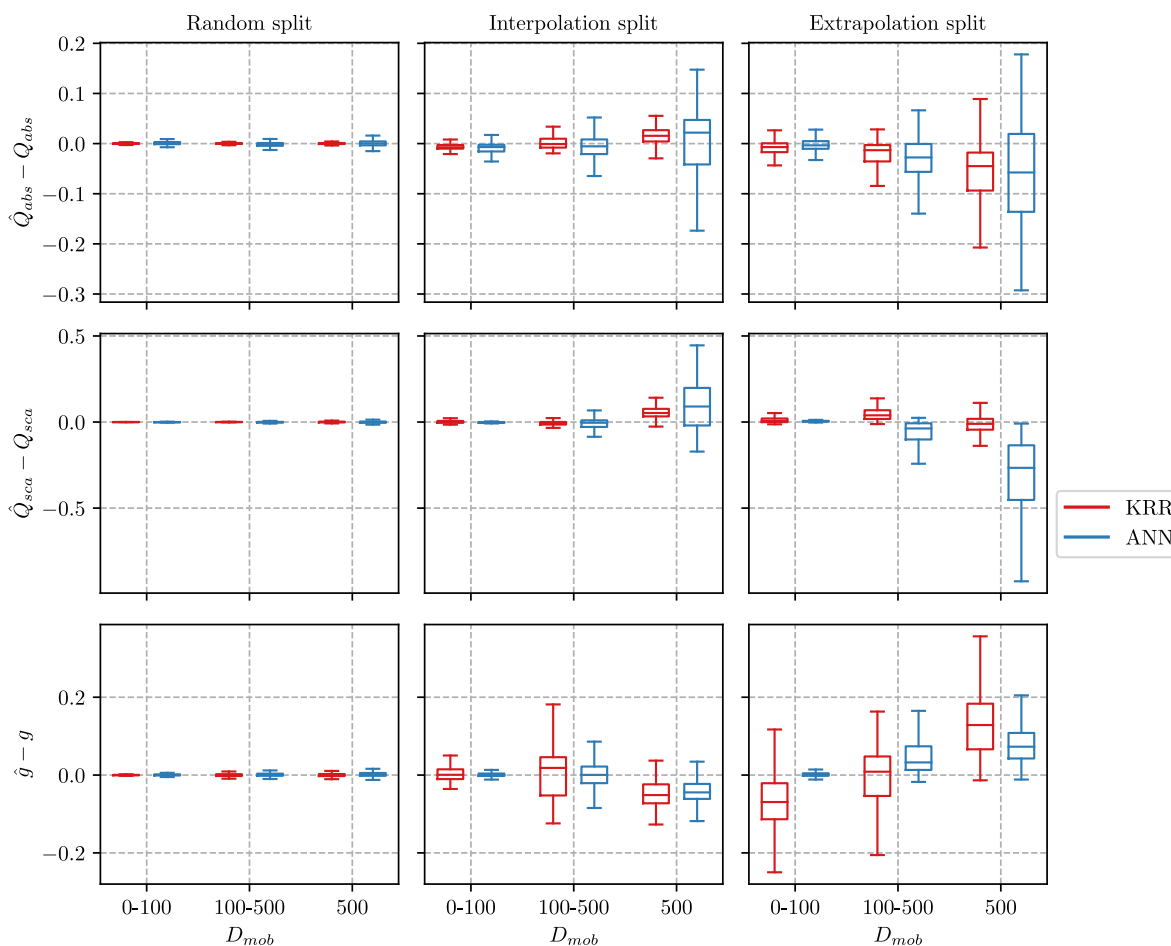
Figure C2. Error between the predicted value ($\hat{Q}_{abs}, \hat{Q}_{sca}, \hat{g}$) and the true value for three optical properties for various cases of mobility diameter (D_{mob}). Note that we do not show outliers to prevent unnecessary clutter.

