

Deep learning representation of the aerosol size distribution

Donifan Barahona¹, Katherine H. Breen^{1,2}, Karoline Block³, and Anton Darmenov¹

¹NASA, Goddard Space Flight Center, Greenbelt, MD, USA

²Morgan State University, Baltimore, MD, USA

³Leipzig Institute for Meteorology, Faculty of Physics and Earth Sciences, University of Leipzig, Leipzig, Germany

Correspondence: Donifan Barahona (donifan.o.barahona@nasa.gov)

Abstract. Aerosols influence Earth’s radiative balance via the scattering and absorbing of solar radiation, affect cloud formation, and play important roles on precipitation, ocean seeding and human health. Accurate modeling of these effects requires knowledge of the ~~the~~ chemical composition and size distribution of aerosol particles present in the atmosphere. Computationally intensive applications like remote sensing and weather forecasting commonly use simplified representations of aerosol microphysics, prescribing the aerosol size distribution (ASD), introducing uncertainty in climate predictions and aerosol retrievals. ~~This work develops~~ In this work, we develop a neural network model, ~~termed~~ MAMnet, to predict the ASD and mixing state ~~using the bulk mass of aerosol for seven lognormal modes based on the bulk aerosol mass~~ and the meteorological state. MAMnet ~~can be driven by the output of single moment~~ is designed to operate with outputs from single-moment, mass-based ~~aerosol schemes or using reanalysis products. We show that MAMnet is able to~~ aerosol schemes, making it compatible with ~~existing models. We demonstrate that MAMnet can~~ accurately reproduce the ~~predictions output~~ of a two-moment ~~microphysics aerosol model as well as field measurements~~ modal aerosol scheme, and also agrees well with field measurements when driven by reanalysis data. Our model paves the way to improve the ~~physical~~ representation of aerosols in ~~physical~~ atmospheric models while maintaining the versatility and efficiency required in large scale applications.

1 Introduction

Aerosols play a crucial role in the Earth’s ~~climate~~ system by influencing radiative forcing (Forster et al., 2007; Bender, 2020), cloud formation and lifetime (Christensen et al., 2020), and precipitation patterns (Stier et al., 2024). Aerosol particle size and composition determine their atmospheric lifetime (Seinfeld and Pandis, 2016), impact on human health (Arfin et al., 2023), long range transport (Uno et al., 2009), and their ability to become cloud droplets and ice crystals (Seinfeld et al., 2016). The size and composition of atmospheric aerosols are critical parameters determining the concentration of cloud condensation nuclei (CCN) in the atmosphere (Seinfeld and Pandis, 2016). Understanding the distribution and composition of atmospheric aerosols is thus essential for accurate climate and weather simulations (Seinfeld et al., 2016).

The ~~representation of the aerosol size distribution (ASD)~~ ASD and mixing state ~~is are~~ at the center of the ability of climate models to accurately simulate the transport and chemical evolution of aerosol species (Aquila et al., 2011; Bender et al., 2019). Variability in the representation of the ASD among models has been shown to drive large differences in cloud droplet number concentration and aerosol–cloud radiative forcing (Virtanen et al., 2025). Explicitly resolving the ASD improves the

representation of nucleation, condensation, and coagulation processes (Zhou et al., 2018), and it is critical for realistically simulating scavenging within clouds, as smaller particles are less efficiently removed than larger ones, affecting global particle number concentrations by up to 20% (Pierce et al., 2015). It has been shown that models that resolve particle-level mixing state and size better represent CCN activity, aerosol aging, and radiative properties (Riemer et al., 2019).

30 Atmospheric models represent the ASD using approximations with different degrees of sophistication. The bulk mass approach ~~aggregates aerosol species into a limited number of prescribed bins based on their mass and composition~~ predicts the transport and evolution of aerosols by tracking the mass concentration of individual chemical species (Jones et al., 1994; Langner and Rodhe, 1991; Ginoux et al., 2001; Chin et al., 2000). ~~This approach inherently assumes externally mixed components, that is, all particles of the same size have the same composition (Seinfeld and Pandis, 2016). Bulk schemes offer low computational~~
35 ~~expense, but they fail to distinguish between processes affecting number concentration and mass, hence cannot explicitly predict the ASD.~~ It inherently treats aerosols as externally mixed, since each particle is assumed to consist of a single chemical component or their surrogate (Riemer et al., 2019). Because each species is typically represented by a single prognostic variable, the bulk approach is not designed to resolve the ASD or the mixing state, which are often prescribed from climatological data. However, ~~they are amenable to data assimilation methods (Randles et al., 2017), and are often~~ due to their low computational
40 cost, bulk schemes are well suited for data assimilation (Randles et al., 2017) and are widely used in forecasting systems, satellite ~~retrievals~~ retrieval algorithms, and reanalysis products (Chu et al., 2002; Gelaro et al., 2017; Inness et al., 2019). For example, ~~the aerosol transport in the MERRA-2 climate reanalysis (Gelaro et al., 2017; Randles et al., 2017) is based on the Goddard Chemistry, Aerosol, Radiation, and Transport model, GOCART (Colarco et al., 2010b), (GOCART) (Colarco et al., 2010b).~~ GOCART is a bulk aerosol scheme that explicitly calculates the mass of major species—dust, black carbon, organic material,
45 sea salt, and sulfate ~~using a bulk, externally-mixed approach, and forms the basis of aerosol assimilation in the MERRA-2 climate reanalysis (Gelaro et al., 2017; Randles et al., 2017)—~~ using an externally mixed representation.

