



RESEARCH ARTICLE

10.1029/2023MS003889

Key Points:

- Machine learning model trained on particle-resolved simulations has been coupled to CAM6 MAM4 to improve BC mixing state representation
- The BC mixing state index is reduced by 19%, and the mass ratio of coating to BC is reduced by 52%, in better agreement with observations
- Due to the drop in hygroscopicity of BC-containing aerosol (9%), BC activation fraction decreased by 20%, and BC burden increased by 4%

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Wang,
Minghui.Wang@nju.edu.cn

Citation:







Shen, W., Wang, M., Riemer, N., Zheng, Z., Liu, Y., & Dong, X. (2024). Improving BC mixing state and CCN activity representation with machine learning in the community atmosphere model version 6 (CAM6). *Journal of Advances in Modeling Earth Systems*, 16, e2023MS003889. <https://doi.org/10.1029/2023MS003889>

Received 17 JUN 2023

Accepted 19 NOV 2023

© 2024 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Improving BC Mixing State and CCN Activity Representation With Machine Learning in the Community Atmosphere Model Version 6 (CAM6)

Wenxiang Shen^{1,2} , Minghui Wang^{1,2} , Nicole Riemer³ , Zhonghua Zheng⁴ , Yawen Liu^{1,2} , and Xinyi Dong^{1,2,5} 

¹School of Atmospheric Sciences, Nanjing University, Nanjing, China, ²Joint International Research Laboratory of Atmospheric and Earth System Sciences and Institute for Climate and Global Change Research, Nanjing University, Nanjing, China, ³Department of Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA, ⁴Department of Earth and Environmental Sciences, The University of Manchester, Manchester, UK, ⁵Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing, China

Abstract Representing mixing state of black carbon (BC) is challenging for global climate models (GCMs). The Community Atmosphere Model version 6 (CAM6) with the four-mode version of the Modal Aerosol Module (MAM4) represents aerosols as fully internal mixtures with uniform composition within each aerosol mode, resulting in high degree of internal mixing of BC with non-BC species and large mass ratio of coating to BC (R_{BC} , the mass ratio of non-BC species to BC in BC-containing particles). To improve BC mixing state representation, we coupled a machine learning (ML) model of BC mixing state index trained on particle-resolved simulations to the CAM6 with MAM4 (MAM4-ML). In MAM4-ML, we use R_{BC} to partition accumulation mode particles into two new modes, BC-free particles and BC-containing particles. We adjust R_{BC} to make the modeled BC mixing state index (χ_{mode}) match the one predicted by the ML model (χ_{ML}). On a global average, the mass fraction of BC-containing particles in accumulation mode decreases from 100% (MAM4-default) to 48% (MAM4-ML). The globally averaged χ_{mode} decreases from 78% (MAM4-default) to 63% (MAM4-ML, 19% reduction) and agrees well with χ_{ML} (66%). The R_{BC} decreases by 52% for accumulation mode and better agrees with observations. The hygroscopicity drops by 9% for BC-containing particles in accumulation mode, leading to a 20% reduction in the BC activation fraction. The surface BC concentration increases most (6.9%) in the Arctic, and the BC burden increases by 4%, globally. Our study highlights the application of the ML model for improving key aerosol processes in GCMs.

Plain Language Summary Black carbon (BC) mixing state is an important property for assessing impacts of BC on human health and climate. It describes how BC is distributed among the particle population and within individual particles. Due to computational costs, global climate models (GCMs) consider aerosols as either external mixtures (e.g., bulk models) or fully internal mixtures with uniform composition within a certain size range (e.g., an aerosol bin in sectional models or an aerosol mode in modal models). To improve BC mixing state representation, we coupled a machine learning (ML) model learned from particle-resolved simulations to the online simulations of the Community Atmosphere Model version 6 (CAM6) with the four-mode version of the Modal Aerosol Module (MAM4-ML). We used the mass ratio of coating to BC (R_{BC}) to determine the mass fraction of BC-free particles and BC-containing particles, and we adjusted R_{BC} to make the modeled BC mixing state match ML model. The MAM4-ML model reduces the degree of internal mixture of BC substantially. In addition, MAM4-ML R_{BC} is reduced by 52%, leading to a 20% reduction in BC activation fraction and a 4% increase in BC burden. This study proposes a method for the application of the ML to improve GCMs.

1. Introduction

Black carbon (BC) has a significant impact on climate change (Bond et al., 2013). BC absorbs solar radiation, heats the atmosphere, and thereby has direct or semi-direct effects on the energy balance of the Earth system (Ramanathan & Carmichael, 2008). BC also affects cloud formation and properties by serving as cloud condensation nuclei (CCN) or ice nuclei (Andreae & Rosenfeld, 2008; Schill et al., 2020). Additionally, BC particles can enhance surface haze pollution and have adverse effects on air quality and human health (Ding et al., 2016; Grahame et al., 2014). Aerosol mixing state is the distribution of chemical composition within a single particle and among the population of particles that partially determines these effects (Riemer & West, 2013). A

large number of laboratory and field observations have documented that freshly emitted BC undergoes a series of physical and chemical processes in the atmosphere (known as BC aging processes) (Liu et al., 2017; Monge et al., 2010; Peng et al., 2016; Tritscher et al., 2011; Zhang et al., 2008), during which the mixing state of BC ranges between the two extremes of an external mixture and an internal mixture (Raatikainen et al., 2015; Yu et al., 2020; Zhao et al., 2021). As BC particles mix with other aerosol species and interact with gas-phase species, BC mixing state and related physicochemical properties, such as particle size distribution (Metcalf et al., 2012; Reddington et al., 2013), morphology (Guo et al., 2016; Slowik et al., 2007), chemical compositions (Lee et al., 2017a; Moffet et al., 2016; Pósfai et al., 1999), and the mass ratio of coating to BC (Cappa et al., 2012; Fierce et al., 2020; Healy et al., 2015), constantly evolve in the atmosphere. In addition, BC mixing state is an important property for effectively assessing climate-relevant properties of BC, such as hygroscopicity (Yuan & Zhao, 2023), CCN activity (Ching et al., 2012, 2017), and optical absorption (Fierce et al., 2016; Zhao et al., 2021).

BC mixing state should be represented in models, which in principle requires an accurate representation of particle-level chemical compositions (Riemer & West, 2013). However, due to computational costs, it is challenging to fully resolve BC mixing state in global climate models (GCMs). Most GCMs usually simplify aerosol representation by tracking some separate aerosol populations instead of individual particle compositions (Riemer et al., 2019). In GCMs, the commonly used aerosol models are bulk, sectional, and modal models, which are computationally efficient for long-term and large-scale simulations. For example, the early bulk models track the mass of individual chemical species with prescribed particle size distributions and assume different chemical species as external mixtures (Koch, 2001; Solomon et al., 2006). This simple assumption neglects the fact that BC is usually internally mixed with other species to varying extents due to aging processes (Bauer et al., 2013; Liu, He, et al., 2020). Univariate sectional models split the particle size distribution into several size bins (typically from 8 to 20 bins) and predict the mass of different chemical species in each size bin (Debry et al., 2007; Spracklen et al., 2005). Univariate sectional models assume that particles have the same composition within each size bin and hence cannot resolve the composition diversity across particles within the bin. To better address this issue, univariate sectional models can be expanded to multivariate bin structures with two or three dimensions by introducing several interacting univariate distributions, such as multi-distribution sectional model (Jacobson, 2002) and higher dimensional sectional model (Matsui et al., 2013; Zhu et al., 2016).

However, the computational cost of these high-dimensional sectional models increases exponentially with the number of dimensions. Modal models use several overlapping log-normal functions to represent the particle size distribution of different aerosol modes (typically 3 to 16 modes), based on which the mass of different chemical species and the total aerosol number of each mode are typically tracked (Bauer et al., 2008; Binkowski & Roselle, 2003; Liu et al., 2016; Stier et al., 2005). Different modes may overlap in a given size range, and hence modal models can capture different aerosol mixtures within the size range. The quality of mixing state resolution depends on the number, definition and placement of the modes (Riemer et al., 2019). However, some inherent assumptions of the modal models may also introduce uncertainties. For multi-modal models, rules need to be defined to handle the interactions of the various modes (e.g., the transfer between modes, and the destination mode after coagulation) in the parameterizations of aerosol processes of nucleation/condensation and coagulation (Sartelet et al., 2006). In addition, the diversity of the particle composition within a mode cannot be fully resolved as modal models assume all particles to be fully internal mixtures with uniform composition within each aerosol mode (referred to as “uniform mixing assumption”).

Zheng, West, et al. (2021) suggested that modal representation generally overestimates the extent of the mixing of BC with other species mainly due to the lack of BC-free particles (defined as particles without BC species) within the modes containing BC species. Recently, observations have shown that there is relatively high abundance of BC-free particles, especially over the remote areas at low and mid-latitudes and polar regions (Croft et al., 2016; Wentworth et al., 2016; Willis et al., 2016). Sources of BC-free particles include direct marine emissions, dust emissions, new particle formation from natural or anthropogenic precursors, and secondary aerosol condensation on pre-existing BC-free particles (Chen et al., 2022; Kodros et al., 2018; Raatikainen et al., 2015; Sharma et al., 2017; Zhang et al., 2017). In addition, the uncertainties of BC mixing state representation may further affect simulations of BC physicochemical properties, including the mass ratio of coating to BC, hygroscopicity, and CCN activity, which have not been adequately discussed in GCMs.

Currently, the most accurate simulations of BC mixing state are achieved through particle-resolved models that track the size and chemical composition of individual particles. For instance, Riemer et al. (2009) have developed

a stochastic particle-resolved box model (PartMC-MOSAIC), which resolves aerosol mixing state explicitly. But integrating particle-resolved models into large-scale three-dimensional simulations would be much more computationally expensive than using the traditional modal or sectional models. Therefore, given the dilemma of the need for particle-resolved properties and consideration of computational efficiency, machine learning (ML) techniques can provide a new approach for integrating particle-resolved models and large-scale GCMs. The use of machine learning in atmospheric sciences has been a new but fast-developing field, with recent applications in many areas, such as atmospheric chemistry (e.g., Kelp et al., 2020; Kelp et al., 2022; Sturm et al., 2023; Sturm & Wexler, 2020), aerosol microphysics (e.g., Yu et al., 2022), air quality (e.g., Ruan et al., 2023; Zheng et al., 2023), and remote-sensing (e.g., Grange et al., 2018).

Some recent studies have attempted to apply machine learning to diagnose aerosol mixing states in GCMs. For example, Hughes et al. (2018) trained a ML model to predict the aerosol mixing-state index, a metric to quantify aerosol mixing state (Riener & West, 2013), using a large ensemble of particle-resolved model simulations as the training data. The ML model by Hughes et al. (2018) takes variables of interest that GCMs already track as inputs and has been successfully applied to predict global maps of aerosol mixing-state index based on the GEOS-Chem-TOMAS model (a sectional model based GCM, the Goddard Earth Observing System chemical-transport model coupled with the TwO Moment Aerosol Sectional microphysics scheme). Then, Zheng, West, et al. (2021) and Zhang, Curtis, et al. (2021) extended the approach by Hughes et al. (2018) by defining the mixing state index in three different ways based on the CAM6 (Community Atmosphere Model version 6) model. The aerosol model of CAM6 is the four-mode version of the Modal Aerosol Module (MAM4) (Liu et al., 2016), which uses the modal method to represent the size distributions with four lognormal modes (i.e., Aitken, accumulation, coarse, and primary carbon modes). However, Hughes et al. (2018), Zheng, West, et al. (2021), and Zhang, Curtis, et al. (2021) only performed offline diagnosis of the mixing-state index based on instantaneous outputs of GCMs, and the ML technique for aerosol mixing state index estimates has not been incorporated to the online simulations of GCMs. Moreover, it has not been quantified how changes in mixing state representation may affect other aspects of model simulations, such as CCN activity.

In this paper, we couple a ML model of BC mixing index trained from particle-resolved simulations to the online simulations of the CAM6 with MAM4. The BC mixing index predicted by the ML model is then used to partition aerosol particles in the accumulation mode of the default MAM4 into two new modes, BC-free particles and BC-containing particles, which helps to improve the mixing state representation in MAM4. Based on the improved mixing state representation, we also updated the aerosol activation parameterization to quantify the impacts of changes in BC mixing states on CCN activity and BC concentrations. The paper is organized as follows. Section 2 introduces the definition of BC mixing state index, the CAM6 model coupled with the ML model, and the description of simulations. In Section 3, we document how the improved model simulates BC mixing state index, mass ratio of coating to BC (R_{BC}), BC hygroscopicity, activation fraction, and BC concentrations. Section 4 summarizes our findings.

2. Methods and Models

2.1. Description of CAM6

The Community Atmosphere Model version 6 with comprehensive chemistry (CAM6-Chem) (Emmons et al., 2020; Tilmes et al., 2019) is an atmospheric component of the Community Earth System Model version 2.1 (CESM2). The CAM6-Chem has 32 vertical layers up to about 40 km. The chemical scheme is the Model for Ozone and Related chemical Tracers covering the troposphere and stratosphere (MOZART-TS1, Emmons et al. (2020)). The MOZART-TS1 scheme uses a sophisticated SOA approach to better simulate the production of ozone and secondary organic aerosol (SOA) gas, which includes oxidation of biogenic and anthropogenic volatile organic compounds (VOCs) and the volatility basis set (VBS) scheme to simulate SOA gas formation (Donahue et al., 2006; Robinson et al., 2007). The aerosol model is MAM4 (Liu et al., 2016), and represents aerosol size distributions with four modes, including Aitken (containing dust, sea salt, SOA, and sulfate), accumulation (containing BC, dust, primary organic matter, sea salt, SOA, and sulfate), coarse (containing dust, sea salt, and sulfate), and primary carbon (containing BC and primary organic matter) modes. All fresh BC/primary organic matter (POM) of biomass burning and fossil fuel emissions are assigned to the hydrophobic primary carbon mode with the same initial volume-mean diameter (i.e., 134 nm), while the geometric standard deviation is different for different sectors of biomass burning and fossil fuel emissions (Dentener et al., 2006; Liu et al., 2005). The aging

of BC/POM in the primary carbon mode is explicitly treated. Specifically, the condensation of H_2SO_4 vapor and SOA gas and the coagulation of primary carbon mode with Aitken mode and of primary carbon mode with accumulation mode can convert fresh BC/POM in the primary carbon mode to the accumulation mode, which represents aged BC/POM. The mass concentration of different aerosol species and the number concentration of each mode are resolved explicitly, while the mode radius is diagnosed based on the log-normal distribution assumption with the fixed geometric standard deviation (i.e., 1.2 for coarse mode and 1.6 for all the Aitken, accumulation, and primary carbon modes). The hygroscopicity (κ) is 0, 0, 0.068, 0.14, 0.507, and 1.16 for BC, POM, dust, SOA, sulfate, and sea salt, respectively.

In the CAM6-Chem, the default anthropogenic and biogenic emissions for aerosols and gas-phase species are from the historical emission data set of CMIP6 (Hoesly et al., 2018; Van Marle et al., 2017), but we replace two anthropogenic aerosols species (BC, POM) and four gas-phase species (SO_2 , NO_x , CO, NH_3) with corresponding species of MEIC emission (Li et al., 2017; Zheng et al., 2018) over China region. This is because previous studies (Fan et al., 2022; Paulot et al., 2018) have indicated that CMIP6 emissions failed to capture the anthropogenic emission trends over China in recent years due to the lack of consideration of emission control measures implemented in China, which had started as early as 2006 in China (Liu & Wang, 2017). According to Shen et al. (2023), the CAM6 model with the MEIC emission produced a better agreement with the observed annual and seasonal trends of surface BC concentrations over China. Note that we multiply the organic carbon (OC) emission in MEIC by a factor of 1.4 to calculate the POM emission used in the model (Fan et al., 2018).

2.2. BC Mixing State Index Calculation

The aerosol mixing state index (χ , Riemer and West (2013)) is a quantitative metric for the mixing state of an aerosol population based on Shannon entropy, and is given by the affine ratio of average per-particle species diversity (D_a) and the bulk population species diversity (D_γ)

$$\chi = \frac{D_a - 1}{D_\gamma - 1}. \quad (1)$$

Here, D_a and D_γ are calculated based on the mixing entropies H_a and H_γ ,

$$D_a = e^{H_a}, \quad (2)$$

$$D_\gamma = e^{H_\gamma}. \quad (3)$$

Here, H_a measures the average per-particle mixing entropy for the population and is calculated as the mass-weighted average of the Shannon entropy of each particle, H_i . H_γ is the Shannon entropy of the species distribution within the bulk population. H_a , H_i , and H_γ are calculated as

$$H_a = \sum_{i=1}^N p_i H_i, \quad (4)$$

$$H_i = \sum_{a=1}^A -p_i^a \ln p_i^a, \quad (5)$$

$$H_\gamma = \sum_{a=1}^A -p^a \ln p^a. \quad (6)$$

Here, N is the number of particles, and A is the number of species. p_i is the mass fraction of particle i in the population. H_i represents the Shannon entropy of particle i . p_i^a is the mass fraction of species a in particle i . p^a is the mass fraction of species a in the population. The detailed definition and calculation process of χ can also be found in Riemer and West (2013).

In the modal model, the aerosol is considered internally mixed within each mode and externally mixed with respect to different modes. Based on modal model information, the H_a in Equation 4 can be calculated using Equations 7 and 8.

$$H_{\alpha} = \sum_{m=1}^M p_m H_m \quad (7)$$

$$H_m = \sum_{a=1}^A -p_m^a \ln p_m^a \quad (8)$$

Here, M is the number of aerosol mode types, and p_m is the mass fraction of mode m in the population. H_m represents the Shannon entropy of the species distribution within mode m . p_m^a is the mass fraction of species a in mode m . The detailed derivation of modal model based H_{α} in Equation 7 from particle-resolved H_{α} in Equation 4 can be found in Zheng, West, et al. (2021).