data for the interpolation split. From the $\hat{Q}_{abs}$ results, the KRR predictions were reasonable for the entire size range. The predictions for $\hat{Q}_{sca}$ were also reasonable for KRR. However, after the particle size increased to larger than 500 nm, the prediction of $\hat{Q}_{sca}$ using KRR was underpredicted. The prediction of $\hat{Q}_{sca}$ using ANN showed a size-dependent behavior, under-predicting the results for certain particle sizes, after which there is an over-prediction. Similar size-dependent behavior was observed in the predictions $\hat{g}$ from ANN and KRR. The $\hat{g}$ predictions showed deviations from their true values as the particle size increased. In the case of interpolation split, the overfitting or underfitting is generally more pronounced in the larger particle size (> 500 nm). The explanation for this could be the lower resolution of the training data for particle size > 500 nm, which was a limitation of large computation time for larger particles and more coating fraction.
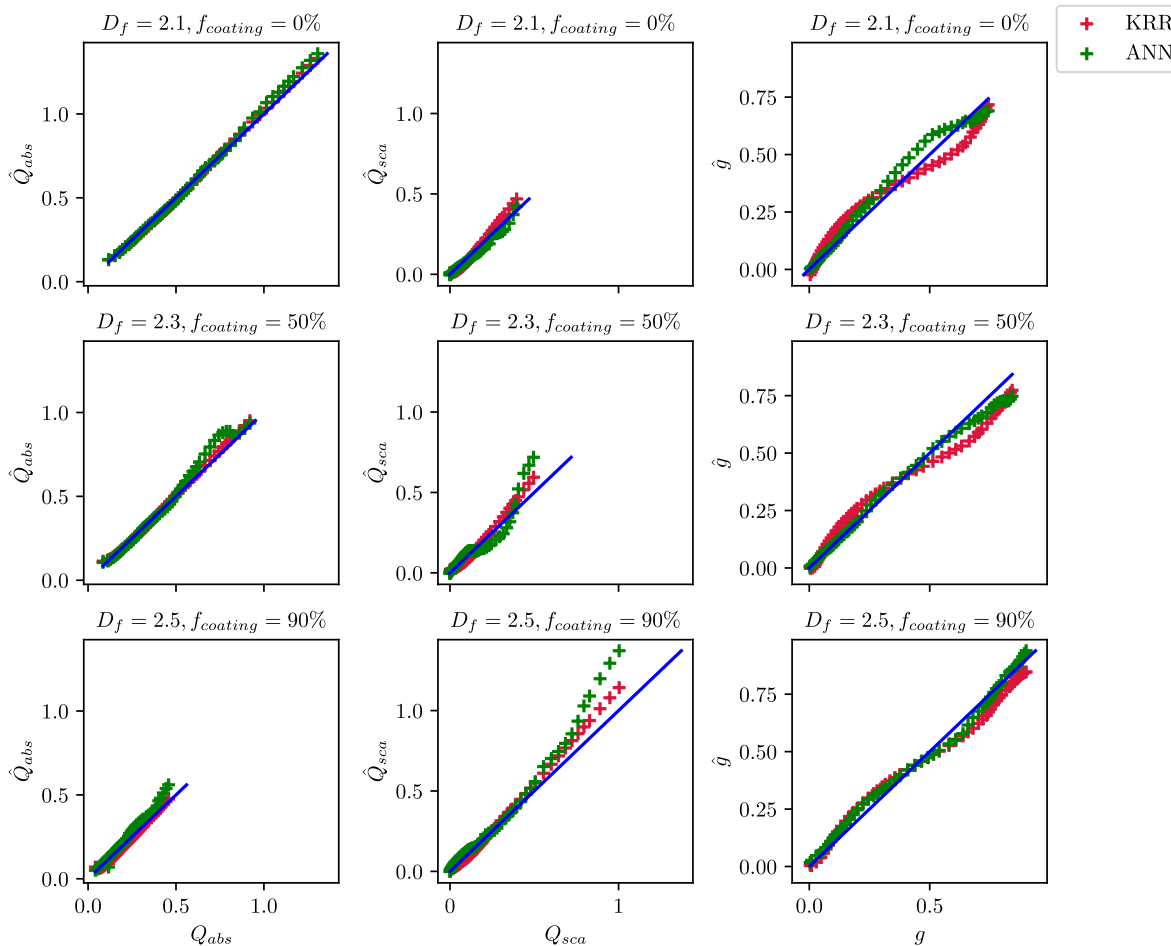
Figure C3. Comparison of the predicted optical properties with their true values for the interpolation split when the ML models have trained on data with boundary fractal dimensions: $D_f = 1.5, 1.7, 1.9, 2.7, 2.9$ and tested the model on data with inner fractal dimensions: $D_f = 2.1, 2.3, 2.5$).