In contrast to bulk methods, modal aerosol schemes estimate both the number concentration and mass of atmospheric aerosol, ~~hence two moments of the ASD (e.g., Whitby and McMurry, 1997; Wilson et al., 2001; Stier et al., 2005; Liu et al., 2012)~~ approximating the ASD as the combination of overlapping populations, termed “modes”, each typically assumed internally-mixed, and
50 following a log-normal distribution with prescribed geometric standard deviation (e.g., Whitby and McMurry, 1997; Wilson et al., 2001; Stier et al., 2005). Because they predict the number concentration and mass independently, modal schemes can ~~handle internally-mixed aerosol species, where aerosol species partition between different modes, better approximating the ASD (Wilson et al., 2001; Herzog et al., 2004; Stier et al., 2005).~~ This leads to a better representation of aerosol-cloud interactions (Adams and Seinfeld, 2002), the variability in net radiative effects (Herzog et al., 2004), and the effects of alterations to emissions on a global scale (Wei et al., 2022). ~~resolve the mixing state of aerosol species particularly when several subpopulations are used (Riemer et al., 2019).~~ More sophisticated aerosol schemes either ~~calculate additional modes~~ compute additional moments of the ASD (Zhang et al., 2020) ~~or~~, explicitly resolve it by using a binned approach (e.g., Adams and Seinfeld, 2002). ~~These models offer (e.g., Adams and Seinfeld, 2002),~~ or represent it on a particle-by-particle basis (Riemer et al., 2019). While these models provide the most physically consistent representation of the ASD, ~~but often are they are often~~ too computationally expensive for operational ~~forecast and long-term~~
60 ~~climate predictions~~ forecasting and long-term climate simulations.

In computationally intensive applications and satellite retrievals it is desirable to maintain the efficiency and simplicity of the bulk schemes, ~~however with a realistic representation of the ASD.~~ However, key processes such as nucleation, coagulation, scavenging, and activation, as well as aerosol radiative properties, are highly sensitive to particle size, requiring the explicit representation ASD (Seinfeld and Pandis, 2016). A common approach to address this is to prescribe a global mean ASD
65 (e.g., Remer et al., 2005; Barahona et al., 2014; Inness et al., 2019; Block et al., 2024). Yet, aerosol size and composition vary substantially across time and space, influenced by both meteorological conditions and natural and anthropogenic sources. As a result, a fixed global ASD can only approximate the actual, locally varying distribution, potentially introducing biases in the simulation of aerosol–radiation and aerosol–cloud interactions.

To address these challenges, there is growing interest in leveraging machine learning (ML) techniques to develop more effi-
70 cient and accurate aerosol models ~~(e.g., Gong et al., 2021; Rasp et al., 2018; Harder et al., 2022; Silva et al., 2021)~~ (e.g., Rasp et al., 2018). ML models, ~~we~~ can in principle capture complex nonlinear relationships between aerosol properties and environmental variables with reduced computational costs. For example, Harder et al. (2022) developed a surrogate of the Modal Aerosol Module ~~(MAM; Liu et al., 2012)~~ (MAM7; Liu et al., 2012) to predict the mass and number tendencies of aerosol species, with the aim to improve computational performance. The emulator replaces computationally intensive parts of ~~MAM~~ MAM7, however does
75 not map the ASD to the mass of the aerosol species, a requirement to many assimilation and remote sensing algorithms (Randles et al., 2017; Buchard et al., 2017). Other work also have sought to use ML to predict the ASD from bulk aerosol mass, focusing on specific species or its effect on cloud formation (Nair et al., 2021; Zhu et al., 2023).

Here we present a novel ML-based approach for predicting the ASD and aerosol mixing state in atmospheric models that run bulk aerosol schemes. This is accomplished by developing a ML-based parameterization that emulates the ASD predicted by the
80 ~~MAM~~ MAM7 model, using as input the total mass of aerosol species from relatively fast single-moment bulk aerosol models like GOCART. By combining the strengths of machine learning and physical principles, the parameterization maps the bulk mass model into the ASD that would be predicted by the modal approach, enhancing the former. Our method offers a promising avenue for advancing aerosol representation in climate predictions, data assimilation and remote sensing applications.

2 Methods and Data

85 We developed a neural network (NN) termed “MAMnet”, to estimate the ASD using as input the total mass of aerosol species ~~and the meteorological state. This was accomplished by training the NN,~~ air density and temperature. This minimal set of inputs ensures that the neural network remains independent of the host model, since including additional meteorological inputs like wind speed and humidity, would introduce sensitivity to model-specific parameterizations. It also makes MAMnet suitable for applications involving satellite aerosol retrievals, where only a limited set of atmospheric variables is typically
90 available. This approach is supported by previous studies showing that the conversion between aerosol mass and number concentrations can be reasonably approximated using spatially varying, but prescribed, ASDs (e.g., Remer et al., 2005; Inness et al., 2019; ~~,~~ suggesting that such relationships can be effectively learned by a neural network. MAMnet was trained on simulated data using the ~~MAM~~ MAM7 model implemented on the NASA’s Global Earth Observing System (GEOS), ~~combining the modal aerosol~~

species predicted by MAM into their total mass across all modes, thus resembling the externally-mixed species predicted by the GOCART scheme. This section details the modeling components as well as the development and evaluation approach of the NN.

2.1 Modeling components

The NASA Goddard Earth Observing System (GEOS), consists of a set of components that numerically represent different aspects of the Earth system (atmosphere, ocean, land, sea-ice, and chemistry), coupled following the Earth System Modeling Framework (https://gmao.gsfc.nasa.gov/GEOS_systems/). In GEOS-AGCM configuration, atmospheric transport of water vapor, condensate and other tracers, and associated land-atmosphere exchanges, is computed explicitly, whereas sea-ice and sea surface temperature (SST) are prescribed as time-dependent boundary conditions (Reynolds et al., 2002; Rienecker et al., 2008). Cloud microphysics in the operational version of GEOS uses a single moment microphysics scheme for short-term weather forecast (Molod et al., 2015), and a two-moment cloud scheme in subseasonal and seasonal prediction (Barahona et al., 2014; Molod et al., 2020). GEOS constitutes the modeling base of MERRA-2 (Modern Era Retrospective analysis for Research and Applications, version 2), the first multidecadal reanalysis to integrate both aerosol and meteorological observations (Gelaro et al., 2017; Randles et al., 2017). In MERRA-2 aerosol fields are described using GOCART. Aerosols are interactive and radiatively active, hence MERRA-2 has a representation of the aerosol direct effect. Aerosol assimilation uses the Goddard Aerosol Assimilation System (GAAS), and the overall assimilation cycle is controlled by the meteorology.