Note that the definition of species depends on the application of χ . For example, species can be defined as a single chemical species or groups of species with similar aerosol properties, such as hygroscopicity, absorption, or volatility. To focus on the mixing of BC mass with non-BC mass, we define the χ for BC (referred to as “BC mixing state index”) by assuming that each particle contains two types of species: BC and all other non-BC species. Note that the definition of χ in this study follows the mixing state index of optically absorbing and non-absorbing species (χ_o) in Zheng, West, et al. (2021) and Zhang, Curtis, et al. (2021). In MAM4, the non-BC species include dust, sea salt, POM, SOA, and sulfate. Because BC particles are usually found as submicron aerosol (PM1) due to their formation mechanism, we focus on the χ in the submicron aerosol population, which includes the Aitken, the accumulation, and the primary carbon modes in MAM4. Finally, the χ in the modal model framework (referred to as “ χ_{mode} ”) can be calculated by Equations 1–8. We add the calculation to the CAM6 source code to obtain χ_{mode} of the default MAM4 treatment. The values for χ_{mode} range from 0 (completely external mixture) to 1 (completely internal mixture), and a higher value of χ_{mode} indicates a more internal mixture of BC.

2.3. CAM6 Coupled With Machine Learning Model of BC Mixing State Index

2.3.1. Machine Learning Model of BC Mixing State Index

Zheng, West, et al. (2021) developed a ML model based on PartMC-MOSAIC simulations, which was able to predict χ (referred to as “ χ_{ML} ”) using the instantaneous outputs of CAM6 with MAM4. Here we briefly introduce the ML model (i.e., the model for χ_o in Zheng, West, et al. (2021)) and the application of the ML model in the offline CAM6 by Zheng, West, et al. (2021). By varying the input parameters for PartMC-MOSAIC (including meteorological parameters, emissions of aerosol and gas phase species, and the aerosol initial condition; 41 parameters in total), a large number of particle populations (45,000) were obtained for the training and testing of the ML model. Each aerosol population was characterized by 16 parameters (χ_{ML} and the other 15 variables as the input parameters for the training of χ_{ML}). The input parameters follow the two criteria, which make it possible to apply the ML model to CAM6 simulations. First, parameters are available within both the PartMC-MOSAIC and CAM6 simulations, so that we have consistent parameters for training, validation, and application processes. Second, the parameters are physically or chemically related to the aerosol mixing state, which enables the ML model to extract the mixing state index from the parameters. Specifically, the 15 input parameters included the aerosol bulk concentrations (BC, dust, sea salt, POM, SOA, and sulfate), gas mixing ratios (dimethyl sulfide, hydrogen peroxide, sulfuric acid, ozone, semi-volatile organic gas, and sulfur dioxide) and the environmental variables (air temperature, relative humidity, and solar zenith angle). The input parameters of non-precursor gas species, such as hydrogen peroxide and ozone, are important products of photochemistry in both the PartMC-MOSAIC and CAM6, which influence the generation of secondary aerosol precursors and the BC aging process. Figure 1 shows the schematic of the inputs and outputs for the ML model. Finally, based on the consistent input parameters obtained from CAM6 simulations, χ_{ML} can be predicted offline by the trained ML model. Table 1 shows the detailed calculations for the 15 input parameters in CAM6. The ML model is a tree-based ensemble model derived by the eXtreme Gradient Boosting (XGBoost; Chen and Guestrin (2016)), and the mean absolute error of the trained ML model is about 5% using the hold-out testing data. For more details on the PartMC-MOSAIC model simulations, the training and testing processes, and the offline application of the ML model, please refer to Zheng, West, et al. (2021).

In this study, we further integrate the ML model into the online simulations of CAM6. Our strategy is to train the ML model in Python and call the ML model from Fortran. Using the Python XGBoost software package

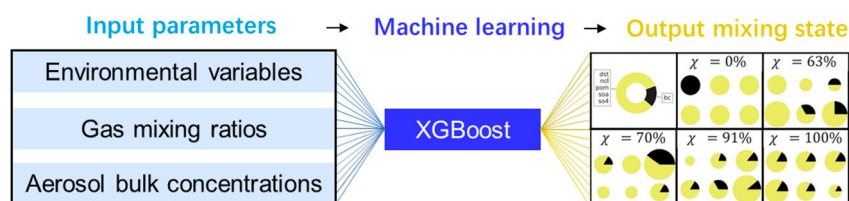


Figure 1. Schematic of the inputs and outputs for the machine learning model. The detailed variables for environmental variables, gas mixing ratios and aerosol bulk concentrations are shown in Table 1.

(Chen & Guestrin, 2016), we can train the ML model offline (i.e., the same as the model for χ_o in Zheng, West, et al. (2021)) and then save the trained model to disk in the XGBoost internal binary format. Then, by invoking the XGBoost C functions in the Fortran wrappers, we can read the saved binary file and call the trained ML model from Fortran. Finally, we embed the Fortran wrappers as a new module into the CAM6 source code. At the runtime of the CAM6, the module is called at every grid cell and timestep to read the binary file and predict χ_{ML} . Directly implementing the PartMC-MOSAIC in a complex GCM would slow down the model with the native scheme prohibitively (Curtis et al., 2017). The ML model emulates the input–output relationships (as shown in Figure 1) based on training data from particle-resolved simulations of PartMC-MOSAIC. Running CAM6 with the ML model online is 7%–8% slower than running CAM6 with the default MAM4 scheme.

Zheng, West, et al. (2021) have successfully applied the ML model to the CAM6 offline at a global scale and found that the diagnosed χ_{ML} from the ML model are similar in magnitude to the observations. When we incorporated the ML model into the source code of CAM6, we have confirmed that, with the same input parameters, the ML model would produce consistent results of χ_{ML} by calling it offline in Python XGBoost software package or online in Fortran integrated with XGBoost C functions in CAM6. The offline application of the ML model by Zheng, West, et al. (2021) was limited to diagnosing aerosol mixing state index only, but was not used to improve the representation of aerosol mixing and aging in CAM6. On the basis of Zheng, West, et al. (2021), we predict χ_{ML} at the runtime of the CAM6 and further use χ_{ML} to improve aerosol representations (as described in

Section 2.3.2) and evaluate its impact on aerosol activation (as described in Section 2.3.3). Note that, in the generation of training data (Zheng, West, et al., 2021), the simplification or absence of some physical-chemical processes in PartMC-MOSAIC may result in uncertainties for the ML model. For example, Zheng, West, et al. (2021) introduced the Aitken mode sulfate particles into PartMC-MOSAIC through emission, rather than explicitly simulating the process of new particle formation and growth. Inadequate representation of this particle type may lead to an underestimation of BC-free particle abundance in certain regions, which in turn may result in an overestimation of the degree of the internal mixing of BC (i.e., the χ_{ML}). Other processes that are not explicitly treated in the generation of training data include nucleation-scavenging of aerosol, cloud processes, and other processes under conditions above the boundary layer. Further efforts are needed in the future to improve the ML model for better addressing these issues.

2.3.2. Improvement of BC Mixing State Representation

In the default MAM4 scheme, BC has two mixing states (i.e., fresh hydrophobic BC in primary carbon mode and aged hydrophilic BC in accumulation mode). After BC aging is completed, the aged BC is assumed to be fully internally mixed with all other aerosol species in the accumulation mode, that is, by definition, all aerosol particles in the accumulation mode are BC-containing particles. Therefore, due to the lack of BC-free particles, the default MAM4 tends to overestimate the χ_{ML} in PM1 compared to the ML model (Zheng, West, et al., 2021), which may lead to a bias in the mass ratio of coating to BC.

Table 1
Input Parameters for the Machine Learning Model

Input parameter name	Variable in CAM6	Type	Unit
Black carbon	bc_a1 + bc_a4	A	kg/kg
Mineral dust	dst_a1 + dst_a2	A	kg/kg
Sea salt	ncl_a1 + ncl_a2	A	kg/kg
Primary organic matter	pom_a1 + pom_a4	A	kg/kg
Secondary organic aerosol	soa_a1 + soa_a2	A	kg/kg
Sulfate	so4_a1 + so4_a2	A	kg/kg
Dimethyl sulfide	DMS	G	mol/mol
Hydrogen peroxide	H ₂ O ₂	G	mol/mol
Sulfuric acid	H ₂ SO ₄	G	mol/mol
Ozone	O ₃	G	mol/mol
SOA gas	SOAG	G	mol/mol
Sulfur dioxide	SO ₂	G	mol/mol
Air temperature	T	E	K
Relative humidity	RELHUM	E	1
Solar zenith angle	SZA	E	Radian

Note. Type: A (Aerosol bulk concentrations), G (Gas mixing ratios), E (Environmental variables). The words “a1,” “a2,” and “a4” refer to Accumulation, Aitken, and Primary carbon mode.

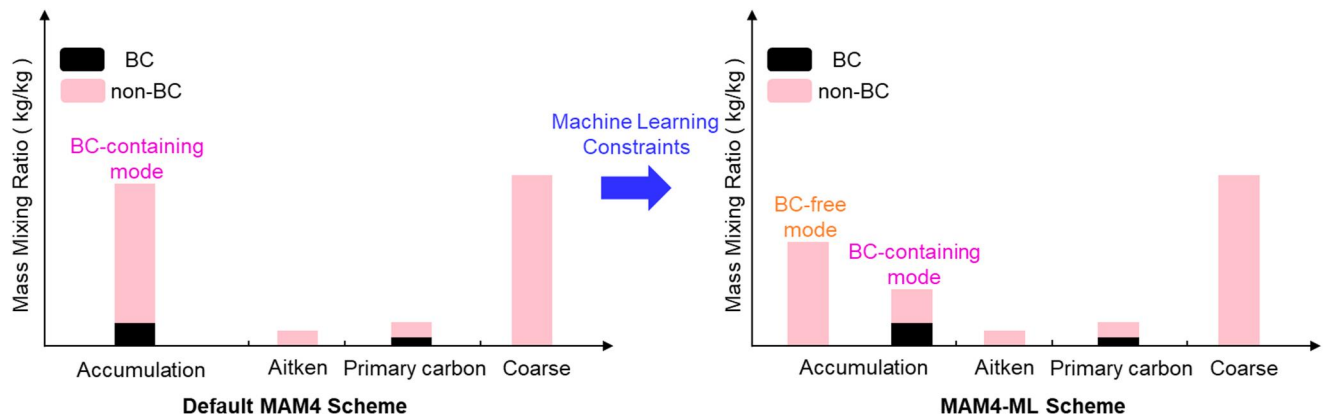


Figure 2. Schematic of aerosol modes in default MAM4 (left panel) and MAM4-ML (right panel). The caption shows BC-free particle mode (orange) and BC-containing particle mode (purple). The legend shows particle composition (BC species in black and non-BC species in pink) in each mode.

To address this issue, we propose a method to improve the BC mixing-state representation based on the difference between the χ_{mode} calculated by MAM4 model information and χ_{ML} predicted by the ML model. Specifically, we partition particles in the accumulation mode in MAM4 into two new modes: BC-containing particles and BC-free particles in the updated MAM4 treatment (see the right panel in Figure 2). As BC mass concentration (m_{BC}) and total aerosol mass concentration (m_{accum}) in the accumulation mode are already tracked by the CAM6 model during each time step, the mass partitioning of non-BC aerosol mass between the BC-containing particles ($m_{\text{non-BC}}$) and BC-free particles ($m_{\text{BC-free}}$) can be determined by the parameter R_{BC} (the mass ratio of non-BC species to BC in BC-containing particles):

$$m_{\text{non-BC}} = R_{\text{BC}} m_{\text{BC}} \quad (9)$$

$$m_{\text{BC-free}} = m_{\text{accum}} - m_{\text{non-BC}} - m_{\text{BC}} \quad (10)$$

To determine R_{BC} , we adjust R_{BC} by the iterative approximation method to make χ_{mode} calculated by Equations 7 and 8 using the updated MAM4 information (referred to as “MAM4-ML χ_{mode} ”) to match the χ_{ML} predicted by the ML model. We use the False Position Method (FPM) to accelerate the convergence of updated R_{BC} . The initial value of MAM4-ML R_{BC} (i.e., the R_{BC} of BC-containing particles in MAM4-ML treatment) equals the R_{BC} in default MAM4 (i.e., the R_{BC} of BC-containing particles in default MAM4 treatment). The convergence criterion is that the difference between the successive approximations of R_{BC} is less than 0.01. Note that the χ_{ML} is directly predicted by the ML model with the input parameters simulated by CAM6. As can be seen from Table 1, the input parameters of aerosol are aerosol bulk concentrations by adding up the aerosol concentrations of each mode in CAM6, which do not carry mixing state information directly by themselves. The R_{BC} and χ_{mode} in the MAM4-ML treatment are always smaller than or equal to those of the default MAM4 treatment. This is because for a given aerosol population, separating out BC-free particles from BC-containing particles in the accumulation mode will result in a smaller R_{BC} and a lower degree of the mixing of BC particles with non-BC particles. This further leads to a lower χ_{mode} . A lower value of χ_{mode} and R_{BC} in MAM4-ML than those in default MAM4 treatment, in turn, means a thinner coating and a less internal mixture of BC in the updated MAM4 treatment.

2.3.3. Updating the Treatment of CCN Activity in CAM6

Changes in BC mixing state affect the hygroscopicity and aerosol activation of BC-containing aerosols, which further influence the in-cloud and below-cloud scavenging of BC directly or indirectly. In the current cloud microphysics of CAM6, stratiform and convective clouds are parameterized separately. In stratiform clouds, aerosol activation is treated explicitly, while there is no explicit treatment of aerosol activation in convective clouds. Therefore, we further modified the treatment of aerosol activation in stratiform clouds based on the MAM4-ML treatment described in Section 2.3.2.

In stratiform clouds, aerosol activation is assumed to occur when clouds expand (i.e., cloud fraction increases) (Ovtchinnikov & Ghan, 2005) or updraft carries air into the base of the cloud (Ghan et al., 1997). When aerosols

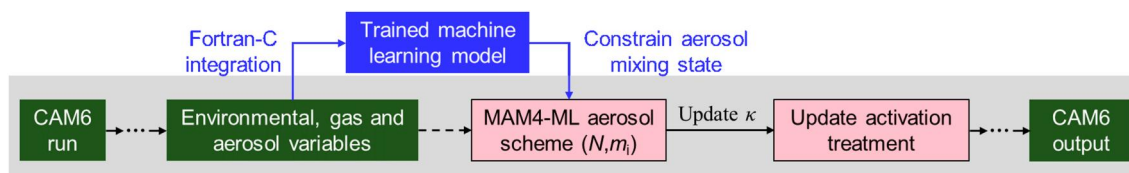


Figure 3. The flow chart of the application of the machine learning (ML) model to the online CAM6. The N , m_i , and κ indicates the aerosol number concentration, aerosol mass concentration of different species, and hygroscopicity of the five aerosol modes in MAM4-ML. The detailed definitions for environmental variables, gas mixing ratios and aerosol bulk concentrations in CAM6 are shown in Table 1.

are activated, the number and species mass of activated interstitial aerosol are transferred to the cloud-borne state (so-called “cloud-borne aerosol” within stratiform clouds, Ghan and Easter (2006)). According to Abdul-Razzak and Ghan (2000), the activated fraction of aerosol number and mass is parameterized based on the updraft velocities and the properties of aerosol modes (i.e., the aerosol number concentration, aerosol species mass concentration, hygroscopicity, aerosol size distribution of each mode). In the default MAM4, the aerosols are involved in activation in four modes (i.e., Aitken, accumulation, coarse, and primary carbon modes).

Here we update the treatment of CCN activity in CAM6 based on the MAM4-ML scheme, and the flow chart is shown in Figure 3. In the module of aerosol activation calculation, we improve BC mixing state representation with ML, as shown in Figure 2 and described in Section 2.3.2. We then updated the aerosol activation treatment by replacing the accumulation mode with two new modes that represent BC-containing particles and BC-free particles, respectively. The size distribution of these two new modes is the same as the size distribution of the accumulation mode. The hygroscopicity is calculated based on the MAM4-ML aerosol representation (i.e., the number concentration and aerosol species mass concentration for five modes) with the volume mixing rules (Liu et al., 2012). We still use the same droplet activation parameterization by Abdul-Razzak and Ghan (2000) for calculating activation fraction of the number and mass of each mode, but we now apply it to five modes in the MAM4-ML treatment instead of four modes in the default MAM4, as we split the accumulation mode into BC-free and BC-containing modes in the MAM4-ML. The activated number and mass for each mode is then transferred to the cloud-borne aerosol. Note that in MAM4-ML, the two new aerosol modes only exist temporarily during the aerosol activation process, and for other model processes, the representations of aerosol modes remain the same as default MAM4.

2.4. Description of Simulations

Two experiments (i.e., the MAM4-default and MAM4-ML) are conducted for the year of 2011. The MAM4-default experiment uses the default CAM6-Chem model. MAM4-ML uses CAM6-Chem coupled with the ML model described in Section 2.3 to examine to what extent the updated CAM6-Chem improves the representation of BC mixing state and CCN activity. Except for the ML model and corresponding modification, the configurations of both simulations are the same. The simulations use FCSO compset and are performed at $0.9^\circ \times 1.25^\circ$ horizontal resolution with prescribed historical sea surface temperature (SST) during 2011. The FCSO compset is a specified dynamics configuration that nudges the meteorological fields toward MERRA2 reanalysis with a 6-hr relaxation timescale (Kooperman et al., 2012). We also use a 2-month spin-up time for both simulations (Barré et al., 2015; Shen et al., 2023).