Similarly, Fig. C6 show the machine-learning predictions compared to the true values for the extrapolation split. To
475 study the performance of the KRR and ANN, the results for $D_f = 2.9$ are interesting since they are the farthest from the training data. The deviations of the $\hat{Q}_{abs}$ are more from the true values in the case of the KRR, which showed better performance in the interpolation-split. However, the results for $D_f = 2.5$ and $D_f = 2.7$ show reasonable results since they are closer to the training data set. The predictions $\hat{Q}_{sca}$ were lower than the true values for ANN, especially as the particle size increased. The prediction $\hat{g}$ was larger than its true value in the case of extrapolation-split. However, the
480 performance of predicting $\hat{g}$ from KRR showed an interesting size dependence over particle size unique to this split. When particle sizes were smaller, $\hat{g}$ was higher than the true value, decreased, and returned to higher levels once a
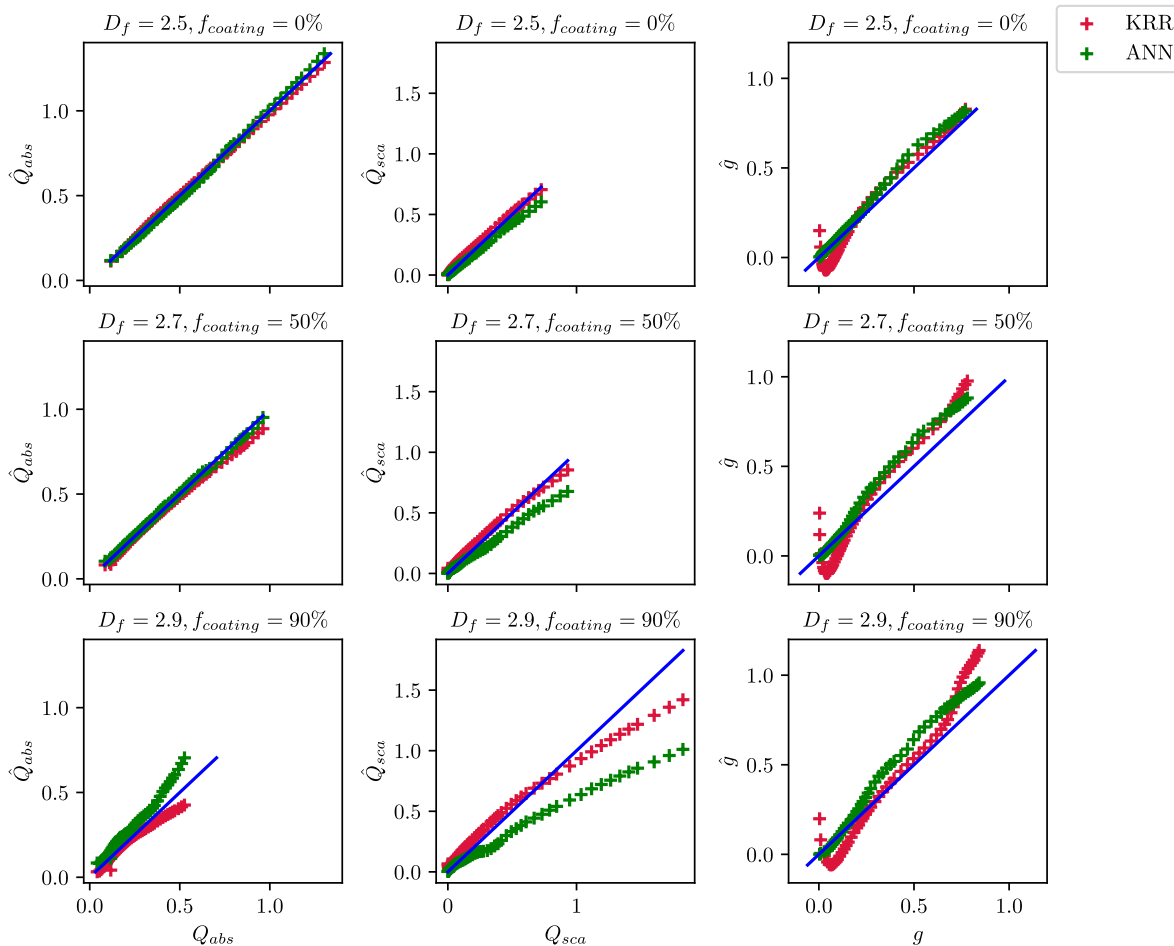
Figure C4. Comparison of the predicted optical properties with their true values for extrapolation split when the ML models are trained on data with smaller fractal dimensions: $D_f = 1.5, 1.7, 1.9, 2.1, 2.3$ and tested the model on data with boundary fractal dimensions: $D_f = 2.5, 2.7, 2.9$).