2.1.1 Aerosol transport schemes

GEOS implements two aerosol schemes to interactively calculate the evolution of aerosol and gaseous tracers. Both include parameterized representation of aerosol formation, growth, aging and wet removal, and differ in their treatment of the ASD and mixing state. GOCART (Chin et al., 2000; Colarco et al., 2010a) is used operationally on weather forecast and data assimilation applications. GOCART is a mass-based aerosol model that explicitly calculates the transport and evolution of dust, black carbon, organic material, sea salt, and sulfate. Aerosol species are assumed externally mixed. Dust and sea salt are represented in five mass bins of different sizes whereas a single bin is assumed for other species. The ASD for each bin is prescribed as a lognormal distribution. Both organics and black carbon are split into hydrophilic hydrophilic and hydrophobic components. Dust and sea salt emissions are prognostic whereas sulfate and biomass burning emissions are obtained from the MERRA-2 dataset (~~Randles et al., 2017~~) (Colarco et al., 2010a; Randles et al., 2017).

GEOS also implements the MAM-MAM7 model (Liu et al., 2012), as an alternative aerosol scheme for research applications. MAM-MAM7 is a modal aerosol scheme that predicts the mass and number concentration of Aitken (AIT), accumulation (ACC), primary carbon (PCM), fine dust (FDU) and sea salt (FSS), and coarse dust (CDU) and sea salt (CSS) aerosol modes. The aerosol representation is internally mixed with aerosol species and modal composition as detailed in Table Tables 1 and 2. The total number of simulated tracers in MAM-MAM7 is 31: 24 modal mass components and seven aerosol number concentrations. The size distributions for each mode is assumed to follow a lognormal distribution, with geometric mean diameter computed diagnostically and prescribed geometric standard deviation for each mode (Liu et al., 2012).

Table 1. Aerosol species considered in this work; κ is the hygroscopicity parameter (Kreidenweis et al., 2005).

Abbreviation	Description	κ	Density Kg m^{-3}
SU	Sulfates, includes ammonium	0.64	1600
SS	Sea salt	1.3	2200
OG	Primary and secondary organics	0.25	900
POM	primary Organic Matter	0.25	900
BC	Black carbon	0.01	1600
DU	Dust	0.1	1700

Table 2. Aerosol modes predicted by ~~MAM~~MAM7 and MAMnet; σ_g is the geometric standard deviation.

Abbreviation	Mode	σ_g	Species in mode
ACC	Accumulation	1.8	SU, OG, POM, BC, SS
AIT	Aitken	1.6	SU, OG, SS
CDU	Coarse dust	1.8	SU, DU
CSS	Coarse sea salt	2.0	SU, SS
FDU	Fine dust	1.8	SU, DU
FSS	Fine sea salt	2.0	SU, SS
PCM	Primary carbon matter	1.6	POM, BC

2.2 Development of the deep learning model

We built a neural network, termed “MAMnet”, to estimate the aerosol number concentration and composition ~~mimicking the~~MAM-emulating the output of the MAM7 model (Table 2), using as input the total mass mixing ratios for dust, sulfates, organics, black carbon and sea salt, and the atmospheric state (temperature, T and air density, ρ_{air}), for a total of 31 predicted tracers as shown in Fig. 1. ~~This~~MAMnet is intended to map the simulated aerosol mass across species into a 7-modal ASD, rather than to fully emulate MAM7. This is arguably a simpler task than emulating the full range of aerosol processes represented in MAM7, since the aerosol mass fields used as input already encapsulate the integrated effects of meteorology, clouds, as well as trends in aerosol emissions. The mass-number relationship on the other hand is not expected to depend strongly on such factors, since in many cases it can be approximated to some degree using prescribed formulations for the ASD (McCoy et al., 2017). ~~This~~ section describes the development of the NN.

2.2.1 Data generation

The AGCM configuration of GEOS, running ~~MAM~~MAM7 (referred to as “GEOS+~~MAM~~MAM7”), was used to develop a robust dataset to train the neural network. We ran a 5-year simulation (2001-2006) at 1-degree horizontal resolution and 72 vertical levels (from the surface to 0.01 hPa), with diurnal, instantaneous outputs at UTC 9:00:00 and 21:00:00. ~~From these simulation, 25 output files were randomly selected without replacement for training, and 10 for the testing of the trained model.~~