3. Results

3.1. Simulation of χ and R_{BC}

3.1.1. Simulated χ With Machine Learning

Figure 4 shows the annual mean χ_{mode} for MAM4-default and MAM4-ML at the surface layer and relative difference in zonal mean annual χ_{mode} between the two models. MAM4-default has a higher χ_{mode} (78%) than χ_{ML} (66%) globally (Figure 4a and Figures S1a and S1b in Supporting Information S1), suggesting a more internal mixture of BC with non-BC species. The overestimation results from the uniform mixing assumption in MAM4-default, which considers 100% mass of accumulation mode as BC-containing particles. In MAM4-ML, χ_{mode} (63%) decreases by 19% and matches χ_{ML} well (Figure S1c in Supporting Information S1, except for the Southern Ocean at 30°S – 60°S , which will be explained later). As described in Section 2.3.2. MAM4-ML reduces the extent

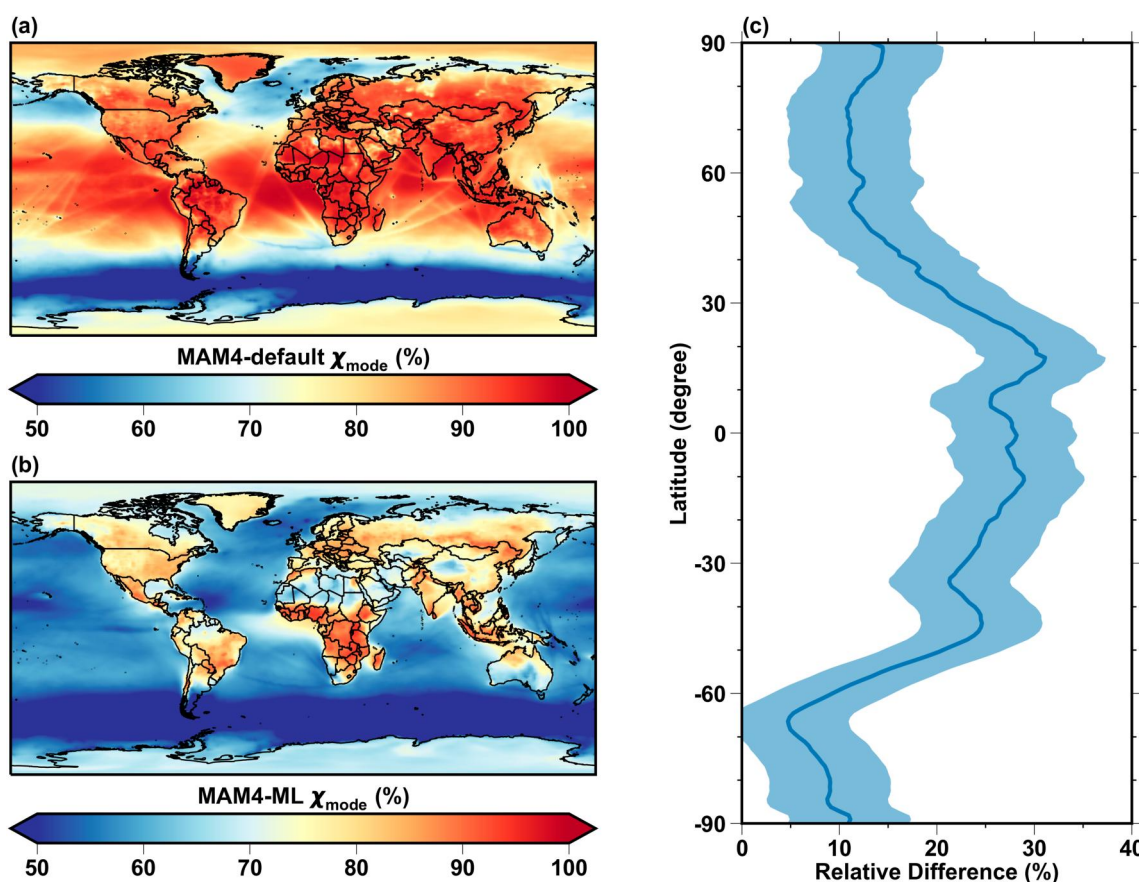


Figure 4. (a), The distribution of annual mean mixing state index (χ_{mode}) at the surface in the MAM4-default. (b), Same as (a), but in MAM4-ML. (c), The zonal mean relative difference in χ_{mode} between (a) and (b). The shaded area in (c) shows the 10th and 90th percent.

of internal mixing of BC by splitting the mass of accumulation mode into two aerosol populations of BC-containing particles (annual mean mass fraction of 48%) and BC-free particles (annual mean mass fraction of 52%) (Figure 5).

The relative difference between χ_{mode} of both models is zonally structured (Figure 4c). The decrease in MAM4-ML χ_{mode} is larger over oceanic and remote regions, as observations have shown that there are more BC-free particles (i.e., fewer BC-containing particles) due to enhanced marine emission, dust emission, and new particle formation (Croft et al., 2016; Kodros et al., 2018; Raatikainen et al., 2015; Sharma et al., 2017; Wentworth et al., 2016; Willis et al., 2016). For example, there is a large reduction ($\sim 27\%$) of MAM4-ML χ_{mode} in low and mid-latitude ocean areas. In these ocean areas, the χ_{mode} is 77% for MAM4-default and 56% for MAM4-ML. This is because most non-BC species (Figure 5) from secondary particle formation (such as sulfate) (Clarke et al., 2013; Kerminen et al., 2018) and primary marine aerosol production (Mårtensson et al., 2003) are moved to BC-free particles in MAM4-ML instead of BC-containing particles in MAM4-default. The considerable difference in χ_{mode} of both models over the oceans suggests that the uniform mixing assumption in the MAM4-default may not hold for the ocean areas. Over low- and mid-latitude source regions where the BC mass fraction is also higher (Figures S2a and S3a in Supporting Information S1), the χ_{mode} is higher for both MAM4-default (90%) and MAM4-ML (78%), and hence the reduction of χ_{mode} is smaller (13%). This phenomenon indicates that BC tends to be more internally mixed with non-BC species in both models, which can be attributed to the mechanism that in polluted source areas, the relatively abundant non-BC particles co-emitted with BC and high concentration of gaseous precursors may enhance the mixing of BC with non-BC species. Yu et al. (2020) found that the observed χ was higher when the pollution was heavier in urban areas, as the non-BC materials have a higher probability of coagulating with or condensing on BC. In polar regions, both MAM4-default χ_{mode} (73%) and MAM4-ML χ_{mode} (65%) are higher than those over low- and mid-latitude oceans, possibly because of the

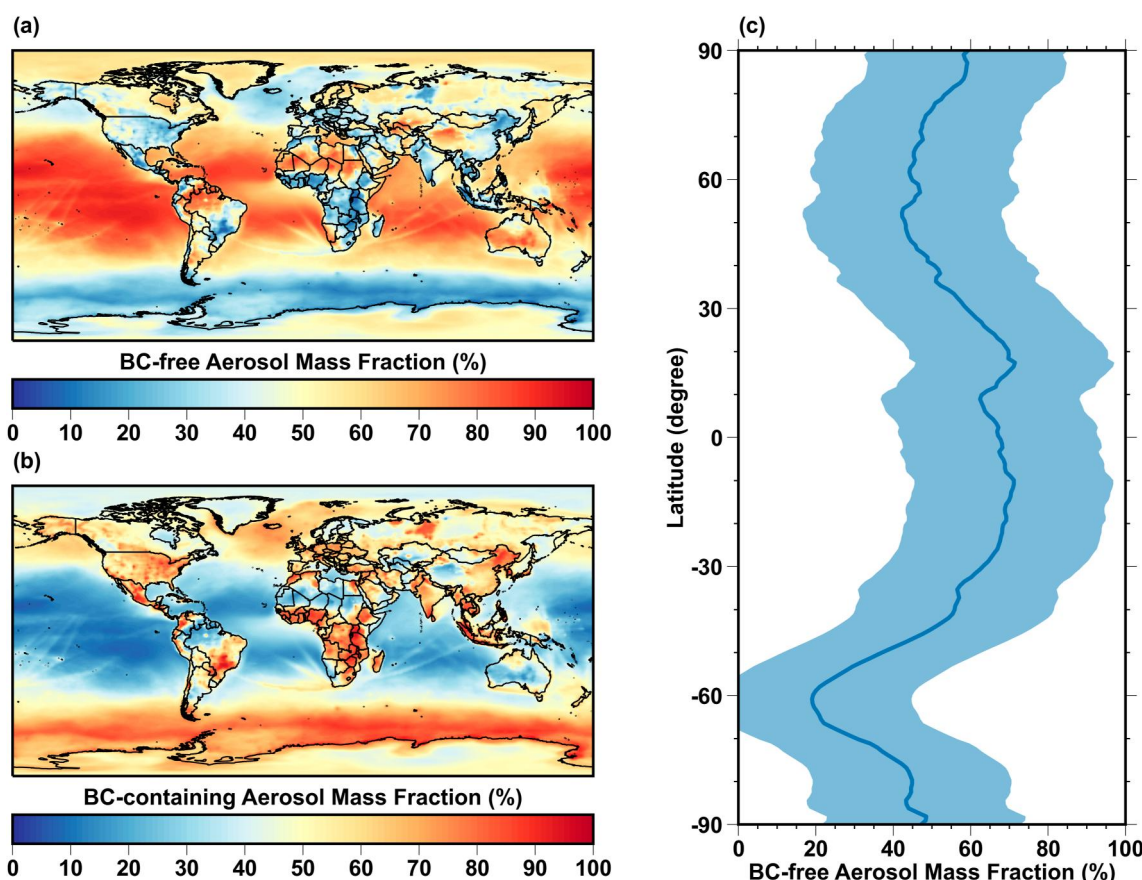


Figure 5. (a), The distribution of annual mean mass fraction of BC-free particles at the surface in the accumulation mode of MAM4-ML. (b), Same as (a), but of BC-containing particles. (c), the zonal mean values of (a). The shaded area in (c) shows the 10th and 90th percentiles.

accumulated mixing of BC with non-BC species and the wet removal and cloud processing of BC-free particles during the transport in both models (Schulz et al., 2019; Willis et al., 2019). As a result, the reduction of χ_{mode} is also smaller (11%) in polar regions. There is a marginal change of χ_{mode} over the Southern Ocean, which is mainly due to the low mass fraction of BC in PM1 ($\sim 1\%$, Figure S3 in Supporting Information S1) in both models. It is interesting to note that Figures 4 and 5 show the variation in χ and BC-containing particles over oceans imposed by international ship emissions, because the BC emissions affect the BC mass concentrations assigned to the primary carbon mode, the prediction of χ_{ML} by the ML model, and the partitioning in the MAM4-ML.

Figure 6 illustrates how χ varies with altitude. The vertical distribution of MAM4-ML χ_{mode} differs significantly from that of MAM4-default. For example, the MAM4-ML χ_{mode} rises from the surface to altitudes of 300 hPa, probably due to the BC aging promoting the mixing of BC with non-BC species during upward transport (Metcalf et al., 2012; Sahu et al., 2012; Schwarz et al., 2008). At altitudes above 300 hPa, MAM4-ML χ_{mode} decreases due to two reasons: (a) most BC is aged (Figure S4b in Supporting Information S1), and (b) strong new particle formation forms sulfate in the tropical upper troposphere and lower stratosphere (Figure S4a in Supporting Information S1) (Bianchi et al., 2016; Williamson et al., 2019; Zhao, Shrivastava, et al., 2020). However, the MAM4-default χ_{mode} increases with altitude, as it treats all increased non-BC species as BC-containing particles. Overall, the MAM4-ML χ_{mode} (63%–77%) agrees well with χ_{ML} (66%–77%) at different altitudes. The MAM4-default χ_{mode} (78%–99%), however, overestimates χ_{ML} by 18%–39% at all altitudes, implying that default MAM4 treatment may overestimate the internal mixing degree of BC in the whole atmosphere. Compared to χ_{ML} , the overestimation of MAM4-default χ_{mode} is relatively constant (18%–23%) at lower altitudes (below 300 hPa). At altitudes above 300 hPa, the overestimation of MAM4-default χ_{mode} becomes large (25%–39%) for both ocean and land (Figure 7). Note that MAM4-ML χ_{mode} is slightly lower (5%) than χ_{ML} near the surface mainly due to low mass fraction of BC over the Southern Ocean.

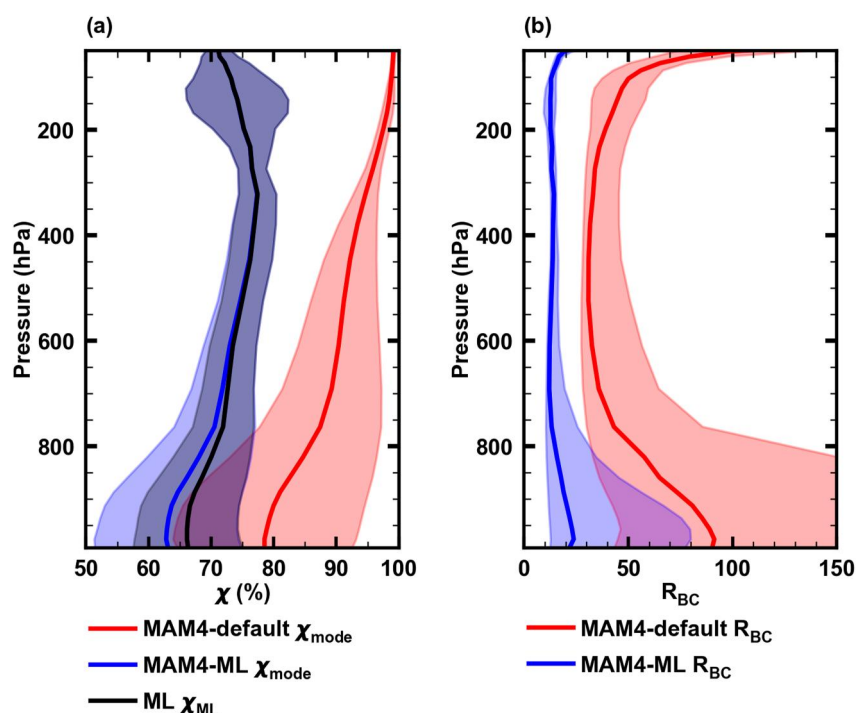


Figure 6. (a), Vertical distribution of the annual mean mixing state indices (χ_{mode}) at each layer in MAM4-default (red) and MAM4-ML (blue), and χ_{ML} in ML model (black). (b), Vertical distribution of the annual median mass ratio of coating to BC (R_{BC}) at each layer in MAM4-default (red) and MAM4-ML (blue). The shaded area shows the 10th and 90th percentiles. In (b), we consider the median R_{BC} instead of mean R_{BC} to avoid anomalies.

3.1.2. Update of R_{BC} With Improved Mixing State Representation

The ensemble mean R_{BC} in the accumulation mode represents the degree of coating for aged BC particles. The uniform mixing assumption in MAM4-default also leads to very large R_{BC} in accumulation mode. Figure 8 shows that MAM4-default R_{BC} is above 10 over most regions and can reach more than 100 over most ocean areas, which is not consistent with observations that R_{BC} is usually lower than 20 (Cappa et al., 2012, 2019; Kompalli et al., 2021; Lee et al., 2017a; Massoli et al., 2015b; McMeeking et al., 2014). The improved BC mixing state representation leads to smaller R_{BC} in MAM4-ML. Figure 8 compares the annual mean MAM4-default R_{BC} and MAM4-ML R_{BC} at the surface and relative difference in zonal mean annual R_{BC} . The global mean R_{BC} decreased by 52% in MAM4-ML, as more than half (52%, Figure 5) of the mass of non-BC species now moves to BC-free particles instead of coating BC in MAM4-ML. The global median R_{BC} decrease from 90 in MAM4-default to 22 (76% reduction) in MAM4-ML. The decline in MAM4-ML R_{BC} has a similar zonal structure to the reduction of χ_{mode} (Figure 4c). Specifically, the difference in R_{BC} between both models is larger over low- and mid-latitude oceans than in the polluted source areas. Fierce et al. (2016) also found that compared to the particle-resolved results, the GISS-MATRIX model (the global model fields generated by Goddard Institute for Space Studies climate modelE coupled to Multiconfiguration Aerosol TRacker of mIXing state) tends to have more bias in simulating BC absorption enhancement over the oceans than in the source regions. Over the Southern Ocean, R_{BC} drops little in the MAM4-ML due to the low mass fraction of BC.