certain threshold was reached. In general, for the results when the $f_{coating}$ is 90, which is the upper limit of the feature, the results for $\hat{Q}_{abs}$, $\hat{Q}_{sca}$, and $\hat{g}$ showed an expected higher deviation from their true values for both interpolation and extrapolation split.
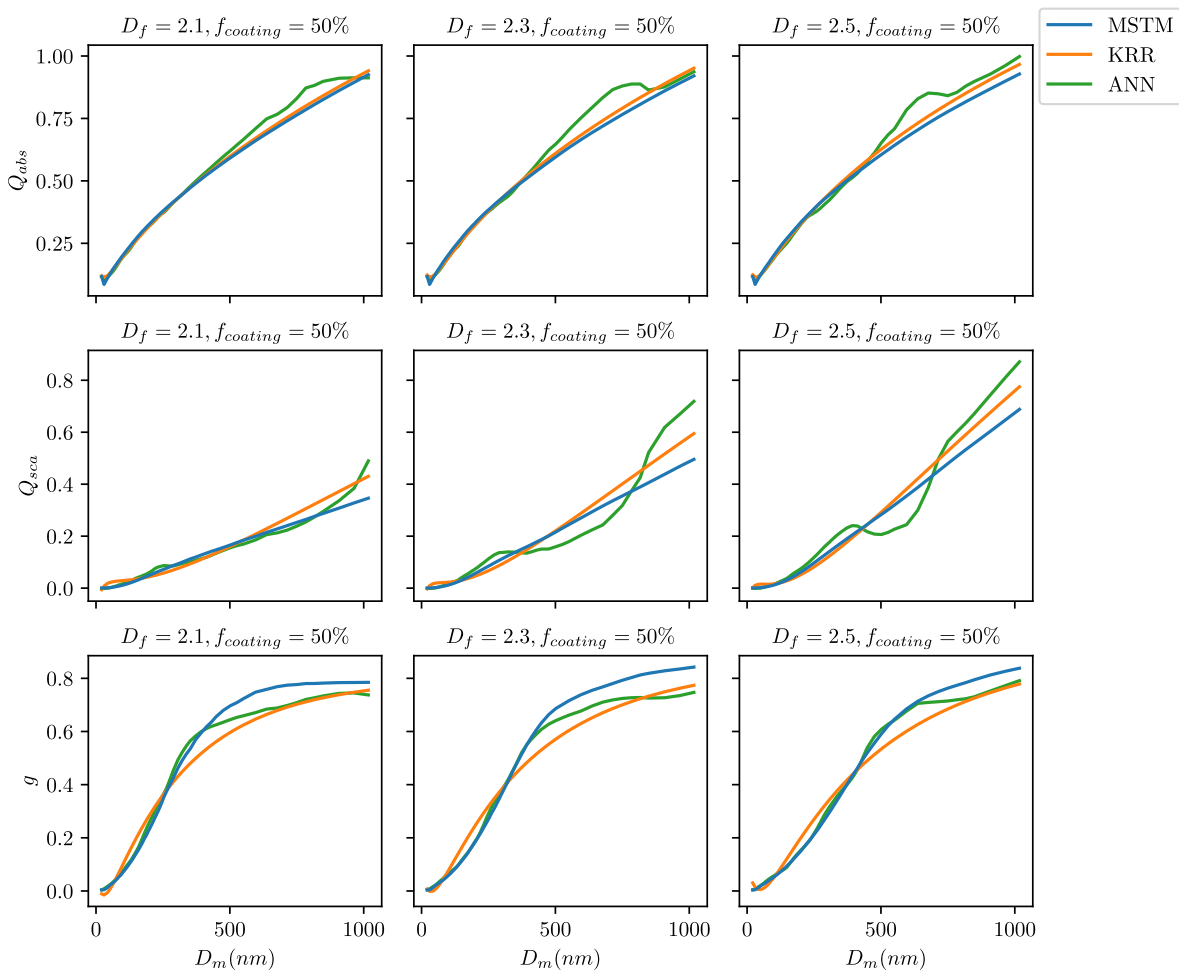
Figure C5. Optical properties of BC fractal aggregates predicted using machine learning methods KRR and NN for the interpolation split: when models have trained on data with boundary fractal dimensions $D_f = 1.5, 1.7, 1.9, 2.7, 2.9$ and tested the model if it fits for the intermediate values of fractal dimensions $D_f = 2.1, 2.3, 2.5$).The three columns show the predicted values of absorption efficiency(Q_{abs}), scattering efficiency(Q_{sca}), and asymmetry parameter(g). Each row corresponds to the predictions for the intermediate values of fractal dimensions $D_f = 2.1, 2.3, 2.5$.
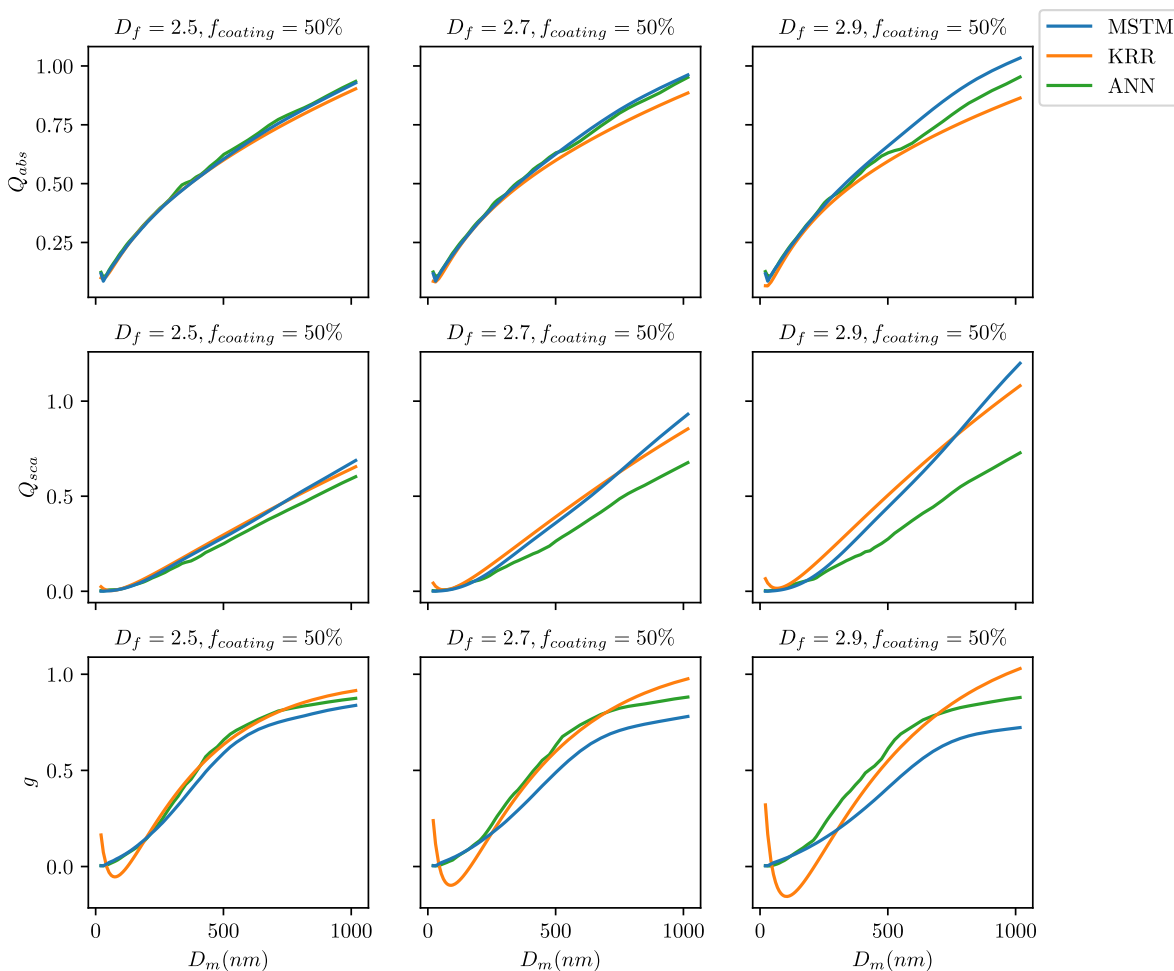
Figure C6. Optical properties of BC fractal aggregates predicted using machine learning methods KRR and NN for the extrapolation split: when models are trained on data with smaller fractal dimensions $D_f = 1.5, 1.7, 1.9, 2.1, 2.3$ and tested the model on data with higher fractal dimensions $D_f = 2.5, 2.7, 2.9$).The three columns show the predicted values of absorption efficiency(Q_{abs}), scattering efficiency(Q_{sca}), and asymmetry parameter(g). Each row corresponds to the predictions for the left-out higher fractal dimensions $D_f = 2.5, 2.7, 2.9$.
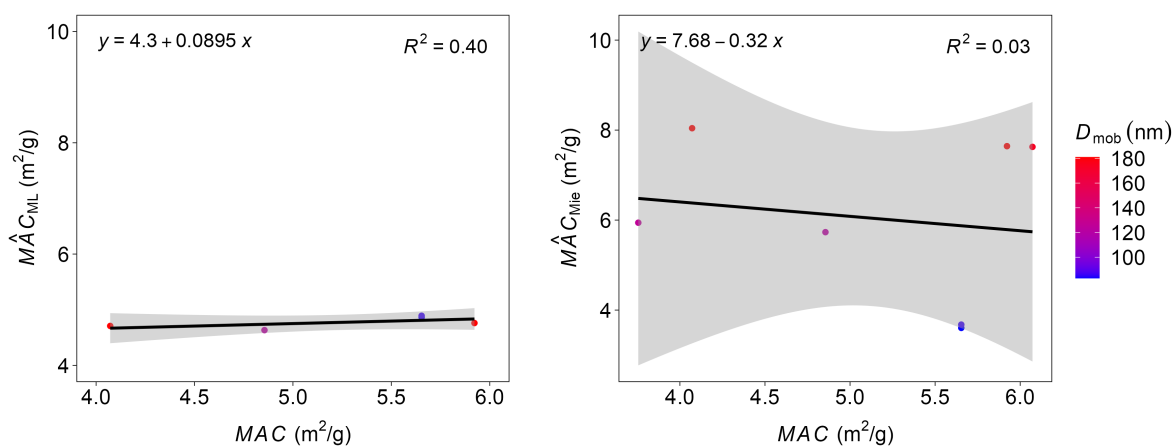
Figure C7. Mass absorption cross-section *(MAC)* for coated BC particles generated in a laboratory study at different D_{mob} (Romshoo et al., 2022). **(A)** compares the $\hat{MAC}_{ML}$ with the measured *MAC* from the laboratory experiment. **(B)** compares the $\hat{MAC}_{Mie}$ with the measured *MAC*.