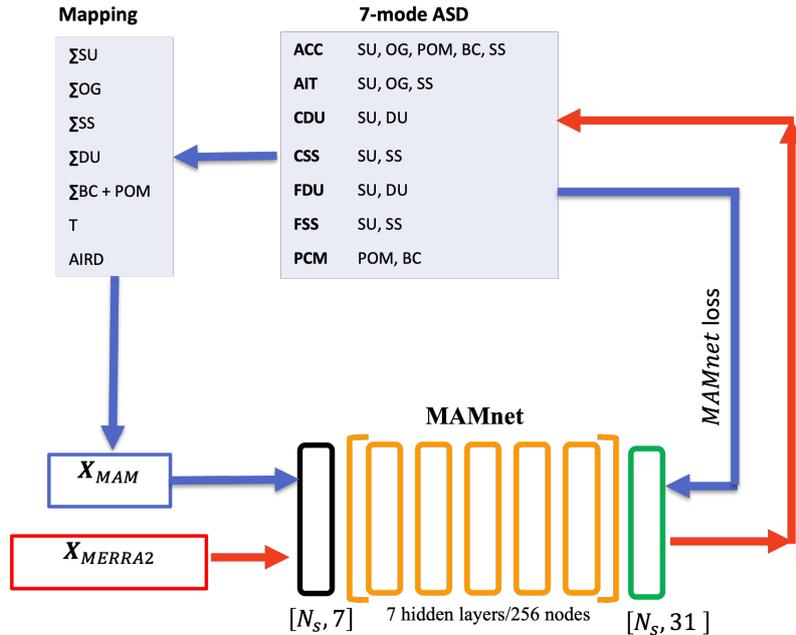


Figure 1. Neural network development workflow. Blue arrows represent the training steps, while red arrows correspond to the inference process. During training, the “mapping” step aggregates aerosol species across modes from the MAM-MAM7 output to construct the input X_{MAM} . The MAM-MAM7 output is then used to calculate the MAMnet loss (calculated as the minimum mean square difference between the model prediction and the MAM7 fields). During inference, input from MERRA-2 or GOCART is used to predict the aerosol size distribution and mixing state, mimicking the MAM output. MAMnet consists of a single input layer (black), seven hidden layers (orange), and one output layer (green). T and ρ_{air} AIRD represent the temperature and air density, respectively. N_s is the number of samples.

Temperature and horizontal winds were “replayed” to MERRA-2. The replay technique is a form of nudging that combines analysis increments with the model results every six hours, at each model grid point, to correct the model state every six hours (Takacs et al., 2018). This ensures that the simulation reproduces the observed meteorological state. Aerosol mass however
 145 evolves freely from observational constraints.

To construct the training dataset, we randomly selected 25 unique output files (without replacement) from the years 2001–2005. An additional 10 unique files were used as validation data to compute the loss during training. Each file represents global instantaneous output from the GEOS+MAM7 model. Although the validation data are not used to update the network parameters, the validation loss informs optimization choices (e.g., early stopping and hyperparameter selection) and is therefore considered
 150 part of the training process. For testing, we used 5 separate output files from the year 2006, which was not included in either the training or validation stages, ensuring a fully independent evaluation set.

Each grid cell in the GEOS+MAM7 output is treated as an independent training example, resulting in a large volume of data: with $N_{time} = 25$ timestamps, $N_{lev} = 72$ vertical levels, $N_{lat} = 181$ latitudes, and $N_{lon} = 360$ longitudes, the training set

contains over 100 million samples. The samples are randomly shuffled in time and space prior to training. This single-cell approach makes the parameterization resolution-independent, facilitating integration into atmospheric models with varying grid resolutions. It also ensures broad coverage of physically plausible combinations of aerosol mass and number. Although this approach omits spatial or vertical correlations, the mass-number relationship depends primarily on the relative abundance of species within a given grid cell. This was tested by using using a full-column input structure, which resulted in no significant gain in accuracy (not shown).

To balance data volume and temporal representativeness, we sampled at 12-hour intervals, which allows the network to capture differences between day and night while maximizing the number of training samples. Higher-frequency sampling could better resolve the diurnal cycle, but this comes at the cost of fewer training time steps due to memory limitations. This is not expected to be critical as the relationship between aerosol mass and number is expected to exhibit weaker diurnal variability than mass itself, which would be already resolved by the host model.

We combined the internally-mixed, modal mass components parameterized by MAM-MAM7 across 5 different species including sulfate, sea salt, dust, primary and secondary organic matter, and black carbon (Table 1 and Fig. 1) to derive the total mass mixing ratios for the input features. ~~The input mass features were log₁₀ mass input variables were first log₁₀-transformed and all input variables (including T and ρ_{air}) were standardized by removing the mean and dividing by the standard deviation.~~ and the resulting values were then standardized by computing Z-scores using the global mean and standard deviation across all levels. Temperature and air density were also standardized using their global mean and standard deviation. Statistics used for normalization were calculated using 100 random instantaneous output files not used during training. Target variables included the mass of each of the MAM species and the number concentration for each mode. Because aerosol mass and number concentration vary over several orders of magnitude, we log₁₀-transformed the targets, and filtered out values less than minimum threshold values, $10^{-20} \mu\text{g Kg}^{-1}$ and 10^{-4}mg^{-1} for mass and number, respectively. The modal aerosol dry diameter (hereafter D_{pg}) was not directly included as a target of MAMnet, and it is not part of the loss function. Instead it was used to check for mass conservation, that is, matching the predicted D_{pg} against the target values indicates that the mass and number concentration remain consistent in the prediction. D_{pg} was derived for each i^{th} mode in the form (Seinfeld and Pandis, 2016),

$$D_{\text{pg},i} = \left(\frac{6}{\pi N_i} \sum_{j=1}^{N_{\text{sp},i}} \frac{M_{j,i}}{\rho_{j,i}} \right)^{1/3} \exp \left(-\frac{3 \ln^2 \sigma_{g,i}}{2} \right) \quad (1)$$

where $\sigma_{g,i}$ and N_i are the geometric standard deviation and number concentration for the i^{th} mode, respectively. $M_{j,i}$ and $\rho_{j,i}$ are the mass and density of the j^{th} species in the i^{th} mode, respectively, and $N_{\text{sp},i}$ is the number of species present in the mode.