Figure 9 compares the R_{BC} between the model and observations at the surface. The observed R_{BC} is measured mainly by soot particle aerosol mass spectrometer (SP-AMS), as summarized in Table S1 in Supporting Information S1. Figure 9 shows the MAM4-default R_{BC} is generally higher than the observed R_{BC} , while the MAM4-ML R_{BC} is much closer to the observations. The observed R_{BC} varies due to BC emission sources, atmospheric environment, and transport (Cappa et al., 2012, 2019; Lee et al., 2017a; Liu et al., 2015, 2017; McMeeking et al., 2014; Zhao et al., 2021), which the MAM4-ML R_{BC} can capture to some extent. For instance, Figure 9 shows the observed R_{BC} was lower in polluted urban cities, such as Beijing (5.0), Shanghai (5.0), Nanjing (4.3), Shenzhen (2.5), and Fresno (2.3). In these urban cities, the MAM4-ML has lower R_{BC} than MAM4-default (e.g.,

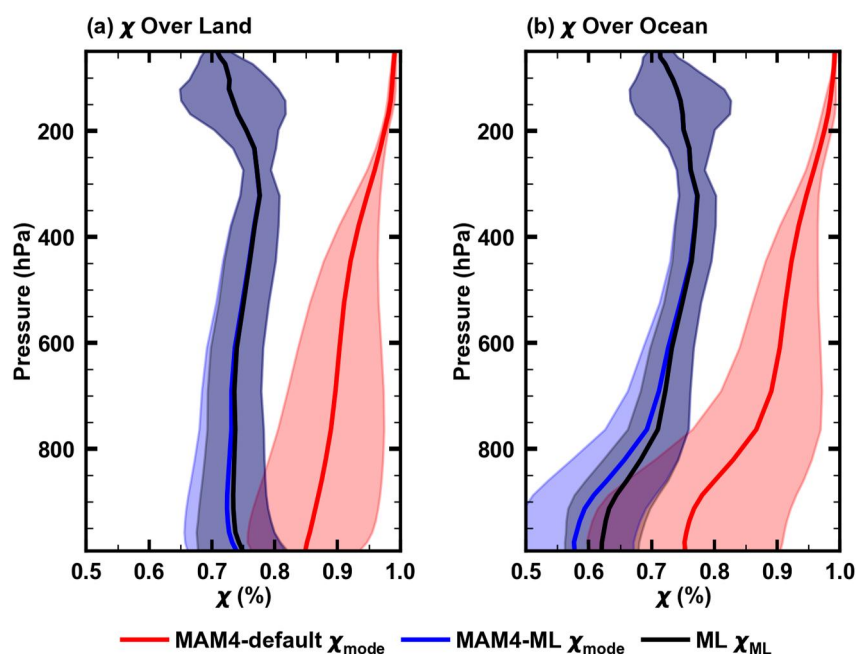


Figure 7. (a), Vertical distribution of the annual mean mixing state indices (χ) at each layer in MAM4-default (red), MAM4-ML (blue) and ML model (black) over the land. (b), Same as (a), but over the ocean. The shaded area shows the 10th and 90th percentiles.

Beijing: 7.0/10.1; Shanghai: 9.6/11.1; Nanjing: 5.9/7.8; Shenzhen: 11.6/16.2; Fresno: 11.6/20.8). The relatively small R_{BC} in observations and MAM4-ML may be due to the lower BC aging degree for freshly emitted BC (Lee et al., 2019; Shen et al., 2023). Zhao et al. (2021) found that the freshly emitted BC from anthropogenic emissions tends to have a low R_{BC} (about 2.56–3.16) in a suburban site in Taizhou. In addition, the comparison of diurnal R_{BC} in Beijing (Xie et al., 2019) and Collier et al. (2018) (Figure S5 in Supporting Information S1) shows that the MAM4-ML R_{BC} has a similar diurnal pattern as the observed ones. Specifically, the R_{BC} is lower at night and morning, and it rises gradually with a peak value in the afternoon (12:00–18:00), which may relate to the diurnal variation of BC aging rate (Chen et al., 2017; Shen et al., 2023). Observations show that due to the BC aging process, there is a general trend for more aged BC-containing particles in polluted urban areas to have a thicker coating (Collier et al., 2018; Lee et al., 2017b; Massoli et al., 2015b; Wang et al., 2019b). Note that in Figure 9 and Figure S5 in Supporting Information S1, the R_{BC} is only ensemble mean R_{BC} of aged BC-containing particles in the accumulation mode for the model but fresh and aged BC-containing particles for the observations. If we also consider the fresh BC-containing particles in primary carbon mode in the model, the conclusions remain similar (Figure S6 in Supporting Information S1).

Previous laboratory and observation studies have shown that R_{BC} is usually associated with emission sources, that is, lower R_{BC} values in regions dominated by fossil fuel emissions and higher R_{BC} values in regions influenced by biomass-burning emissions (Holanda et al., 2020; Ko et al., 2020; Liu et al., 2017; Metcalf et al., 2012). In our study, the MAM4-ML R_{BC} can also capture the similar feature. For example, Figure 10 shows that the median values of R_{BC} in accumulation mode are generally higher than 15/30 for MAM4-ML/MAM4-default over source regions influenced by biomass burning, such as the United States (median R_{BC} = 27/79), Middle Africa (median R_{BC} = 20/30), and the Amazon (median R_{BC} = 17/36). In contrast, the median values of R_{BC} are ~10/20 for MAM4-ML/MAM4-default in source regions dominated by anthropogenic fossil fuel emission, such as Europe (median R_{BC} = 13/25), India (median R_{BC} = 10/19), and Eastern China (median R_{BC} = 7/15). Previous observational studies have also measured higher R_{BC} values (>10, Cappa et al. (2012), Massoli et al. (2015a), McMeeking et al. (2014)) in North America than those observed at polluted urban cities of China (<5, Cao et al. (2022), Cui et al. (2022), Wang et al. (2019b), Wu et al. (2019)). Furthermore, laboratory experiments by Liu et al. (2017) showed that R_{BC} from biomass burning was larger than that from solid fuel and traffic sources, and they also simulated the higher R_{BC} over the United States, Amazon, and Western and Middle Africa by the 3D Global Model of Aerosol Processes model (GLOMAP).

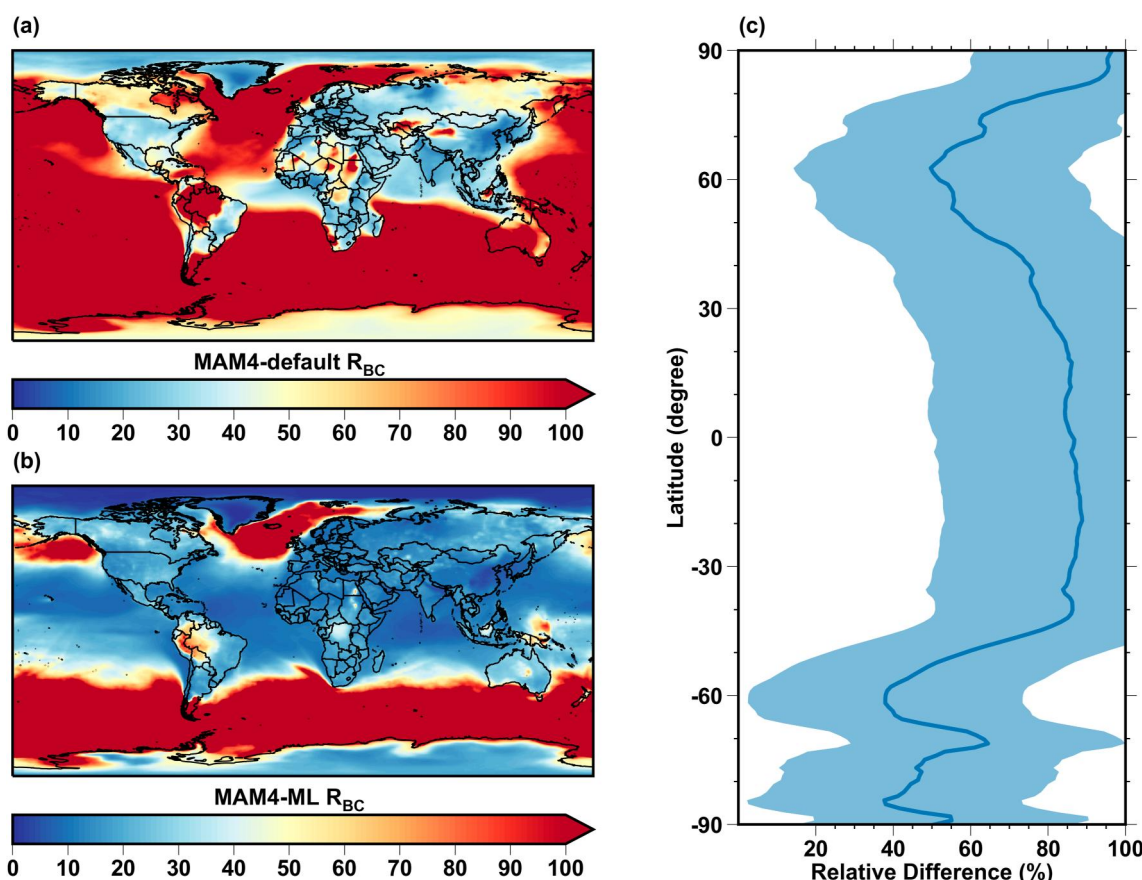


Figure 8. (a), The distribution of annual median mass ratio of coating to BC (R_{BC}) at the surface in MAM4-default. (b), Same as (a), but in MAM4-ML. (c), The zonal mean relative difference between (a) and (b). The shaded area in (c) shows the 10th and 90th percentiles. In (a) and (b), we consider the median R_{BC} instead of mean R_{BC} to avoid anomalies.

The MAM4-ML R_{BC} is higher over remote land and oceans than that in source regions because of BC aging during long-range transport, which has also been documented by previous observational studies (Kompalli et al., 2021; Wang et al., 2017). For example, Zhai et al. (2022) found an R_{BC} of 1.36 for fresh local traffic emission in Shanghai, but it increased to 5.95 for long-range transported aged air from northern China. Wang et al. (2017) detected a larger R_{BC} of 7.0 in the remote Tibetan Plateau, and our corresponding R_{BC} is 26.4 for MAM4-default and 12.7 for MAM4-ML (Figure 9). Kompalli et al. (2021) measured R_{BC} values of 4.24, 5.76, 6.91, and 8.31 from ship campaign covering the Arabian Sea and the remote Indian Ocean, and our simulated R_{BC} values are 20.79, 19.81, 17.38, and 18.29 for MAM4-default, and 6.67, 7.99, 8.43, and 6.71 for MAM4-ML, respectively (Figure 9). We also see from Figure 9 that the improvement of R_{BC} is larger over the oceans. Furthermore, from oceans to polar regions, the R_{BC} in both models drops significantly, which may result from easier activation and removal of BC-containing particles with larger R_{BC} at high latitudes (Discussed later in Section 3.2).

Above the surface layer, the MAM4-ML R_{BC} declines gradually at altitudes below 600 hPa (Figure 6b), especially in the ocean (Figure S7 in Supporting Information S1). Due to the lack of direct measurements, we can hardly compare the vertical profiles of R_{BC} between the model and observations directly. The decline trend in MAM4-ML R_{BC} likely resulted from the fact that BC-containing particles with higher CCN activities (i.e., the larger R_{BC}) are more easily removed due to higher hygroscopicity when they are uplifted to higher altitudes (Moteki et al., 2012, 2019; Ohata et al., 2016, 2021a). We notice that the general decline of MAM4-ML R_{BC} with altitudes is consistent with that of the shell/core ratio of BC diameter (D_p/D_c), as aircraft observations show that D_p/D_c measured by SP2 (the Single-Particle Soot Photometer) usually drops with altitudes (Ding et al., 2019; Liu, Ding, et al., 2020; Ohata, Mori, et al., 2021; Sahu et al., 2012; Zhao et al., 2019, 2020b). Above the surface, the global mean R_{BC} in MAM4-default is greater than 30 at all the altitudes, and global mean R_{BC} in MAM4-ML is reduced

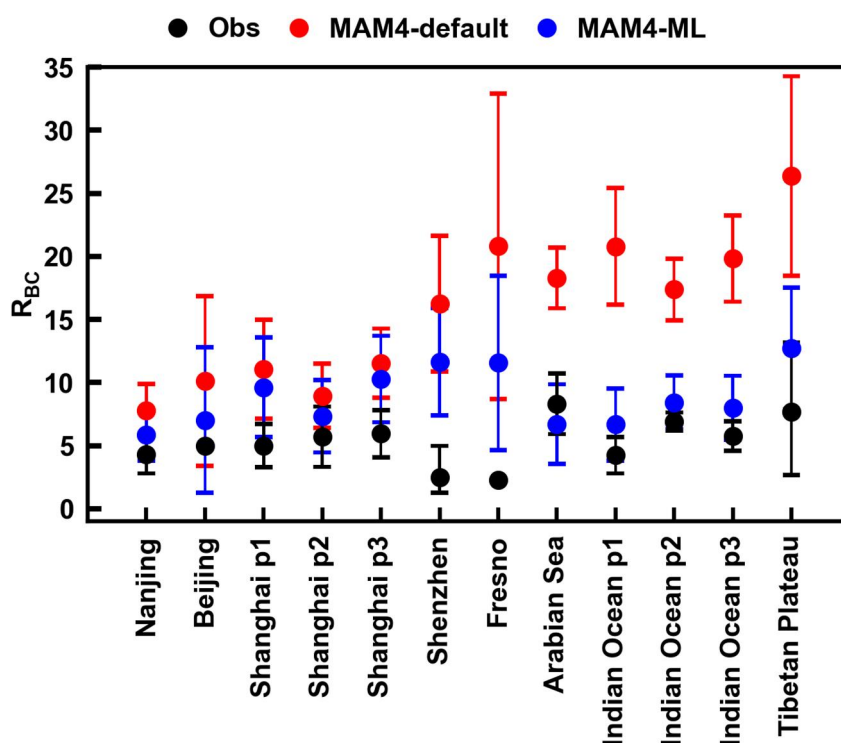


Figure 9. (a), Comparisons of mass ratio of coating to BC (R_{BC}) at the surface between observations (black) and model results from MAM4-default (red) and MAM4-ML (blue). The dots indicate the mean R_{BC} and the lines indicate the standard deviation, except that the lines in Shenzhen and Tibetan Plateau indicate the range of R_{BC} in observations. The observed R_{BC} (Table S1 in Supporting Information S1) is from Wu et al. (2019) (Nanjing), Wang et al. (2019a) (Beijing), Cui et al. (2022) (Shanghai p1), Zhai et al. (2022) (Shanghai p2-p3), Cao et al. (2022) (Shenzhen), Collier et al. (2018) (Fresno), Kompalli et al. (2021) (Arabian Sea and Indian Ocean p1-p3), and Wang et al. (2017) (Tibetan Plateau), respectively. The R_{BC} is from aged BC-containing particles in the accumulation mode for the model results but fresh and aged BC-containing particles for the observations. Simulation period is when the corresponding observation is available as shown in Table S1 in Supporting Information S1.

to 10–20. From surface observations around the world, it is unlikely that the R_{BC} could reach more than 20 (Cappa et al., 2012, 2019; Healy et al., 2015; Liu et al., 2015, 2017; McMeeking et al., 2014; Peng et al., 2016), so we consider it may be the right direction for MAM4-ML to reduce R_{BC} at higher altitudes. As R_{BC} at higher altitudes is rarely observed, direct measurements of R_{BC} are urgently needed to evaluate MAM4-ML R_{BC} at higher altitudes. From the surface to altitudes of 600 hPa, the relative difference in R_{BC} between both models decreases mainly over the ocean areas (Figure S7 in Supporting Information S1). While at higher altitudes (above 600 hPa), the update of R_{BC} changes little across altitudes.

Our results indicate that uniform mixing assumption leads to the higher χ_{mode} and R_{BC} in MAM4-default. By improving BC mixing state representation with ML, MAM4-ML reduces χ_{mode} and R_{BC} significantly. Meanwhile, the amount of BC-free particles in accumulation mode is also determined in MAM4-ML. The MAM4-default only has BC-free particles in Aitken mode for PM1, which make up 64% of PM1's total number (Figure 11a). The MAM4-ML has BC-free particles in both Aitken and accumulation modes, which account for 79% of PM1's total number (Figure 11b). Recently, BC-free particles have been observed to have a substantial number fraction of PM1 (Chen et al., 2022; Collier et al., 2018; Zhang et al., 2017). In urban Beijing, Chen et al. (2022) reported that the number fraction of BC-free particles (40–300 nm diameter range) was 23%–33% in the winter of 2018 and 52%–69% in the summer of 2019. At the same site but in 2011, the MAM4-ML simulates that 31% of PM1's number is BC-free particles in winter and 35% in summer, while the MAM4-default simulates 21% in winter and 15% in summer, respectively. The BC-free particles may also account for a significant mass fraction of PM1. For example, in an urban Fresno, Collier et al. (2018) found that 83% of PM1's mass was BC-free particles in the Winter of 2014. In the Winter of 2011 at Fresno, the simulated BC-free particles have a mass fraction of 2% in PM1 for MAM4-default and 24% for MAM4-ML. Though not a direct comparison, both

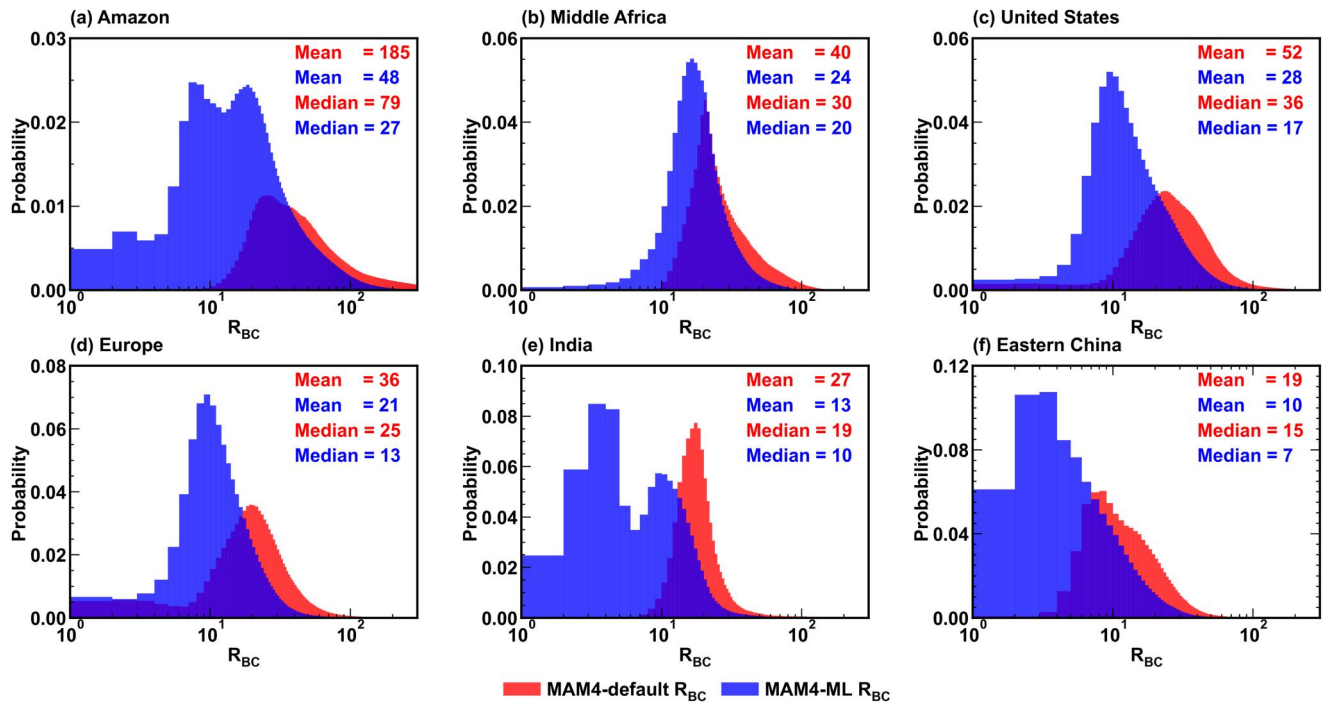


Figure 10. Probability density of R_{BC} at the surface in MAM4-default (red) and MAM4-ML (blue). The regions are Amazon (a, 25°S–10°N, 80°W–45°W), Middle Africa (b, 10°S–10°N, 10°E–35°E), United States (c, 25°N–50°N 75°W–125°W), Europe (d, 35°N–55°N, 0°E–40°E), India (e, 15°N–30°N, 70°E–90°E), and Eastern China (f, 25°N–40°N, 100°E–120°E), respectively.