485   *Author contributions.*   The study was designed by BR, ThM, JP, ToM, MP, and MK. BR and ThM developed the optical simulations and database. The machine learning experiments were conducted by JP and ToM, with help from BR and ThM. The results were prepared by JP and ToM, with help from BR. The paper was written by BR, JP, and ToM. The paper was reviewed, commented on, and edited by ThM, MK, and MP.

Competing interests.   The authors declare that none of the authors have any competing interests.

490   *Acknowledgements.*   This research has been supported by the "Metrology for light absorption by atmospheric aerosols" project funded by the European Metrology Programme for Innovation and Research (EMPIR, grant no. 16ENV02 Black Carbon). We would like to thank the members of the European Metrology Programme for Innovation and Research EMPIR 16ENV02 Black Carbon project for their support and feedback. MK acknowledges support by the Carl-Zeiss Foundation, the DFG awards KL 2698/2-1, KL 2698/5-1, KL 2698/6-1, and KL 2698/7-1, and the BMBF awards 03|B0770E and 01|S21010C.

References

Adachi, K., Chung, S. H., and Buseck, P. R.: Shapes of soot aerosol particles and implications for their effects on climate, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD012868, 2010.

Bambha, R. P., Dansson, M. A., Schrader, P. E., and Michelsen, H. A.: Effects of volatile coatings on the laser-induced incandescence of soot, Applied Physics B, 112, 343–358, https://doi.org/10.1007/s00340-013-5463-9, 2013.