~~During training data was fed to the NN using a “single-level” approach where the 4-dimensional geospatial fields are flattened into one-dimensional arrays with a total number of samples given by:-~~

$$N_{\text{samples}} = N_{\text{lev}} \times N_{\text{time}} \times N_{\text{lat}} \times N_{\text{lon}}$$

185 where for each of N_{time} time steps, N_{lev} is the number of model levels (72), N_{lat} is the number of latitudinal grid points (181), and N_{lon} is the number of longitudinal grid points (360). The single-level approach makes the parameterization resolution-independent facilitating its potential incorporation into GCMs with various resolutions. It also provides a large volume of training data (> 100 M samples) capturing a wide array of physically plausible instances in the training data for MAMnet. It however has the disadvantage that potentially significant spatial relations may be missing in the input. Tests using
 190 a whole-column approach suggested that given the large number of samples such caveat did not result in loss of accuracy (not shown).

2.2.2 Model architecture

Various levels of complexity were tested for the MAMnet architecture, including Multilayer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs) (Bengio et al., 2017). These architectures have demonstrated success in capturing multi-scale
 195 behaviors of GCMs for different physical properties (e.g., Brenowitz and Bretherton, 2019; Rasp et al., 2018; Barahona et al., 2024). MLPs extract global patterns from the entirety of the input feature vector simultaneously, resulting in a greater number of model parameters for optimization. This approach compels the NN to make localized decisions, considering what occurs at an individual model level within each grid cell and time step, utilizing global information encompassing all grid cells and time steps. In contrast, CNNs extract features from smaller spatiotemporal blocks, enabling local decisions to be influenced by
 200 nearby information where the receptive field of each sample is a hyperparameter. Testing of both architectures showed that the MLP configuration exhibited superior performance and was easier to optimize. The final architecture is shown in Fig. 1.

~~The MAMnet model was trained using the Keras library with Tensorflow backend (Chollet et al., 2015). Optimization was carried out with the Adam algorithm (Kingma and Ba, 2014) using the minimum mean square error (MSE) as the loss function. Hyperparameter optimization for MAMnet was performed using the Keras Tuner software (O'Malley et al., 2019). Approximately 1500 optimization trials were performed using random configurations of the hyperparameters in Table A1, using a subset of the training data as in Yu et al. (2024). All trials used the same subset of the training/validation data (5 output files for training, 2 for validation). For each parameter set, a new model was built and trained for up to 100 epochs with the same early stopping criteria used during the training of MAMnet. For each trial, a custom metric was recorded at the end of each epoch, the convergence loss ($\mathcal{L}_{\text{conv}}$), defined as the absolute difference between the training and validation losses. Using this custom metric allowed us to select the model that generalizes the best over both the training and validation data sets. The best set of hyperparameters was selected by choosing the configuration that minimized the MSE on the validation set and had the lowest $\mathcal{L}_{\text{conv}}$ as detailed in the Appendix.~~

205 Hyperparameters targeting generalization (dropout rate, activation function, batch size) were tuned such that the optimized model not only minimized error on the validation data, but minimized the

215 Parameters used during hyperparameter tuning for MAMnet. Optimal hyperparameters are shown in bold. **Hyperparameter Values Interrogated** Number of dense layers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 Number of nodes per layer 32, 64, 128, **256**, 512 Batch normalization **True**, **False**, Dropout **True**, **False**, Dropout rate **0.1**, 0.2, 0.3, 0.5 Initial learning rate $1e-3$, $1e-4$, **$1e-5$** , $1e-6$ Activation function ReLU, ELU, **Leaky ReLU** Activation α 0.1, 0.2, **0.3** Optimizer **Adam**, SGD, RMSprop Batch size **64×72** , 128×72 , 2

Table 3. Datasets for the period 2008–2009 used for comparison with surface aerosol size distributions predicted by MAMnet. The original data reference is given, although all data sets used in this work were curated by Asmi et al. (2011).

Station Name	Station Code	Altitude (m.a.s.l.)	Reference
Aspvreten	ASP	30	Tunved et al. (2004)
Birkenes	BIR	190	Amunsen et al. (1992)
Pallas	PAL	560	Lihavainen et al. (2008)
Preila	PLA	5	Ulevicius et al. (2010)
SMEAR II	SMR	181	Hari et al. (2013)
Vavihill	VHL	172	Kristensson et al. (2008)
Bösel	BOE	16	Birmili et al. (2009)
K-Puszt	KPO	125	Kiss et al. (2002)
Melpitz	MPZ	87	Engler et al. (2007)
Kosetice	OBK	534	Červenková and Váňa (2010)
Hohenpeissenberg	SMPS	988	Birmili et al. (2003)
Waldhof	WAL	70	Birmili et al. (2009)
Cabauw	CBW	60	Russchenberg et al. (2005)
Harwell	HWL	60	Charron et al. (2007)
Mace Head	MHD	5	Jennings et al. (1991)
Finokalia	FKL	250	Mihalopoulos et al. (1997)
JRC-Ispra	ISP	209	Gruening et al. (2009)
Zeppelin	ZEP	474	Ström et al. (2003)
Puy de Dôme	PDD	1465	Venzac et al. (2009)
Schauninsland	SCH	1210	Birmili et al. (2009)
Zugspitze	ZSF	2650	Birmili et al. (2009)
Jungfraujoch	JFJ	3580	Jurányi et al. (2011)
BEO Moussala	BEO	2971	Nojarov et al. (2009)
Monte Cimone	CMN	2165	Marinoni et al. (2008)