observation and MAM4-ML suggest that a large amount of BC-free particles in PM₁ may be externally mixed with BC-containing particles near source regions. In addition, measurements in the remote Arctic showed that BC-free particles dominated (76%–90%) the particle number in the 200–400 nm diameter range (Kodros et al., 2018; Raatikainen et al., 2015; Sharma et al., 2017). And in MAM4-ML in 2011, a majority (85%, ranging from 59% to 98%) of the PM₁'s number are BC-free particles at high latitudes in the Arctic (70°N–90°N). Indeed, Figure 11 also shows the dominant number fraction of BC-free particles in the low- and mid-latitude oceans (85%) in MAM4-ML, where the overestimation of χ_{mode} for MAM4-default is also larger. The zonal distribution of MAM4-ML BC-free particle mass fraction in the accumulation mode is generally consistent with the zonal patterns of the decrease in χ_{mode} and R_{BC} (Figures 4c and 8c). The latitudinal dependence of BC-containing particle mass fractions might be determined by the convective transport and wet removal of hydrophilic BC-free particles during the transport to higher latitudes. A higher fraction of BC-free particles in the accumulation mode usually implies a larger bias in BC mixing states of MAM4-default, as there are no BC-free particles in accumulation mode of MAM4-default.

3.2. Impact of Updated R_{BC} on Aerosol Activation and Concentration

Figure 12a–12c compare the hygroscopicity of BC-containing particles of accumulation mode in MAM4-default and MAM4-ML. Figures 12d and 12e show the mean activation fraction of the mass of BC-containing particles in accumulation mode at the surface for MAM4-default and MAM4-ML. Both models show a similar spatial pattern of hygroscopicity, but the global mean hygroscopicity decreases by 9% in MAM4-ML due to the lower R_{BC} (Figure 8). As a result, the global mean activation fraction in MAM4-ML is reduced by 20% at the surface layer. In polar regions, the decrease in hygroscopicity is most significant, such as the Arctic (32%) and Antarctic (16%, 65–90S). The reduction in activation fraction is also most significant in polar regions, such as the Arctic (45%) and Antarctic (38%). This is because in polar regions, R_{BC} is small for both models (i.e., they have a high BC mass fraction in BC-containing aerosol, which can reach up to 27% in the Arctic), and the hygroscopicity of BC-containing aerosols is sensitive to the decrease in R_{BC} (i.e., the increase in BC mass fraction). In contrast, over the low- and mid-latitude oceans, the drop of R_{BC} is relatively significant in MAM4-ML, but the hygroscopicity decreases little. This is mainly because R_{BC} in both models is still large over these regions (i.e., they have low BC

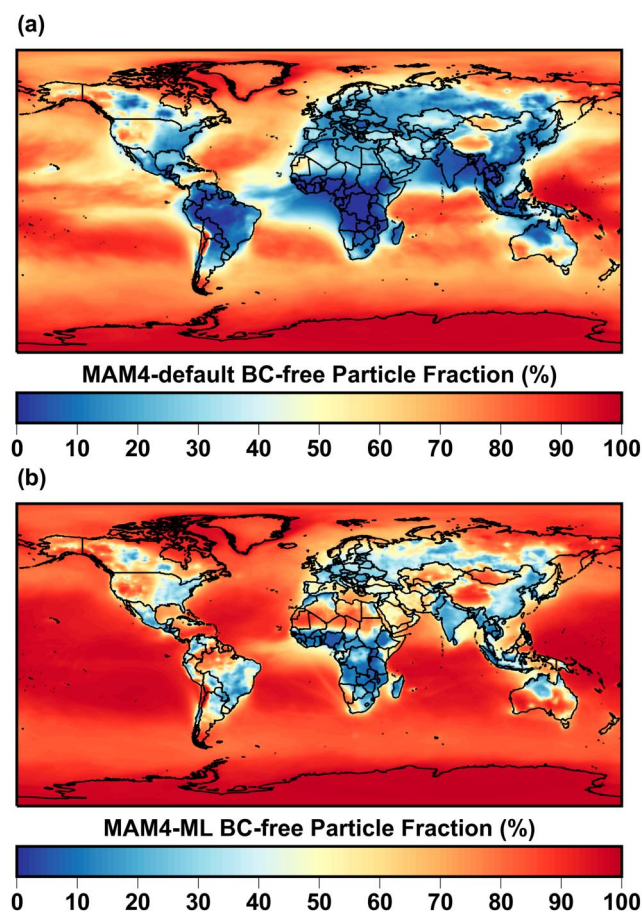


Figure 11. (a), The annual mean number fraction of BC-free particles in PM1 at the surface in MAM4-default. (b), Same as (a), but in MAM4-ML.

Information S1). There is a substantial increase of MAM4-ML surface BC concentration over the North Pacific Ocean, which may be due to the transport of continentally derived BC (such as Eastern China, Russia, and Southeast Asia) into the North Pacific atmosphere (Wu et al., 2020; Zhu et al., 2020). Simulation time does not significantly affect the implications of the MAM4-ML scheme in our study. For example, the increase in surface BC concentration of MAM4-ML in different seasons exhibits similar spatial patterns to the annual results (Figure S12 in Supporting Information S1).

Figure 14c shows the vertical profiles of the relative difference in BC concentration between both models. The increase in MAM4-ML BC concentration rises from the surface to altitudes of 600 hPa and then decreases gradually at altitudes above 600 hPa (Figure 14c). As a result, the increase is significant (5%–10%) at altitudes between 800 and 300 hPa. The vertical distribution of the increase in MAM4-ML BC concentration is generally consistent with that of the lower BC activation fraction (except for the surface). In the surface layer, there is a significant decrease in MAM4-ML BC activation fraction, but a 0.6% increase in global mean surface BC concentration, mainly because of the strong influence of emissions over source regions. According to Shen et al. (2023), the in-cloud scavenging at the surface layer is less significant than in the middle and upper troposphere for BC simulation near source regions in CAM6 model. In fact, for remote Arctic regions, MAM4-ML BC concentration increases significantly from the surface to altitudes of 200 hPa mainly due to accumulation mode BC (Figure S13 in Supporting Information S1), which may help reduce the underestimation of observed BC concentrations at lower altitudes in the Northern Hemisphere high latitudes for MAM4 (Liu et al., 2016). Figure 14b shows the MAM4-ML BC burden increases in most of the Northern Hemisphere, depending on the vertical distribution of the decrease in BC activation fractions and the transport of BC. Globally, MAM4-ML BC burden increases by 4%, and 93% of the increase can be attributed to the accumulation mode BC burden (Table 2). The spatial distribution of the increase in MAM4-ML BC burden is also consistent with that of the accumulation

mass fractions in BC-containing aerosol, which are 1% for MAM4-default and 4% for MAM4-ML). Therefore, the hygroscopicity of BC-containing aerosols depends mainly on the relative percentage of various non-BC species in the coatings of BC, which are similar for both models, as described in Section 2.3.2. Over low-latitude oceans (30°S–30°N), the activation fraction hardly changes. This is because CAM6 model simulates few stratiform clouds there throughout the year (Li et al., 2022), and hence the accumulation mode particles are not activated in either model. In low- and mid-latitude lands, the decline in hygroscopicity and activation fraction is small because of the slight reduction of MAM4-ML R_{BC} .

Above the surface, the difference in hygroscopicity between both models decreases with altitude (Figure 13), which is generally consistent with that of R_{BC} between both models (Figure 6b). The activation fraction in MAM4-default is very high (>95%) across the whole atmosphere, while it drops significantly in MAM4-ML with the lowest mean value of 78% at the surface. Compared to MAM4-default, MAM4-ML has a larger spread of BC activation fraction, which means that the BC activation varies more widely across different regions of the world due to the updated aerosol mixing state representation. The drop in BC activation fraction is 10%–20% at lower altitudes (below 600 hPa). At altitudes above 300 hPa, there is no significant change in BC activation fraction because the aerosols are mostly (>95%) activated in both models.

Figure 14 shows the relative differences in BC mass concentration and burden between MAM4-default and MAM4-ML. The increase in surface BC concentration is most evident in the Arctic (6.9%), especially for the accumulation mode BC (23.9%) (Figure S9a in Supporting Information S1). Although the MAM4-ML BC activation fraction decreases in Eurasia, the relative increase of surface BC concentration is only 0.1%, probably due to the relatively low activation frequency for stratiform clouds (Figure S11 in Supporting Information S1). However, the absolute increase in MAM4-ML surface BC concentration is notable in Eurasia (Figure S10a in Supporting

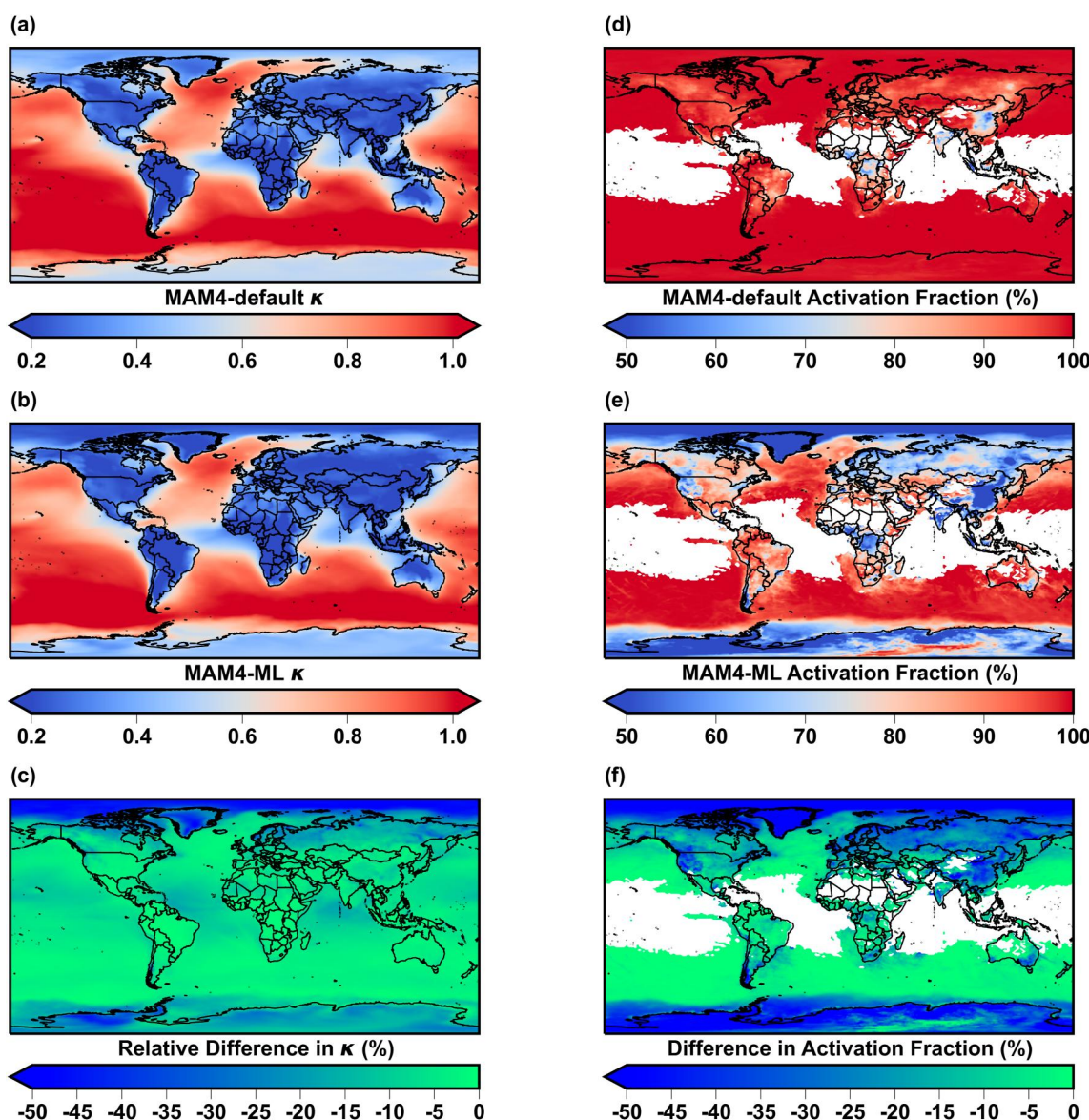


Figure 12. The distribution of annual mean hygroscopicity (κ) of BC-containing particles at the surface in the accumulation mode of MAM4-default. (b), Same as (a), but in MAM4-ML. (c), the relative difference in κ between (a) and (b), that is, $(\kappa_{\text{MAM4-ML}} - \kappa_{\text{MAM4-default}}) / \kappa_{\text{MAM4-default}}$. (d), The distribution of annual mean activation fraction of the mass of BC-containing particles at the surface in the accumulation mode of MAM4-default. (e), Same as (d), but of MAM4-ML. (f), the absolute difference in activation fraction (AF, %) between (d) and (e), that is, $\text{AF}_{\text{MAM4-ML}} - \text{AF}_{\text{MAM4-default}}$.

mode BC (Figure S9 in Supporting Information S1). The slight increase in primary carbon mode BC burden may be due to the lower dry and wet deposition in MAM4-ML (Table 2). BC lifetime is longer due to smaller BC activation fraction in MAM4-ML.

Note that non-BC aerosols in accumulation mode do not differ much between MAM4-default and MAM4-ML (Figures S9 and S10 in Supporting Information S1). This is mainly because their activation fraction in MAM4-ML decreases for the BC-containing mode, while it increases for the BC-free mode (Figure S14 in Supporting Information S1) due to the absence of hydrophobic BC. Aerosols in MAM4-ML Aitken mode (except for dust) increase slightly (Figure S9 in Supporting Information S1), mainly due to the decrease in the maximum supersaturation (Figure S8 in Supporting Information S1). The relative increase in Aitken mode dust is large over the ocean and at high latitudes, but its mass concentration in these areas is several orders of magnitude lower than that of other aerosols in MAM4-default Aitken mode (Figure S2 in Supporting Information S1).

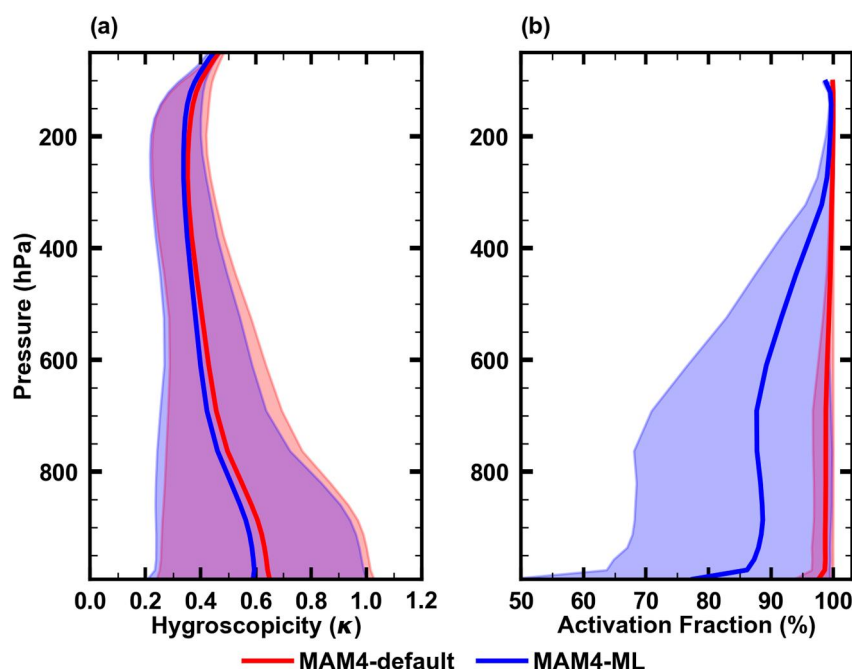


Figure 13. (a), Vertical distribution of annual mean hygroscopicity (κ) of BC-containing particles in the accumulation mode at each layer. (b), Vertical distribution of annual mean activation fraction of the mass of BC-containing particles in the accumulation mode at each layer. Results are from MAM4-default (red) and MAM4-ML (blue). The shaded area shows the 10th and 90th percentiles.