Bescond, A., Yon, J., Ouf, F. X., Ferry, D., Delhaye, D., Gaffié, D., Coppalle, A., and Rozé, C.: Automated Determination of Aggregate Primary Particle Size Distribution by TEM Image Analysis: Application to Soot, Aerosol Science and Technology, 48, 831–841, https://doi.org/10.1080/02786826.2014.932896, 2014.

Bhandari, J., China, S., Chandrakar, K. K., Kinney, G., Cantrell, W., Shaw, R. A., Mazzoleni, L. R., Girotto, G., Sharma, N., Gorkowski, K., Gilardoni, S., Decesari, S., Facchini, M. C., Zanca, N., Pavese, G., Esposito, F., Dubey, M. K., Aiken, A. C., Chakrabarty, R. K., Moosmüller, H., Onasch, T. B., Zaveri, R. A., Scarnato, B. V., Fialho, P., and Mazzoleni, C.: Extensive Soot Compaction by Cloud Processing from Laboratory and Field Observations, Scientific Reports, 9, 11 824, https://doi.org/10.1038/s41598-019-48143-y, 2019.

Bohren, C. F. and Huffman, D. R.: Absorption and scattering of light by small particles, John Wiley & Sons, 2008.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research: Atmospheres, 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, 2013.

Bottou, L., Curtis, F. E., and Nocedal, J.: Optimization Methods for Large-Scale Machine Learning, SIAM Review, 60, 223–311, https://doi.org/10.1137/16M1080173, 2018.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, Journal of the Royal Statistical Society. Series B (Methodological), 26, 211–252, http://www.jstor.org/stable/2984418, 1964.

Chakrabarty, R. K., Moosmüller, H., Garro, M. A., Arnott, W. P., Walker, J., Susott, R. A., Babbitt, R. E., Wold, C. E., Lincoln, E. N., and Hao, W. M.: Emissions from the laboratory combustion of wildland fuels: Particle morphology and size, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006659, 2006.

Clevert, D., Unterthiner, T., and Hochreiter, S.: Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), in: 4th International Conference on Learning Representations, San Juan, Puerto Rico, May 2–4, 2016, edited by Bengio, Y. and LeCun, Y., http://arxiv.org/abs/1511.07289, 2016.

Cortes, C. and Vapnik, V.: Support-vector networks, Machine Learning, 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Coz, E. and Leck, C.: Morphology and state of mixture of atmospheric soot aggregates during the winter season over Southern Asia-a quantitative approach, Tellus B: Chemical and Physical Meteorology, 63, 107–116, https://doi.org/10.1111/j.1600-0889.2010.00513.x, 2011.

Dong, Z., Kang, S., Qin, D., Shao, Y., Ulbrich, S., and Qin, X.: Variability in individual particle structure and mixing states between the glacier–snowpack and atmosphere in the northeastern Tibetan Plateau, The Cryosphere, 12, 3877–3890, https://doi.org/10.5194/tc-12-3877-2018, 2018.

Flagan, R.: Electrical techniques, Aerosol Measurement: Principles, Techniques, and Applications, pp. 537–568, 2001.

Gentner, D. R., Jathar, S. H., Gordon, T. D., Bahreini, R., Day, D. A., El Haddad, I., Hayes, P. L., Pieber, S. M., Platt, S. M., de Gouw, J., Goldstein, A. H., Harley, R. A., Jimenez, J. L., Prévôt, A. S. H., and Robinson, A. L.: Review of Urban Secondary Organic Aerosol Formation from Gasoline and Diesel Motor Vehicle Emissions, Environmental Science & Technology, 51, 1074–1093, https://doi.org/10.1021/acs.est.6b04509, pMID: 28000440, 2017.

Guarieiro, A. L. N., Eiguren-Fernandez, A., da Rocha, G. O., and de Andrade, J. B.: An investigation on morphology and fractal dimension of diesel and diesel-biodiesel soot agglomerates, Journal of the Brazilian Chemical Society, 28, 1351–1362, 2017.

Huber, P. J.: Robust Estimation of a Location Parameter, The Annals of Mathematical Statistics, 35, 73 – 101, https://doi.org/10.1214/aoms/1177703732, 1964.

Jacobson, M. Z.: Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols, Nature, 409, 695–697, https://doi.org/10.1038/35055518, 2001.

Kahnert, M.: On the Discrepancy between Modeled and Measured Mass Absorption Cross Sections of Light Absorbing Carbon Aerosols, Aerosol Science and Technology, 44, 453–460, https://doi.org/10.1080/02786821003733834, 2010a.

Kahnert, M.: Numerically exact computation of the optical properties of light absorbing carbon aggregates for wavelength of 200 nm–12.2 μm, Atmospheric Chemistry and Physics, 10, 8319–8329, https://doi.org/10.5194/acp-10-8319-2010, 2010b.

Kahnert, M. and Kanngießer, F.: Modelling optical properties of atmospheric black carbon aerosols, Journal of Quantitative Spectroscopy and Radiative Transfer, 244, 106 849, https://doi.org/10.1016/j.jqsrt.2020.106849, 2020.