2.3 Observational data

220 Besides synthetic data the neural network was evaluated on its ability to reproduce observations when driven by the MERRA-
 225 2 reanalysis output. This was important for testing the reliability of MAMnet when applied outside of the purely simulated
 environment. Near-surface aerosol number concentrations ranging from 30 nm to 500 nm, compiled by Asmi et al. (2011),
 were used for model evaluation. The dataset includes two years (2008–2009) of hourly measurements from 24 sites across
 Western Europe, as detailed in Table 3. These measurements were collected from two major monitoring networks: the European
 Supersites for Atmospheric Aerosol Research (EUSAAR) project, part of the Sixth Framework Programme of the European
 Commission (Philippin et al., 2009), and the German Ultrafine Aerosol Network (GUAN) (Birmili et al., 2009). The data are

reported as cumulative number concentrations for four aerosol size ranges, N_{30} , N_{50} , N_{100} , and N_{250} , defined as,

$$N_X = \sum_{D_p=X}^Y N(D_p) \quad (2)$$

where the subscript X represents the aerosol number concentration for size range defined by threshold X , and $Y = 500$ nm for $X = 50, 100$ and 250 nm and $Y = 50$ nm for $X = 30$ nm. Equivalently, these can be calculated from the predicted ASD in the form (Seinfeld and Pandis, 2016),

$$N_X = \sum_{i=1}^{N_{\text{mod}}} \frac{N_i}{2} \left[\text{erf} \left(\frac{\ln Y - \ln D_{\text{pg},i}}{\sqrt{2} \ln \sigma_{g,i}} \right) - \text{erf} \left(\frac{\ln X - \ln D_{\text{pg},i}}{\sqrt{2} \ln \sigma_{g,i}} \right) \right] \quad (3)$$

where $N_{\text{mod}} = 7$, is the number of lognormal modes. For each site, MERRA-2 derived aerosol mass concentration, temperature and air density are interpolated at the location and time of the measurements, then used in MAMnet to predict the ASD. Using Eqs. 1 and 3, ~~the predicted N_X can be~~ is predicted as the average of the two lowermost model levels (i.e., nearest the surface) and compared against the observations.

~~Further evaluation was performed by comparing N_{CCN} derived from MAMnet against global datasets. Watson-Parris et al. (2019) utilized the Global Aerosol Synthesis and Science Project (GASSP) dataset (Reddington et al., 2017) to assess the vertical distributions of aerosol number and N_{CCN} in the global aerosol climate model ECHAM-HAM. The GASSP database contains aerosol measurements from 37 field campaigns and over 1000 flights, largely concentrated over North America and Western Europe. to place MAMnet in the context of observationally-constrained estimates. To calculate N_{CCN} from MAMnet we followed the method of Fountoukis and Nenes (2005) to estimate N_{CCN} from the derived 7-modal size distribution and modal composition and the MERRA-2 fields as inputs. Hygroscopicity parameters, (κ) for each mode were obtained by volume-weighting the values for each aerosol species as listed in Table 1. Because N_{CCN} is strongly influenced by the ASD and composition, it serves as a useful diagnostic for evaluating the estimation of particle size. CCN concentrations are highly sensitive to aerosol size (Lee et al., 2013), as larger and more hygroscopic particles are more likely to activate into cloud droplets. As a result, N_{CCN} tends to be enhanced in populations dominated by such particles. Underestimation of particle size therefore translates into a lower N_{CCN} .~~

Global N_{CCN} datasets are typically derived from bulk aerosol mass using simplified assumptions about the ASD. For example, Choudhury and Tesche (2022) estimated N_{CCN} from spaceborne CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) lidar measurements using pre-computed conversion factors. Block et al. (2024) derived N_{CCN} based on the latest Copernicus Atmosphere Monitoring Service (CAMS) reanalysis provided by the European Centre for Medium-Range Weather Forecast (ECMWF) (Block, 2023). Finally, we compare MAMnet results against the GiOcean atmosphere-ocean-aerosol by introducing assumptions on ASD and composition (Block, 2023). Similarly, The GiOcean atmosphere-ocean-aerosol reanalysis (Song et al., 2025), derived from the NASA GEOS-S2S system (Molod et al., 2020), incorporates a more advanced model framework. Unlike MERRA-2, which only assimilates the atmospheric state, GiOcean is a coupled atmosphere-ocean reanalysis that incorporates an updated model framework. This framework atmosphere-ocean reanalysis that includes two-moment cloud microphysics, enabling the explicit calculation of N_{CCN} . The evaluation of N_{CCN} also serves as a test for

the estimation of particle size, as CCN concentrations are highly sensitive to the aerosol size (Lee et al., 2013). We used the approach of Fountoukis and Nenes (2005) to estimate N_{CCN} from the derived 7-modal size distribution and modal composition. Hygroscopicity parameters, κ , for each mode were obtained by volume-weighting the values for each aerosol species as listed in Table 1. Although GiOcean still relies on assumptions about aerosol size and composition, its CCN fields directly influence cloud processes and, in turn, aerosol evolution. In addition to comparing with reanalysis-based products, we evaluated MAMnet’s CCN predictions against in situ measurements from the Global Aerosol Synthesis and Science Project (GASSP) (Watson-Parris et al., 2019; Reddington et al., 2017). The GASSP dataset compiles aerosol and CCN measurements from 37 field campaigns and over 1,000 aircraft flights, primarily concentrated over North America and Western Europe.

3 Results and Discussion

We evaluated the MAMnet model for both its ability to reproduce the original MAM-GEOS+MAM7 model when driven by the testing data set, and to reproduce observations when driven by aerosol concentrations derived from the MERRA-2 reanalysis. We assessed whether MAMnet reproduces the spatial distribution of aerosol variables in GEOS+MAM-MAM7 using the the mean Pearson’s spatial correlation coefficient (R). We also calculated the mean log-bias, i.e.,

$$MLB = \frac{\sum_{n=1}^{N_{\text{samples}}} \log_{10}(\hat{Y}) - \log_{10}(Y)}{N_{\text{samples}}}, \quad (4)$$

where \hat{Y} and Y correspond to the predicted variables by MAMnet and GEOS+MAM-MAM7, respectively. In general MLB in the range $[-0.5, 0, 5]$ indicates a prediction within an order of magnitude of window around the target value. These metrics are summarized in Fig. 3 for each pressure level and output variable.