Simulated CCN concentrations at the surface in MAM4-ML are similar to that in MAM4-default (Figure S15 in Supporting Information S1). The MAM4-ML scheme mainly affects BC concentrations and has little effect on non-BC aerosols (Figures S9 and S10 in Supporting Information S1). The overall impact of MAM4-ML scheme on the BC budget is small (Table 2), because the MAM4-ML does not affect in-cloud scavenging from convective clouds and below-cloud scavenging of BC. Although the drop in BC activation fraction for stratiform clouds can be up to 10%–20% throughout the atmosphere (Figure 13b) in MAM4-ML compared to that in MAM4-default, the activation frequency is low over low- and mid-latitude regions (Figure S11 in Supporting Information S1).

A large number of observations show that BC optical absorption enhancement can vary from 1 to ~3 with the R_{BC} ranging from 0 to ~20 (Cappa et al., 2012, 2019; Healy et al., 2015; Lee et al., 2017a; Liu et al., 2017; Peng et al., 2016). As the MAM4-ML scheme reduces the global mean R_{BC} to lower than 20 throughout the atmosphere, the MAM4-ML scheme could have significant implications for BC optical absorption and radiative effects, which need to be further evaluated.

Several factors may introduce uncertainties for the online application of the ML model. For example, the calculation of χ_{mode} in MAM4-ML and MAM4-default, as well as the ML model itself, may be uncertain in regions where the BC mass fraction is extremely low (i.e., less than 1%, such as the Southern Ocean), because the quantification of aerosol mixing state only makes sense when there are at least two species present at a given location. Moreover, due to the simplification or lack of some physical-chemical processes (as described in Section 3.1) in the generation of training data (Zheng, West, et al., 2021), applying the ML model in the whole atmosphere in CAM6 would lead to potential bias in MAM4-ML in higher altitudes (e.g., vertical profiles of χ_{mode} and R_{BC} in MAM4-ML), which necessitates further improvement of the ML model itself.

4. Conclusions

In this work, we coupled a ML model of BC mixing state index learned from particle-resolved simulations into CAM6 MAM4-ML. In MAM4-ML, we partition accumulation mode particles into two new modes, BC-free particles and BC-containing particles. We use the parameter R_{BC} to determine the mass fraction of BC-free particles and BC-containing particles in accumulation mode, and we adjust R_{BC} to make the MAM4-ML χ_{mode}

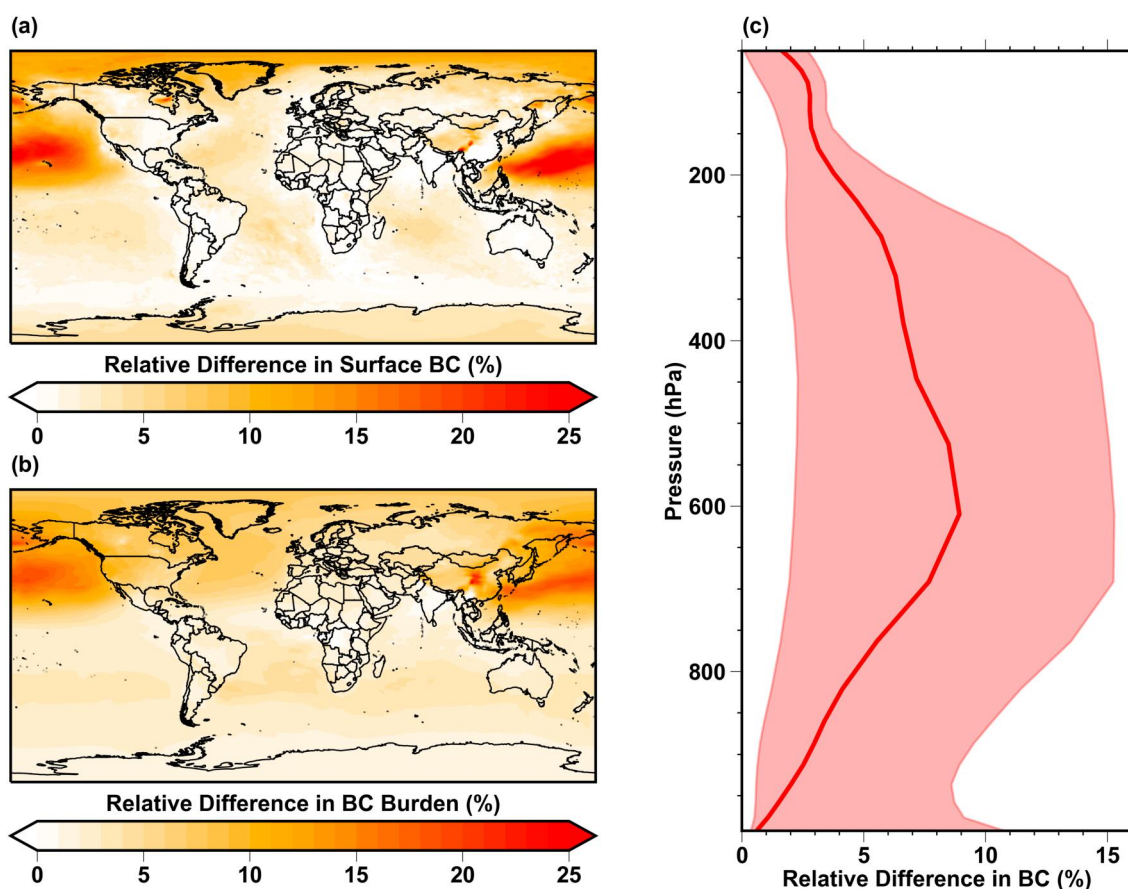


Figure 14. (a), Relative differences in annual mean surface BC mass concentration between MAM4-default and MAM4-ML. (b) same as (a), but in BC burden. (c), Vertical distribution of the relative differences in annual mean BC mass concentration at each layer between MAM4-default and MAM4-ML. The shaded area in (c) shows the 10th and 90th percentiles.

match χ_{ML} . We updated the aerosol activation parameterization based on the improved mixing state representation and quantified the impacts of changes in BC mixing states on CCN activity and BC concentrations.

MAM4-default considers aerosols as fully internal mixtures with uniform composition within each aerosol mode and has therefore a higher χ_{mode} (78%) than χ_{ML} (66%). In contrast, MAM4-ML splits the accumulation mode into BC-containing particles (48%) and BC-free particles (52%), and MAM4-ML χ_{mode} (63%) is reduced by 19% compared to that of MAM4-default and matches the χ_{ML} values well. The relative difference in χ_{mode} between both models is zonally structured, which is larger over the low- and mid-latitude oceans (24%) but smaller over the low- and mid-latitude source regions (13%) and the polar regions (11%). Above the surface, the MAM4-ML χ_{mode} increases from the surface to altitudes of 300 hPa, and then decreases with altitude. However, the MAM4-default χ_{mode} keeps increasing across altitudes.

Compared to χ_{ML} , MAM4-default χ_{mode} overestimates χ_{ML} by 18%–39% throughout the atmosphere, especially at higher altitudes (25%–39%, above 300 hPa), while MAM4-ML χ_{mode} agrees well with χ_{ML} across altitudes.

The uniform mixing assumption in MAM4-default also leads to very large R_{BC} in the accumulation mode (>10 over most regions and >100 over most oceans), which is not consistent with observations. MAM4-ML reduces the global mean R_{BC} by 52% as more than half of the mass of accumulation mode moves to BC-free particles instead of coating BC. The global median R_{BC} also decreases from 90 in MAM4-default to 22 (76% reduction) in MAM4-ML. Similar to χ_{mode} , the decrease in MAM4-ML R_{BC} is more significant over low- and mid-latitude oceans than in polluted source areas. The MAM4-ML

Table 2

Global BC Budgets for MAM4-Default and MAM4-ML in 2011

	MAM4-default	MAM4-ML
Burden (Tg)	0.1500	0.1561
in accumulation	0.1278	0.1335
in primary carbon	0.0221	0.0226
Dry deposition (Tg yr ⁻¹)	2.5656	2.5626
Wet deposition (Tg yr ⁻¹)	6.0927	6.0921
Lifetime (days)	6.32	6.58

R_{BC} agrees well with measurements and exhibits behavior consistent with observations. Specifically, (a) the MAM4-ML R_{BC} is lower for fossil fuel sources and higher for biomass-burning sources; (b) in the source area, the MAM4-ML R_{BC} is relatively small, and it peaks in the afternoon due to BC aging; (c) after long-range transport, the MAM4-ML R_{BC} becomes larger over remote oceans and land than over source areas; (d) the MAM4-ML R_{BC} decreases slightly with altitude.

The decrease in MAM4-ML R_{BC} leads to a 9% reduction in the hygroscopicity of BC-containing aerosol in accumulation mode. Further, the mean activation fraction of accumulation mode BC is reduced by 20% in MAM4-ML. The surface Arctic shows the largest changes, with a 32% drop in hygroscopicity, a 45% drop in activation fraction, and a 6.9% rise in BC concentration. The drop in BC activation fraction is substantial (10%–20%) at lower altitudes (below 600 hPa), and the increase in MAM4-ML BC concentration is significant (5%–10%) at altitudes between 800 and 300 hPa. The global mean BC burden has increased by 4%, and 93% of the increase can be attributed to the accumulation mode BC. The lower activation fraction of accumulation mode BC increases BC lifetime.

Currently, modeling the aerosol mixing states is a challenge for GCMs due to computational efficiency. The PartMC-MOSAIC model resolves aerosol mixing state explicitly, but the model is too expensive to use it directly in GCMs with simulation times on the order of months or years. Coupling a ML model trained from observations or high-detail simulations into the online simulations of GCMs is a promising method for improvements of model treatment. We provide a framework for improving model process representations by online simulations with ML. The ML model constrains the coupled GCMs toward a more realistic model process treatment, such as BC mixing states and the mass ratio of coating to BC. We highlight the model bias in aerosol mixing states for current GCMs and its impacts on aerosol activation in stratiform clouds, which may help reduce the uncertainty of aerosol-cloud interactions. It is also important to further investigate the impact of updated aerosol mixing states on aerosol activation in convective clouds in the future (Shan et al., 2021; Wang et al., 2013; Yu et al., 2019). Our method can be further used to address another important issue of aerosol climate forcing, that is, aerosol optical parameterization and radiative effect. Future studies can also use aerosol mixing states defined in different ways (such as hygroscopicity, absorption, and volatility) to jointly constrain other aspects of model physical and chemical processes.

Data Availability Statement

The code of Community Earth System Model (CESM, Version 2.1.0) is publicly available online (<https://www.cesm.ucar.edu/models/cesm2/>, CESM Working Groups, 2020). The Coupled Model Intercomparison Project Phase 6 (CMIP6) emission inventories and the historical sea surface temperature (SST) data used by CESM are publicly available online (<https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/atm/cam/>, CESM Working Groups, 2019). The Multi-resolution Emission Inventory for China (MEIC) anthropogenic emissions used by CESM are available online (<http://meicmodel.org/>, Tsinghua University, 2010). The Modern-Era Retrospective analysis for Research and Applications (MERRA2) reanalysis data used by CESM are available online (<https://disc.gsfc.nasa.gov/datasets>, GMAO, 2015). The software Python (Version 3.9.12) is used to make all the Figures in this work and the Python code is available online (<https://www.python.org/downloads/>, Python Software Foundation, 2022). The ML model training/testing codes and related data are available online (<https://doi.org/10.5281/zenodo.4731385>, Zheng, 2022). The XGBoost C functions used to integrate with Fortran are available online (<https://xgboost.readthedocs.io/en/stable/c.html>, XGBoost Developers, 2022). Observation and model results in this paper are available on Zenodo (<https://zenodo.org/record/8363703>, Shen, 2023).

Acknowledgments

This work is supported by the National Natural Science Foundation of China (41925023, 4210050028, 91744208, and 42105069). We also acknowledge funding from DOE Grant DE-SC0022130. We greatly thank the High Performance Computing Center of Nanjing University for providing the computational resources used in this work. We appreciate the support by the Frontiers Science Center for Critical Earth Material Cycling, Nanjing University. We thank the editor Dr. Jiwen Fan for handling the publication of this manuscript, and three anonymous reviewers for their constructive comments on improving the quality of the manuscript.

References

- Abdul-Razzak, H., & Ghan, S. J. (2000). A parameterization of aerosol activation: 2. Multiple aerosol types. *Journal of Geophysical Research*, 105(D5), 6837–6844. <https://doi.org/10.1029/1999jd901161>
- Andreae, M., & Rosenfeld, D. (2008). Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active aerosols. *Earth-Science Reviews*, 89(1–2), 13–41. <https://doi.org/10.1016/j.earscirev.2008.03.001>
- Barré, J., Gaubert, B., Arellano, A. F. J., Worden, H. M., Edwards, D. P., Deeter, M. N., et al. (2015). Assessing the impacts of assimilating IASI and MOPITT CO retrievals using CESM-CAM-chem and DART. *Journal of Geophysical Research: Atmospheres*, 120(19), 10501–10529. <https://doi.org/10.1002/2015JD023467>
- Bauer, S., Wright, D., Koch, D., Lewis, E., McGraw, R., Chang, L.-S., et al. (2008). MATRIX (Multiconfiguration Aerosol TRacker of mIXing state): An aerosol microphysical module for global atmospheric models. *Atmospheric Chemistry and Physics*, 8(20), 6003–6035. <https://doi.org/10.5194/acp-8-6003-2008>

- Bauer, S. E., Ault, A., & Prather, K. A. (2013). Evaluation of aerosol mixing state classes in the GISS modelE-MATRIX climate model using single-particle mass spectrometry measurements. *Journal of Geophysical Research: Atmospheres*, 118(17), 9834–9844. <https://doi.org/10.1002/jgrd.50700>
- Bianchi, F., Tröstl, J., Junninen, H., Frege, C., Henne, S., Hoyle, C. R., et al. (2016). New particle formation in the free troposphere: A question of chemistry and timing. *Science*, 352(6289), 1109–1112. <https://doi.org/10.1126/science.1223447>
- Binkowski, F. S., & Roselle, S. J. (2003). Models-3 community multiscale air quality (CMAQ) model aerosol component 1. Model description. *Journal of Geophysical Research*, 108(D6), 4183. <https://doi.org/10.1029/2001jd001409>
- Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Bernsten, T., DeAngelo, B. J., et al. (2013). Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, 118(11), 5380–5552. <https://doi.org/10.1002/jgrd.50171>
- Cao, L.-M., Wei, J., He, L.-Y., Zeng, H., Li, M.-L., Zhu, Q., et al. (2022). Aqueous aging of secondary organic aerosol coating onto black carbon: Insights from simultaneous L-ToF-AMS and SP-AMS measurements at an urban site in southern China. *Journal of Cleaner Production*, 330, 129888. <https://doi.org/10.1016/j.jclepro.2021.129888>
- Cappa, C. D., Onasch, T. B., Massoli, P., Worsnop, D. R., Bates, T. S., Cross, E. S., et al. (2012). Radiative absorption enhancements due to the mixing state of atmospheric black carbon. *Science*, 337(6098), 1078–1081. <https://doi.org/10.1126/science.1223447>
- Cappa, C. D., Zhang, X., Russell, L. M., Collier, S., Lee, A. K., Chen, C. L., et al. (2019). Light absorption by ambient black and brown carbon and its dependence on black carbon coating state for two California, USA, cities in winter and summer. *Journal of Geophysical Research: Atmospheres*, 124(3), 1550–1577. <https://doi.org/10.1029/2018jd029501>
- CESM Working Groups of National Center for Atmospheric Research. (2019). CESM input data repository: CMIP6 emission inventories and sea surface temperature [Dataset]. National Center for Atmospheric Research. Retrieved from <https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/atm/cam/>
- CESM Working Groups of National Center for Atmospheric Research. Community Earth System Model. (2020). Release (version 2.1.3) [Software]. National Center for Atmospheric Research. Retrieved from <https://www.cesm.ucar.edu/models/cesm2/>
- Chen, L., Zhang, F., Collins, D., Ren, J., Liu, J., Jiang, S., & Li, Z. (2022). Characterizing the volatility and mixing state of ambient fine particles in the summer and winter of urban Beijing. *Atmospheric Chemistry and Physics*, 22(4), 2293–2307. <https://doi.org/10.5194/acp-22-2293-2022>
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM sigkdd International Conference on Knowledge Discovery and Data Mining*. paper presented at.
- Chen, X., Wang, Z., Yu, F., Pan, X., Li, J., Ge, B., et al. (2017). Estimation of atmospheric aging time of black carbon particles in the polluted atmosphere over central-eastern China using microphysical process analysis in regional chemical transport model. *Atmospheric Environment*, 163, 44–56. <https://doi.org/10.1016/j.atmosenv.2017.05.016>
- Ching, J., Fast, J., West, M., & Riemer, N. (2017). Metrics to quantify the importance of mixing state for CCN activity. *Atmospheric Chemistry and Physics*, 17(12), 7445–7458. <https://doi.org/10.5194/acp-17-7445-2017>
- Ching, J., Riemer, N., & West, M. (2012). Impacts of black carbon mixing state on black carbon nucleation scavenging: Insights from a particle-resolved model. *Journal of Geophysical Research*, 117(D23), D23209. <https://doi.org/10.1029/2012jd018269>
- Clarke, A. D., Freitag, S., Simpson, R. M. C., Hudson, J. G., Howell, S. G., Brekhovskikh, V. L., et al. (2013). Free troposphere as a major source of CCN for the equatorial pacific boundary layer: Long-range transport and teleconnections. *Atmospheric Chemistry and Physics*, 13(15), 7511–7529. <https://doi.org/10.5194/acp-13-7511-2013>
- Collier, S., Williams, L. R., Onasch, T. B., Cappa, C. D., Zhang, X., Russell, L. M., et al. (2018). Influence of emissions and aqueous processing on particles containing black carbon in a polluted urban environment: Insights from a soot particle-aerosol mass spectrometer. *Journal of Geophysical Research: Atmospheres*, 123(12), 6648–6666. <https://doi.org/10.1002/2017JD027851>
- Croft, B., Wentworth, G. R., Martin, R. V., Leaitch, W. R., Murphy, J. G., Murphy, B. N., et al. (2016). Contribution of Arctic seabird-colony ammonia to atmospheric particles and cloud-albedo radiative effect. *Nature Communications*, 7(1), 13444. <https://doi.org/10.1038/ncomms13444>
- Cui, S., Huang, D. D., Wu, Y., Wang, J., Shen, F., Xian, J., et al. (2022). Chemical properties, sources and size-resolved hygroscopicity of submicron black-carbon-containing aerosols in urban Shanghai. *Atmospheric Chemistry and Physics*, 22(12), 8073–8096. <https://doi.org/10.5194/acp-22-8073-2022>
- Curtis, J. H., Riemer, N., & West, M. (2017). A single-column particle-resolved model for simulating the vertical distribution of aerosol mixing state: WRF-PartMC-MOSAIC-SCM v1.0, geosci. *Model Dev*, 10(11), 4057–4079. <https://doi.org/10.5194/gmd-10-4057-2017>
- Debry, E., Fahey, K., Sartelet, K., Sportisse, B., & Tombette, M. (2007). A new Size REsolved aerosol model (SIREAM). *Atmospheric Chemistry and Physics*, 7(6), 1537–1547. <https://doi.org/10.5194/acp-7-1537-2007>
- Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., et al. (2006). Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. *Atmospheric Chemistry and Physics*, 6(12), 4321–4344. <https://doi.org/10.5194/acp-6-4321-2006>
- Ding, A., Huang, X., Nie, W., Sun, J., Kerminen, V. M., Petäjä, T., et al. (2016). Enhanced haze pollution by black carbon in megacities in China. *Geophysical Research Letters*, 43(6), 2873–2879. <https://doi.org/10.1002/2016gl067745>
- Ding, S., Liu, D., Zhao, D., Hu, K., Tian, P., Zhou, W., et al. (2019). Size-related physical properties of black carbon in the lower atmosphere over Beijing and Europe. *Environmental Science & Technology*, 53(19), 11112–11121. <https://doi.org/10.1021/acs.est.9b03722>
- Donahue, N. M., Robinson, A., Stanier, C., & Pandis, S. (2006). Coupled partitioning, dilution, and chemical aging of semivolatile organics. *Environmental Science & Technology*, 40(8), 2635–2643. <https://doi.org/10.1021/es052297c>
- Emmons, L. K., Schwantes, R. H., Orlando, J. J., Tyndall, G., Kinnison, D., Lamarque, J., et al. (2020). The chemistry mechanism in the community Earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(4), e2019MS001882. <https://doi.org/10.1029/2019MS001882>
- Fan, T., Liu, X., Ma, P.-L., Zhang, Q., Li, Z., Jiang, Y., et al. (2018). Emission or atmospheric processes? An attempt to attribute the source of large bias of aerosols in eastern China simulated by global climate models. *Atmospheric Chemistry and Physics*, 18(2), 1395–1417. <https://doi.org/10.5194/acp-18-1395-2018>
- Fan, T., Liu, X., Wu, C., Zhang, Q., Zhao, C., Yang, X., & Li, Y. (2022). Comparison of the anthropogenic emission inventory for CMIP6 models with a country-level inventory over China and the simulations of the aerosol properties. *Advances in Atmospheric Sciences*, 39(1), 80–96. <https://doi.org/10.1007/s00376-021-1119-6>
- Fierce, L., Bond, T. C., Bauer, S. E., Mena, F., & Riemer, N. (2016). Black carbon absorption at the global scale is affected by particle-scale diversity in composition. *Nature Communications*, 7(1), 12361. <https://doi.org/10.1038/ncomms12361>