Kim, J., Bauer, H., Dobovičnik, T., Hitzenberger, R., Lottin, D., Ferry, D., and Petzold, A.: Assessing Optical Properties and Refractive Index of Combustion Aerosol Particles Through Combined Experimental and Modeling Studies, Aerosol Science and Technology, 49, 340–350, https://doi.org/10.1080/02786826.2015.1020996, 2015.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in: 3rd International Conference on Learning Representations, San Diego, CA, USA, May 7–9, edited by Bengio, Y. and LeCun, Y., http://arxiv.org/abs/1412.6980, 2015.

Lamb, K. D. and Gentine, P.: Zero-shot learning of aerosol optical properties with graph neural networks, arXiv preprint arXiv:2107.10197, 2021.

Liati, A., Brem, B. T., Durdina, L., Vögtli, M., Arroyo Rojas Dasilva, Y., Dimopoulos Eggenschwiler, P., and Wang, J.: Electron Microscopic Study of Soot Particulate Matter Emissions from Aircraft Turbine Engines, Environmental Science & Technology, 48, 10 975–10 983, https://doi.org/10.1021/es501809b, pMID: 25180674, 2014.

Liu, C., Yin, Y., Hu, F., Jin, H., and Sorensen, C. M.: The Effects of Monomer Size Distribution on the Radiative Properties of Black Carbon Aggregates, Aerosol Science and Technology, 49, 928–940, https://doi.org/10.1080/02786826.2015.1085953, 2015.

Liu, C., Xu, X., Yin, Y., Schnaiter, M., and Yung, Y. L.: Black carbon aggregates: A database for optical properties, Journal of Quantitative Spectroscopy and Radiative Transfer, 222–223, 170–179, https://doi.org/10.1016/j.jqsrt.2018.10.021, 2019.

Liu, L. and Mishchenko, M. I.: Scattering and Radiative Properties of Morphologically Complex Carbonaceous Aerosols: A Systematic Modeling Study, Remote Sensing, 10, https://doi.org/10.3390/rs10101634, 2018.

Luo, J., Zhang, Y., Wang, F., Wang, J., and Zhang, Q.: Applying machine learning to estimate the optical properties of black carbon fractal aggregates, Journal of Quantitative Spectroscopy and Radiative Transfer, 215, 1–8, https://doi.org/10.1016/j.jqsrt.2018.05.002, 2018.

Maas, A. L., Hannun, A. Y., and Ng, A. Y.: Rectifier nonlinearities improve neural network acoustic models, in: Proceedings of the 30th International Conference on Machine Learning, edited by Dasgupta, S. and McAllester, D., vol. 28 of *Proceedings of Machine Learning Research*, PMLR, Atlanta, Georgia, USA, 2013.

570   Mackowski, D. W.: A multiple sphere T-matrix Fortran code for use on parallel computer clusters, Version 3.0, https://www.eng. auburn.edu/~dmckwski/scatcodes/, [Accessed: 05.07.2022], 2013.

Mackowski, D. W. and Mishchenko, M. I.: A multiple sphere T-matrix Fortran code for use on parallel computer clusters, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 2182–2192, https://doi.org/10.1016/j.jqsrt.2011.02.019, polarimetric Detection, Characterization, and Remote Sensing, 2011.

575   Mie, G.: Considerations on the optics of turbid media, especially colloidal metal solutions, Ann. Phys, 25, 377–442, 1908.

Mishchenko, M. I., Travis, L. D., and Mackowski, D. W.: T-matrix computations of light scattering by nonspherical particles: A review, Journal of Quantitative Spectroscopy and Radiative Transfer, 55, 535–575, https://doi.org/10.1016/0022-4073(96)00002-7, 1996.

Mishchenko, M. I., Travis, L. D., and Lacis, A. A.: Scattering, absorption, and emission of light by small particles, Cambridge
580   university press, 2002.

Mishchenko, M. I., Liu, L., Travis, L. D., and Lacis, A. A.: Scattering and radiative properties of semi-external versus external mixtures of different aerosol types, Journal of Quantitative Spectroscopy and Radiative Transfer, 88, 139–147, https://doi.org/10.1016/j.jqsrt.2003.12.032, photopolarimetry in remote sensing, 2004.

Park, K., Kittelson, D. B., and McMurry, P. H.: Structural Properties of Diesel Exhaust Particles Measured by Transmission
585   Electron Microscopy (TEM): Relationships to Particle Mass and Mobility, Aerosol Science and Technology, 38, 881–889, https://doi.org/10.1080/027868290505189, 2004.

Purcell, E. M. and Pennypacker, C. R.: Scattering and absorption of light by nonspherical dielectric grains, The Astrophysical Journal, 186, 705–714, 1973.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and
590   Sutskever, I.: Learning Transferable Visual Models From Natural Language Supervision, in: Proceedings of the 38th International Conference on Machine Learning, edited by Meila, M. and Zhang, T., vol. 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, PMLR, https://proceedings.mlr.press/v139/radford21a.html, 2021.

Ramanathan, V. and Carmichael, G.: Global and regional climate changes due to black carbon, Nature Geoscience, 1, 221–227, https://doi.org/10.1038/ngeo156, 2008.