3.1 Evaluation against GEOS+MAM7

We first tested whether MAMnet had learned the physical relationships underlying the ASD or simply memorized the training data, and whether the model is able to conserve mass. To investigate this, we run MAMnet using MERRA-2 inputs, then recover the total mass of each species by mapping from the seven modes produced by MAMnet, as depicted in Fig. 1. Since aerosol concentrations in GEOS+MAM7 are not assimilated, they are expected to differ from MERRA-2, which incorporates observational constraints. If MAMnet had merely memorized the GEOS+MAM outputs, these same biases would persist when MERRA-2 inputs were used. Such a discrepancy would also indicate that MAMnet does not conserve mass as the total mass of each species would differ from the inputs.

Figure 2 compares the total aerosol mass column from MERRA-2 (left), MAMnet driven by MERRA-2 inputs (center), and GEOS+MAM (right). It is evident that GEOS+MAM7 tends to underestimate black carbon (BC) and sea salt (SS) over the ocean compared to MERRA2. If MAMnet had merely memorized the GEOS+MAM7 outputs, these same biases would persist when MERRA-2 inputs were used. Instead, MAMnet accurately reproduces the MERRA-2 aerosol concentrations when driven by MERRA-2 inputs, highlighting the internal consistency of the model as MAMnet is able to generalize to new, unseen data.

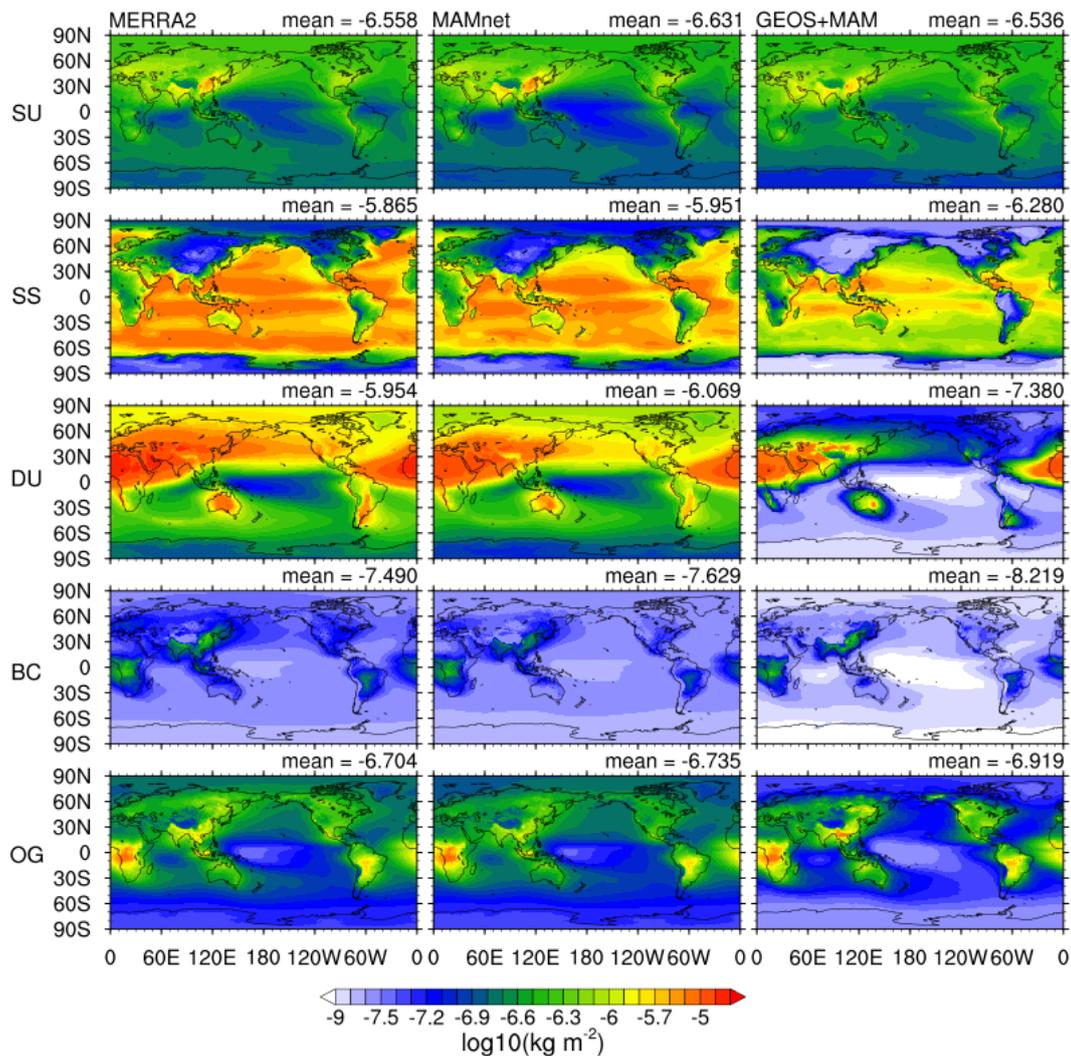


Figure 2. ~~Pearson's spatial correlation~~ Comparison of column-integrated bulk aerosol (P) (from top: sulfates, sea salt, dust, black carbon, organic matter) and mean log-bias represented by (Eq. 4; bottom from left) predicted by MAMnet calculated on MERRA-2, the reserved test set trained MAMnet model applied to MERRA-2 inputs, against and the reserved GEOS+MAM simulation test data. Results are shown for mass (Table 1 and 2) and number (NUM) concentration for each mode, as well as for the derived modal diameter (DPG).