- Fierce, L., Onasch, T. B., Cappa, C. D., Mazzoleni, C., China, S., Bhandari, J., et al. (2020). Radiative absorption enhancements by black carbon controlled by particle-to-particle heterogeneity in composition. *Proceedings of the National Academy of Sciences of the United States of America*, 117(10), 5196–5203. <https://doi.org/10.1073/pnas.1919723117>
- Ghan, S. J., & Easter, R. C. (2006). Impact of cloud-borne aerosol representation on aerosol direct and indirect effects. *Atmospheric Chemistry and Physics*, 6(12), 4163–4174. <https://doi.org/10.5194/acp-6-4163-2006>
- Ghan, S. J., Leung, L. R., Easter, R. C., & Abdul-Razzak, H. (1997). Prediction of cloud droplet number in a general circulation model. *Journal of Geophysical Research*, 102(D18), 21777–21794. <https://doi.org/10.1029/97jd01810>
- Global Modeling, & Assimilation Office (GMAO). (2015). The modern-era retrospective analysis for research and applications reanalysis [Dataset]. Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC). Retrieved from <https://disc.gsfc.nasa.gov/datasets>
- Grahame, T. J., Klemm, R., & Schlesinger, R. B. (2014). Public health and components of particulate matter: The changing assessment of black carbon. *Journal of the Air & Waste Management Association*, 64(6), 620–660. <https://doi.org/10.1080/10962247.2014.912692>
- Grange, S. K., Carslaw, D. C., Lewis, A. C., Boletí, E., & Hueglin, C. (2018). Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmospheric Chemistry and Physics*, 18(9), 6223–6239. <https://doi.org/10.5194/acp-18-6223-2018>
- Guo, S., Hu, M., Lin, Y., Gomez-Hernandez, M., Zamora, M. L., Peng, J., et al. (2016). OH-initiated oxidation of m-xylene on black carbon aging. *Environmental Science & Technology*, 50(16), 8605–8612. <https://doi.org/10.1021/acs.est.6b01272>
- Healy, R. M., Wang, J. M., Jeong, C. H., Lee, A. K., Willis, M. D., Jaroudi, E., et al. (2015). Light-absorbing properties of ambient black carbon and brown carbon from fossil fuel and biomass burning sources. *Journal of Geophysical Research: Atmospheres*, 120(13), 6619–6633. <https://doi.org/10.1002/2015jd023382>
- Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, 11(1), 369–408. <https://doi.org/10.5194/gmd-11-369-2018>
- Holanda, B. A., Pöhlker, M. L., Walter, D., Saturno, J., Sörgel, M., Ditas, J., et al. (2020). Influx of African biomass burning aerosol during the Amazonian dry season through layered transatlantic transport of black carbon-rich smoke. *Atmospheric Chemistry and Physics*, 20(8), 4757–4785. <https://doi.org/10.5194/acp-20-4757-2020>
- Hughes, M., Kodros, J. K., Pierce, J. R., West, M., & Riemer, N. (2018). Machine learning to predict the global distribution of aerosol mixing state metrics. *Atmosphere*, 9(1), 15. <https://doi.org/10.3390/atmos9010015>
- Jacobson, M. Z. (2002). Analysis of aerosol interactions with numerical techniques for solving coagulation, nucleation, condensation, dissolution, and reversible chemistry among multiple size distributions. *Journal of Geophysical Research*, 107(D19), AAC2-1–AAC2-23. <https://doi.org/10.1029/2001jd002044>
- Kelp, M., Jacob, D. J., Lin, H., & Sulprizio, M. P. (2022). An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry. *Journal of Advances in Modeling Earth Systems*, 14(6), e2021MS002926. <https://doi.org/10.1029/2021MS002926>
- Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., & Tessum, C. W. (2020). Toward stable, general machine-learned models of the atmospheric chemical system. *Journal of Geophysical Research: Atmospheres*, 125(23), e2020JD032759. <https://doi.org/10.1029/2020JD032759>
- Kerminen, V.-M., Chen, X., Vakkari, V., Petäjä, T., Kulmala, M., & Bianchi, F. (2018). Atmospheric new particle formation and growth: Review of field observations. *Environmental Research Letters*, 13(10), 103003. <https://doi.org/10.1088/1748-9326/aadf3c>
- Ko, J., Krasowsky, T., & Ban-Weiss, G. (2020). Measurements to determine the mixing state of black carbon emitted from the 2017–2018 California wildfires and urban Los Angeles. *Atmospheric Chemistry and Physics*, 20(24), 15635–15664. <https://doi.org/10.5194/acp-20-15635-2020>
- Koch, D. (2001). Transport and direct radiative forcing of carbonaceous and sulfate aerosols in the GISS GCM. *Journal of Geophysical Research*, 106(D17), 20311–20332. <https://doi.org/10.1029/2001jd900038>
- Kodros, J. K., Hanna, S. J., Bertram, A. K., Leaitch, W. R., Schulz, H., Herber, A. B., et al. (2018). Size-resolved mixing state of black carbon in the Canadian high Arctic and implications for simulated direct radiative effect. *Atmospheric Chemistry and Physics*, 18(15), 11345–11361. <https://doi.org/10.5194/acp-18-11345-2018>
- Kompalli, S. K., Babu, S. N. S., Moorthy, K. K., Satheesh, S. K., Gogoi, M. M., Nair, V. S., et al. (2021). Mixing state of refractory black carbon aerosol in the South Asian outflow over the northern Indian Ocean during winter. *Atmospheric Chemistry and Physics*, 21(11), 9173–9199. <https://doi.org/10.5194/acp-21-9173-2021>
- Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C., & Russell, L. M. (2012). Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5. *Journal of Geophysical Research*, 117(D23). <https://doi.org/10.1029/2012jd018588>
- Lee, A. K., Chen, C.-L., Liu, J., Price, D. J., Betha, R., Russell, L. M., et al. (2017a). Formation of secondary organic aerosol coating on black carbon particles near vehicular emissions. *Atmospheric Chemistry and Physics*, 17(24), 15055–15067. <https://doi.org/10.5194/acp-17-15055-2017>
- Lee, A. K., Rivellini, L.-H., Chen, C.-L., Liu, J., Price, D. J., Betha, R., et al. (2019). Influences of primary emission and secondary coating formation on the particle diversity and mixing state of black carbon particles. *Environmental Science & Technology*, 53(16), 9429–9438. <https://doi.org/10.1021/acs.est.9b03064>
- Lee, A. K. Y., Chen, C. L., Liu, J., Price, D. J., Betha, R., Russell, L. M., et al. (2017b). Formation of secondary organic aerosol coating on black carbon particles near vehicular emissions. *Atmospheric Chemistry and Physics*, 17(24), 15055–15067. <https://doi.org/10.5194/acp-17-15055-2017>
- Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., et al. (2017). Anthropogenic emission inventories in China: A review. *National Science Review*, 4(6), 834–866. <https://doi.org/10.1093/nsr/nwx150>
- Li, T., Wang, M., Guo, Z., Yang, B., Xu, Y., Han, X., & Sun, J. (2022). An updated CLUBB PDF closure scheme to improve low cloud simulation in CAM6. *Journal of Advances in Modeling Earth Systems*, 14(12), e2022MS003127. <https://doi.org/10.1029/2022ms003127>
- Liu, D., Ding, S., Zhao, D., Hu, K., Yu, C., Hu, D., et al. (2020). Black carbon emission and wet scavenging from surface to the top of boundary layer over Beijing region. *Journal of Geophysical Research: Atmospheres*, 125(17), e2020JD033096. <https://doi.org/10.1029/2020jd033096>
- Liu, D., He, C., Schwarz, J. P., & Wang, X. (2020). Lifecycle of light-absorbing carbonaceous aerosols in the atmosphere. *NPJ Climate and Atmospheric Science*, 3(1), 40. <https://doi.org/10.1038/s41612-020-00145-8>
- Liu, D., Whitehead, J., Alfara, M. R., Reyes-Villegas, E., Spracklen, D. V., Reddington, C. L., et al. (2017). Black-carbon absorption enhancement in the atmosphere determined by particle mixing state. *Nature Geoscience*, 10(3), 184–188. <https://doi.org/10.1038/ngeo2901>
- Liu, Q., & Wang, Q. (2017). How China achieved its 11th five-year plan emissions reduction target: A structural decomposition analysis of industrial SO₂ and chemical oxygen demand. *Science of the Total Environment*, 574, 1104–1116. <https://doi.org/10.1016/j.scitotenv.2016.08.176>

- Liu, S., Aiken, A. C., Gorkowski, K., Dubey, M. K., Cappa, C. D., Williams, L. R., et al. (2015). Enhanced light absorption by mixed source black and brown carbon particles in UK winter. *Nature Communications*, 6(1), 8435. <https://doi.org/10.1038/ncomms9435>
- Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., et al. (2012). Toward a minimal representation of aerosols in climate models: Description and evaluation in the Community Atmosphere Model CAM5. *Geoscientific Model Development*, 5(3), 709–739. <https://doi.org/10.5194/gmd-5-709-2012>
- Liu, X., Ma, P.-L., Wang, H., Tilmes, S., Singh, B., Easter, R., et al. (2016). Description and evaluation of a new four-mode version of the modal aerosol module (MAM4) within version 5.3 of the community atmosphere model. *Geoscientific Model Development*, 9(2), 505–522. <https://doi.org/10.5194/gmd-9-505-2016>
- Liu, X., Penner, J. E., & Herzog, M. (2005). Global modeling of aerosol dynamics: Model description, evaluation, and interactions between sulfate and nonsulfate aerosols. *Journal of Geophysical Research*, 110(D18). <https://doi.org/10.1029/2004JD005674>
- Mårtensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H., & Hansson, H.-C. (2003). Laboratory simulations and parameterization of the primary marine aerosol production. *Journal of Geophysical Research*, 108(D9). <https://doi.org/10.1029/2002JD002263>
- Massoli, P., Onasch, T. B., Cappa, C. D., Nuamaan, I., Hakala, J., Hayden, K., et al. (2015b). Characterization of black carbon-containing particles from soot particle aerosol mass spectrometer measurements on the R/V Atlantis during CalNex 2010. *Journal of Geophysical Research: Atmospheres*, 120(6), 2575–2593. <https://doi.org/10.1002/2014JD022834>
- Massoli, P., Onasch, T. B., Cappa, C. D., Nuamaan, I., Hakala, J., Hayden, K., et al. (2015a). Characterization of black carbon-containing particles from soot particle aerosol mass spectrometer measurements on the R/V Atlantis during CalNex 2010. *Journal of Geophysical Research: Atmospheres*, 120(6), 2575–2593. <https://doi.org/10.1002/2014jd022834>
- Matsui, H., Koike, M., Kondo, Y., Moteki, N., Fast, J. D., & Zaveri, R. A. (2013). Development and validation of a black carbon mixing state resolved three-dimensional model: Aging processes and radiative impact. *Journal of Geophysical Research: Atmospheres*, 118(5), 2304–2326. <https://doi.org/10.1029/2012jd018446>
- McMeeking, G., Fortner, E., Onasch, T., Taylor, J., Flynn, M., Coe, H., & Kreidenweis, S. (2014). Impacts of nonrefractory material on light absorption by aerosols emitted from biomass burning. *Journal of Geophysical Research: Atmospheres*, 119(21), 12272–12286. <https://doi.org/10.1002/2014jd021750>
- Metcalfe, A., Craven, J., Ensberg, J., Brioude, J., Angevine, W., Sorooshian, A., et al. (2012). Black carbon aerosol over the Los Angeles basin during CalNex. *Journal of Geophysical Research*, 117(D21). <https://doi.org/10.1029/2011jd017255>
- Moffet, R. C., O'Brien, R. E., Alpert, P. A., Kelly, S. T., Pham, D. Q., Gilles, M. K., et al. (2016). Morphology and mixing of black carbon particles collected in central California during the CARES field study. *Atmospheric Chemistry and Physics*, 16(22), 14515–14525. <https://doi.org/10.5194/acp-16-14515-2016>
- Monge, M. E., D'Anna, B., Mazri, L., Giroir-Fendler, A., Ammann, M., Donaldson, D., & George, C. (2010). Light changes the atmospheric reactivity of soot. *Proceedings of the National Academy of Sciences*, 107(15), 6605–6609. <https://doi.org/10.1073/pnas.0908341107>
- Moteki, N., Kondo, Y., Oshima, N., Takegawa, N., Koike, M., Kita, K., et al. (2012). Size dependence of wet removal of black carbon aerosols during transport from the boundary layer to the free troposphere. *Geophysical Research Letters*, 39(13). <https://doi.org/10.1029/2012gl052034>
- Moteki, N., Mori, T., Matsui, H., & Ohata, S. (2019). Observational constraint of in-cloud supersaturation for simulations of aerosol rainout in atmospheric models. *Npj Climate and Atmospheric Science*, 2(1), 6. <https://doi.org/10.1038/s41612-019-0063-y>
- Ohata, S., Koike, M., Yoshida, A., Moteki, N., Adachi, K., Oshima, N., et al. (2021). Arctic black carbon during PAMARCMIP 2018 and previous aircraft experiments in spring. *Atmospheric Chemistry and Physics*, 21(20), 15861–15881. <https://doi.org/10.5194/acp-21-15861-2021>
- Ohata, S., Mori, T., Kondo, Y., Sharma, S., Hyvärinen, A., Andrews, E., et al. (2021). Estimates of mass absorption cross sections of black carbon for filter-based absorption photometers in the Arctic. *Atmospheric Measurement Techniques*, 14(10), 6723–6748. <https://doi.org/10.5194/amt-14-6723-2021>
- Ohata, S., Moteki, N., Mori, T., Koike, M., & Kondo, Y. (2016). A key process controlling the wet removal of aerosols: New observational evidence. *Scientific Reports*, 6(1), 34113. <https://doi.org/10.1038/srep34113>
- Ovtchinnikov, M., & Ghan, S. J. (2005). Parallel simulations of aerosol influence on clouds using cloud-resolving and single-column models. *Journal of Geophysical Research*, 110(D15). <https://doi.org/10.1029/2004jd005088>
- Paulot, F., Paynter, D., Ginoux, P., Naik, V., & Horowitz, L. W. (2018). Changes in the aerosol direct radiative forcing from 2001 to 2015: Observational constraints and regional mechanisms. *Atmospheric Chemistry and Physics*, 18(17), 13265–13281. <https://doi.org/10.5194/acp-18-13265-2018>
- Peng, J., Hu, M., Guo, S., Du, Z., Zheng, J., Shang, D., et al. (2016). Markedly enhanced absorption and direct radiative forcing of black carbon under polluted urban environments. *Proceedings of the National Academy of Sciences*, 113(16), 4266–4271. <https://doi.org/10.1073/pnas.1602310113>
- Pósfai, M., Anderson, J. R., Buseck, P. R., & Sievering, H. (1999). Soot and sulfate aerosol particles in the remote marine troposphere. *Journal of Geophysical Research*, 104(D17), 21685–21693. <https://doi.org/10.1029/1999jd900208>
- Raatikainen, T., Brus, D., Hyvärinen, A., Svensson, J. M., Asmi, E., & Lihavainen, H. (2015). Black carbon concentrations and mixing state in the Finnish Arctic. *Atmospheric Chemistry and Physics*, 15(17), 10057–10070. <https://doi.org/10.5194/acp-15-10057-2015>
- Ramanathan, V., & Carmichael, G. (2008). Global and regional climate changes due to black carbon. *Nature Geoscience*, 1(4), 221–227. <https://doi.org/10.1038/ngeo156>
- Reddington, C. L., McMeeking, G., Mann, G. W., Coe, H., Frontoso, M. G., Liu, D., et al. (2013). The mass and number size distributions of black carbon aerosol over Europe. *Atmospheric Chemistry and Physics*, 13(9), 4917–4939. <https://doi.org/10.5194/acp-13-4917-2013>
- Riener, N., Ault, A., West, M., Craig, R., & Curtis, J. (2019). Aerosol mixing state: Measurements, modeling, and impacts. *Reviews of Geophysics*, 57(2), 187–249. <https://doi.org/10.1029/2018rg000615>
- Riener, N., & West, M. (2013). Quantifying aerosol mixing state with entropy and diversity measures. *Atmospheric Chemistry and Physics*, 13(22), 11423–11439. <https://doi.org/10.5194/acp-13-11423-2013>
- Riener, N., West, M., Zaveri, R. A., & Easter, R. C. (2009). Simulating the evolution of soot mixing state with a particle-resolved aerosol model. *Journal of Geophysical Research*, 114(D9). <https://doi.org/10.1029/2008jd011073>
- Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., et al. (2007). Rethinking organic aerosols: Semivolatile emissions and photochemical aging. *Science*, 315(5816), 1259–1262. <https://doi.org/10.1126/science.1133061>
- Ruan, Z., Sun, W., Yuan, Y., & Tan, H. (2023). Accurately forecasting solar radiation distribution at both spatial and temporal dimensions simultaneously with fully-convolutional deep neural network model. *Renewable and Sustainable Energy Reviews*, 184, 113528. <https://doi.org/10.1016/j.rser.2023.113528>
- Sahu, L., Kondo, Y., Moteki, N., Takegawa, N., Zhao, Y., Cubison, M., et al. (2012). Emission characteristics of black carbon in anthropogenic and biomass burning plumes over California during ARCTAS-CARB 2008. *Journal of Geophysical Research*, 117(D16). <https://doi.org/10.1029/2011jd017401>