595   Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M.: Hierarchical text-conditional image generation with clip latents, arXiv preprint arXiv:2204.06125, 2022.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, The MIT Press, https://doi.org/10.7551/mitpress/3206.001.0001, 2005.

Romshoo, B., Müller, T., Pfeifer, S., Saturno, J., Nowak, A., Ciupek, K., Quincey, P., and Wiedensohler, A.: Optical properties of
600   coated black carbon aggregates: numerical simulations, radiative forcing estimates, and size-resolved parameterization scheme, Atmospheric Chemistry and Physics, 21, 12 989–13 010, https://doi.org/10.5194/acp-21-12989-2021, 2021.

Romshoo, B., Pöhlker, M., Wiedensohler, A., Pfeifer, S., Saturno, J., Nowak, A., Ciupek, K., Quincey, P., Vasilatou, K., Ess, M., Gini, M., Eleftheriadis, K., Gaie-Levrel, F., and Müller, T.: Importance of size representation and morphology in modelling optical properties of black carbon: comparison between laboratory measurements and model simulations, EGUsphere, 2022, 1–26,
605   https://doi.org/10.5194/egusphere-2022-573, 2022.

Romshoo, B., Patil, J., Michels, T., Müller, T., Kloft, M., and Pöhlker, M.: Database of physicochemical and optical properties of black carbon fractal aggregates, Dataset, https://doi.org/10.5281/zenodo.7523058, 2023a.

Romshoo, B., Patil, J., Michels, T., Müller, T., Kloft, M., and Pöhlker, M.: jaikrishnap/Optical-properties-of-black-carbon- aggregates: Initial release, https://doi.org/10.5281/zenodo.8071901, 2023b.

Romshoo, B., Patil, J., Michels, T., Müller, T., Kloft, M., and Pöhlker, M.: jaikrishnap/Machine-learning-for-prediction-of- BCFAs: Initial release, https://doi.org/10.5281/zenodo.8060206, 2023c.

Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain, Psychological Review, 65, 386–408, https://doi.org/10.1037/h0042519, 1958.

Schkolnik, G., Chand, D., Hoffer, A., Andreae, M., Erlick, C., Swietlicki, E., and Rudich, Y.: Constraining the density and complex refractive index of elemental and organic carbon in biomass burning aerosol using optical and chemical measurements, Atmospheric Environment, 41, 1107–1118, https://doi.org/10.1016/j.atmosenv.2006.09.035, 2007.

Shawe-Taylor, J. and Cristianini, N.: Kernel Methods for Pattern Analysis, Cambridge University Press, https://doi.org/10.1017/CBO9780511809682, 2004.

Smith, A. J. A. and Grainger, R. G.: Simplifying the calculation of light scattering properties for black carbon fractal aggregates, Atmospheric Chemistry and Physics, 14, 7825–7836, https://doi.org/10.5194/acp-14-7825-2014, 2014.

Sorensen, C. M.: Light Scattering by Fractal Aggregates: A Review, Aerosol Science and Technology, 35, 648–687, https://doi.org/10.1080/027868201117868, 2001.

Sun, J., Sun, Y., Xie, C., Xu, W., Chen, C., Wang, Z., Li, L., Du, X., Huang, F., Li, Y., Li, Z., Pan, X., Ma, N., Xu, W., Fu, P., and Wang, Z.: The chemical composition and mixing state of BC-containing particles and the implications on light absorption enhancement, Atmospheric Chemistry and Physics, 22, 7619–7630, https://doi.org/10.5194/acp-22-7619-2022, 2022.

Wang, Y., Liu, F., He, C., Bi, L., Cheng, T., Wang, Z., Zhang, H., Zhang, X., Shi, Z., and Li, W.: Fractal Dimensions and Mixing Structures of Soot Particles during Atmospheric Processing, Environmental Science & Technology Letters, 4, 487–493, https://doi.org/10.1021/acs.estlett.7b00418, 2017.

Wentzel, M., Gorzawski, H., Naumann, K.-H., Saathoff, H., and Weinbruch, S.: Transmission electron microscopical and aerosol dynamical characterization of soot aerosols, Journal of Aerosol Science, 34, 1347–1370, https://doi.org/10.1016/S0021-8502(03)00360-4, intercomparison of Soot Measurement Techniques, 2003.

Wozniak, M., Onofri, F., Barbosa, S., Yon, J., and Mroczka, J.: Comparison of methods to derive morphological parameters of multi-fractal samples of particle aggregates from TEM images, Journal of Aerosol Science, 47, 12–26, https://doi.org/10.1016/j.jaerosci.2011.12.008, 2012.

Wu, Y., Cheng, T., Liu, D., Allan, J. D., Zheng, L., and Chen, H.: Light Absorption Enhancement of Black Carbon Aerosol Constrained by Particle Morphology, Environmental Science & Technology, 52, 6912–6919, https://doi.org/10.1021/acs.est.8b00636, pMID: 29783837, 2018.

Xu, Y.-l. and Gustafson, B. r.: A generalized multiparticle Mie-solution: further experimental verification, Journal of Quantitative Spectroscopy and Radiative Transfer, 70, 395–419, https://doi.org/10.1016/S0022-4073(01)00019-X, light Scattering by Non-Spherical Particles, 2001.