- Sartelet, K., Hayami, H., Albriet, B., & Sportisse, B. (2006). Development and preliminary validation of a modal aerosol model for tropospheric chemistry: MAM. *Aerosol Science and Technology*, 40(2), 118–127. <https://doi.org/10.1080/02786820500485948>
- Schill, G. P., DeMott, P. J., Emerson, E. W., Rauker, A. M. C., Kodros, J. K., Suski, K. J., et al. (2020). The contribution of black carbon to global ice nucleating particle concentrations relevant to mixed-phase clouds. *Proceedings of the National Academy of Sciences*, 117(37), 22705–22711. <https://doi.org/10.1073/pnas.2001674117>
- Schulz, H., Zannata, M., Bozem, H., Leaitch, W. R., Herber, A. B., Burkart, J., et al. (2019). High Arctic aircraft measurements characterising black carbon vertical variability in spring and summer. *Atmospheric Chemistry and Physics*, 19(4), 2361–2384. <https://doi.org/10.5194/acp-19-2361-2019>
- Schwarz, J. P., Gao, R., Spackman, J., Watts, L., Thomson, D., Fahey, D., et al. (2008). Measurement of the mixing state, mass, and optical size of individual black carbon particles in urban and biomass burning emissions. *Geophysical Research Letters*, 35(13). <https://doi.org/10.1029/2008gl033968>
- Shan, Y., Liu, X., Lin, L., Ke, Z., & Lu, Z. (2021). An improved representation of aerosol wet removal by deep convection and impacts on simulated aerosol vertical profiles. *Journal of Geophysical Research: Atmospheres*, 126(13), e2020JD034173. <https://doi.org/10.1029/2020jd034173>
- Sharma, S., Leaitch, W. R., Huang, L., Veber, D., Kolonjari, F., Zhang, W., et al. (2017). An evaluation of three methods for measuring black carbon in Alert, Canada. *Atmospheric Chemistry and Physics*, 17(24), 15225–15243. <https://doi.org/10.5194/acp-17-15225-2017>
- Shen, W. (2023). Observation and model data for this paper: R_{BC} [Dataset]. Zenodo. Retrieved from <https://zenodo.org/record/8363703>
- Shen, W., Wang, M., Liu, Y., Dong, X., Zhao, D., Yue, M., et al. (2023). Evaluating BC aging processes in the community atmosphere model version 6 (CAM6). *Journal of Geophysical Research: Atmospheres*, 128(3), e2022JD037427. <https://doi.org/10.1029/2022JD037427>
- Slowik, J. G., Cross, E. S., Han, J.-H., Kolucki, J., Davidovits, P., Williams, L. R., et al. (2007). Measurements of morphology changes of fractal soot particles using coating and denuding experiments: Implications for optical absorption and atmospheric lifetime. *Aerosol Science and Technology*, 41(8), 734–750. <https://doi.org/10.1080/02786820701432632>
- Solmon, F., Giorgi, F., & Liousse, C. (2006). Aerosol modelling for regional climate studies: Application to anthropogenic particles and evaluation over a European/African domain. *Tellus B: Chemical and Physical Meteorology*, 58(1), 51–72. <https://doi.org/10.1111/j.1600-0889.2005.00155.x>
- Spracklen, D., Pringle, K., Carslaw, K., Chipperfield, M., & Mann, G. (2005). A global off-line model of size-resolved aerosol microphysics: I. Model development and prediction of aerosol properties. *Atmospheric Chemistry and Physics*, 5(8), 2227–2252. <https://doi.org/10.5194/acp-5-2227-2005>
- Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., et al. (2005). The aerosol-climate model ECHAM5-HAM. *Atmospheric Chemistry and Physics*, 5(4), 1125–1156. <https://doi.org/10.5194/acp-5-1125-2005>
- Sturm, P. O., Manders, A., Janssen, R., Segers, A., Wexler, A. S., & Lin, H. X. (2023). Advecting superspecies: Efficiently modeling transport of organic aerosol with a mass-conserving dimensionality reduction method. *Journal of Advances in Modeling Earth Systems*, 15(3), e2022MS003235. <https://doi.org/10.1029/2022ms003235>
- Sturm, P. O., & Wexler, A. S. (2020). A mass-and energy-conserving framework for using machine learning to speed computations: A photochemistry example. *Geoscientific Model Development*, 13(9), 4435–4442. <https://doi.org/10.5194/gmd-13-4435-2020>
- Tilmes, S., Hodzic, A., Emmons, L. K., Mills, M. J., Gettelman, A., Kinnison, D. E., et al. (2019). Climate forcing and trends of organic aerosols in the community Earth system model (CESM2). *Journal of Advances in Modeling Earth Systems*, 11(12), 4323–4351. <https://doi.org/10.1029/2019MS001827>
- Tritscher, T., Jurányi, Z., Martin, M., Chirico, R., Gysel, M., Hering, M. F., et al. (2011). Changes of hygroscopicity and morphology during ageing of diesel soot. *Environmental Research Letters*, 6(3), 034026. <https://doi.org/10.1088/1748-9326/6/3/034026>
- Van Marle, M. J., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., et al. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, 10(9), 3329–3357. <https://doi.org/10.5194/gmd-10-3329-2017>
- Wang, H., Easter, R. C., Rasch, P. J., Wang, M., Liu, X., Ghan, S. J., et al. (2013). Sensitivity of remote aerosol distributions to representation of cloud–aerosol interactions in a global climate model. *Geoscientific Model Development*, 6(3), 765–782. <https://doi.org/10.5194/gmd-6-765-2013>
- Wang, J., Liu, D., Ge, X., Wu, Y., Shen, F., Chen, M., et al. (2019b). Characterization of black carbon-containing fine particles in Beijing during wintertime. *Atmospheric Chemistry and Physics*, 19(1), 447–458. <https://doi.org/10.5194/acp-19-447-2019>
- Wang, J., Liu, D., Ge, X., Wu, Y., Shen, F., Chen, M., et al. (2019a). Characterization of black carbon-containing fine particles in Beijing during wintertime. *Atmospheric Chemistry and Physics*, 19(1), 447–458. <https://doi.org/10.5194/acp-19-447-2019>
- Wang, J., Zhang, Q., Chen, M., Collier, S., Zhou, S., Ge, X., et al. (2017). First chemical characterization of refractory black carbon aerosols and associated coatings over the Tibetan Plateau (4730 m asl). *Environmental Science & Technology*, 51(24), 14072–14082. <https://doi.org/10.1021/acs.est.7b03973>
- Wentworth, G. R., Murphy, J. G., Croft, B., Martin, R. V., Pierce, J. R., Côté, J.-S., et al. (2016). Ammonia in the summertime Arctic marine boundary layer: Sources, sinks, and implications. *Atmospheric Chemistry and Physics*, 16(4), 1937–1953. <https://doi.org/10.5194/acp-16-1937-2016>
- Williamson, C. J., Kupc, A., Axisa, D., Bilsback, K. R., Bui, T., Campuzano-Jost, P., et al. (2019). A large source of cloud condensation nuclei from new particle formation in the tropics. *Nature*, 574(7778), 399–403. <https://doi.org/10.1038/s41586-019-1638-9>
- Willis, M. D., Bozem, H., Kunkel, D., Lee, A. K., Schulz, H., Burkart, J., et al. (2019). Aircraft-based measurements of High Arctic springtime aerosol show evidence for vertically varying sources, transport and composition. *Atmospheric Chemistry and Physics*, 19(1), 57–76. <https://doi.org/10.5194/acp-19-57-2019>
- Willis, M. D., Burkart, J., Thomas, J. L., Köllner, F., Schneider, J., Bozem, H., et al. (2016). Growth of nucleation mode particles in the summertime Arctic: A case study. *Atmospheric Chemistry and Physics*, 16(12), 7663–7679. <https://doi.org/10.5194/acp-16-7663-2016>
- Wu, Y., Liu, D., Wang, J., Shen, F., Chen, Y., Cui, S., et al. (2019). Characterization of size-resolved hygroscopicity of black carbon-containing particle in urban environment. *Environmental Science & Technology*, 53(24), 14212–14221. <https://doi.org/10.1021/acs.est.9b05546>
- Wu, Z., Hu, L., Guo, T., Lin, T., & Guo, Z. (2020). Aeolian transport and deposition of carbonaceous aerosols over the Northwest Pacific Ocean in spring. *Atmospheric Environment*, 223, 117209. <https://doi.org/10.1016/j.atmosenv.2019.117209>
- XGBoost Developers. (2022). The XGBoost C functions used to integrate with Fortran in this paper [Software]. Zenodo. Retrieved from <https://xgboost.readthedocs.io/en/stable/c.html>
- Xie, C., Xu, W., Wang, J., Liu, D., Ge, X., Zhang, Q., et al. (2019). Light absorption enhancement of black carbon in urban Beijing in summer. *Atmospheric Environment*, 213, 499–504. <https://doi.org/10.1016/j.atmosenv.2019.06.041>

- Yu, C., Liu, D., Broda, K., Joshi, R., Olfert, J., Sun, Y., et al. (2020). Characterising mass-resolved mixing state of black carbon in Beijing using a morphology-independent measurement method. *Atmospheric Chemistry and Physics*, 20(6), 3645–3661. <https://doi.org/10.5194/acp-20-3645-2020>
- Yu, F., Luo, G., Nair, A. A., Tsigaridis, K., & Bauer, S. E. (2022). Use of machine learning to reduce uncertainties in particle number concentration and aerosol indirect radiative forcing predicted by climate models. *Geophysical Research Letters*, 49(16), e2022GL098551. <https://doi.org/10.1029/2022GL098551>
- Yu, P., Froyd, K. D., Portmann, R. W., Toon, O. B., Freitas, S. R., Bardeen, C. G., et al. (2019). Efficient in-cloud removal of aerosols by deep convection. *Geophysical Research Letters*, 46(2), 1061–1069. <https://doi.org/10.1029/2018GL080544>
- Yuan, L., & Zhao, C. (2023). Quantifying particle-to-particle heterogeneity in aerosol hygroscopicity. *Atmospheric Chemistry and Physics*, 23(5), 3195–3205. <https://doi.org/10.5194/acp-23-3195-2023>
- Zhai, J., Yang, X., Li, L., Bai, B., Liu, P., Huang, Y., et al. (2022). Absorption enhancement of black carbon aerosols constrained by mixing-state heterogeneity. *Environmental Science & Technology*, 56(3), 1586–1593. <https://doi.org/10.1021/acs.est.1c06180>
- Zhang, R., Khalizov, A. F., Pagels, J., Zhang, D., Xue, H., & McMurry, P. H. (2008). Variability in morphology, hygroscopicity, and optical properties of soot aerosols during atmospheric processing. *Proceedings of the National Academy of Sciences*, 105(30), 10291–10296. <https://doi.org/10.1073/pnas.0804860105>
- Zhang, Y., Su, H., Kecorius, S., Wang, Z., Hu, M., Zhu, T., et al. (2017). Mixing state of refractory black carbon of the North China plain regional aerosol combining a single particle soot photometer and a volatility tandem differential mobility analyzer. *Atmospheric Chemistry and Physics Discussions*, 1–27.
- Zhao, B., Shrivastava, M., Donahue, N. M., Gordon, H., Schervish, M., Shilling, J. E., et al. (2020). High concentration of ultrafine particles in the Amazon free troposphere produced by organic new particle formation. *Proceedings of the National Academy of Sciences*, 117(41), 25344–25351. <https://doi.org/10.1073/pnas.2006716117>
- Zhao, D., Huang, M., Tian, P., He, H., Lowe, D., Zhou, W., et al. (2019). Vertical characteristics of black carbon physical properties over Beijing region in warm and cold seasons. *Atmospheric Environment*, 213, 296–310. <https://doi.org/10.1016/j.atmosenv.2019.06.007>
- Zhao, D., Liu, D., Yu, C., Tian, P., Hu, D., Zhou, W., et al. (2020). Vertical evolution of black carbon characteristics and heating rate during a haze event in Beijing winter. *Science of the Total Environment*, 709, 136251. <https://doi.org/10.1016/j.scitotenv.2019.136251>
- Zhao, G., Tan, T., Zhu, Y., Hu, M., & Zhao, C. (2021). Method to quantify black carbon aerosol light absorption enhancement with a mixing state index. *Atmospheric Chemistry and Physics*, 21(23), 18055–18063. <https://doi.org/10.5194/acp-21-18055-2021>
- Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., et al. (2018). Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmospheric Chemistry and Physics*, 18(19), 14095–14111. <https://doi.org/10.5194/acp-18-14095-2018>
- Zheng, Z. (2022). The training/testing codes and related data for the machine learning model used in this paper [Software]. Zenodo. <https://doi.org/10.5281/zenodo.4731385>
- Zheng, Z., Curtis, J. H., Yao, Y., Gasparik, J. T., Anantharaj, V. G., Zhao, L., et al. (2021b). Estimating submicron aerosol mixing state at the global scale with machine learning and Earth system modeling. *Earth and Space Science*, 8(2), e2020EA001500. <https://doi.org/10.1029/2020ea001500>
- Zheng, Z., Fiore, A. M., Westervelt, D. M., Milly, G. P., Goldsmith, J., Karambelas, A., et al. (2023). Automated machine learning to evaluate the information content of tropospheric trace gas columns for fine particle estimates over India: A modeling testbed. *Journal of Advances in Modeling Earth Systems*, 15(3), e2022MS003099. <https://doi.org/10.1029/2022MS003099>
- Zheng, Z., West, M., Zhao, L., Ma, P.-L., Liu, X., & Riener, N. (2021a). Quantifying the structural uncertainty of the aerosol mixing state representation in a modal model. *Atmospheric Chemistry and Physics*, 21(23), 17727–17741. <https://doi.org/10.5194/acp-21-17727-2021>
- Zhu, Q., Liu, Y., Shao, T., & Tang, Y. (2020). Transport of asian aerosols to the Pacific Ocean. *Atmospheric Research*, 234, 104735. <https://doi.org/10.1016/j.atmosres.2019.104735>
- Zhu, S., Sartelet, K., Zhang, Y., & Nenes, A. (2016). Three-dimensional modeling of the mixing state of particles over Greater Paris. *Journal of Geophysical Research: Atmospheres*, 121(10), 5930–5947. <https://doi.org/10.1002/2015jd024241>