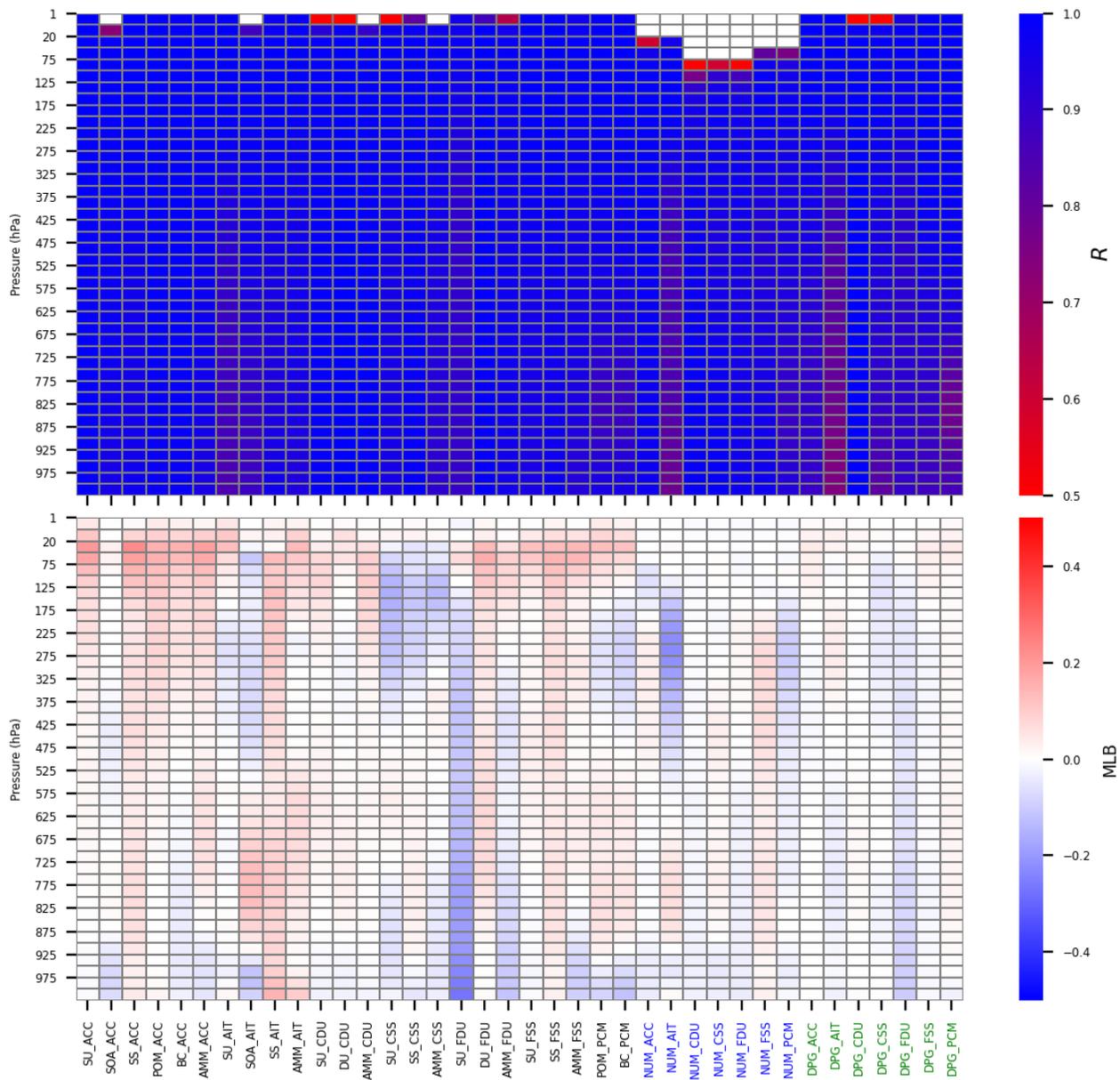


Figure 3. Pearson's spatial correlation (R) (top) and mean log-bias (Eq. 4; bottom) predicted by MAMnet calculated on the reserved test set, against the GEOS+MAM7 simulation. Results are shown for mass (Table 1 and 2) and number (NUM) concentration for each mode, as well as for the derived modal diameter (DPG). Label color on the horizontal axis is added for emphasis only.

290 This test also demonstrates that mass is conserved, as the total mass of each aerosol species is recovered showing very little bias.

~~MAMnet demonstrates strong performance in reproducing~~ [Fig. 3 summarizes the performance of MAMnet compared to GEOS+MAM7 across all output number and mass variables, as well modal size, for all model levels. MAMnet is able to reproduce](#) the modal number concentrations (“NUM” variables in Fig. 3) from the GEOS+MAM simulations, with high spatial correlations $R > 0.9$ and mean log-bias (MLB) within ± 0.1 across most pressure levels. However, performance slightly degrades at pressures below $p < 100$ hPa, particularly for the Aitken (NUM_AIT) and coarse dust modes (NUM_CDU), where correlations drop slightly ($R > 0.7$) and MLB increases to ± 0.3 . The largest discrepancies occur near the surface ($p > 900$ hPa) and in the upper troposphere (150 – 400 hPa). Specifically, NUM_AIT shows underprediction between 150 – 400 hPa while it overpredicts from 700 – 850 hPa, indicating that MAMnet tends to underestimate fine particles near the tropopause and overestimate them in the mid-to-lower troposphere. Similarly, NUM_PCM exhibits negative biases near 150 – 400 hPa, suggesting underprediction of particle number in the primary carbon mode at higher altitudes. The MLB patterns (Fig. 3, bottom panel) reveal localized biases at specific pressure ranges. Positive MLBs (orange shading) occur in NUM_PCM and NUM_ACC, indicating slight overestimation at mid-to-lower pressure levels. In contrast, NUM_CDU displays small negative MLB (blue shading) around 500 – 700 hPa, suggesting underprediction of coarse dust particles in the mid-troposphere. In summary, ~~while MAMnet captures the overall trends in the modal number concentrations with high accuracy, systematic errors emerge for specific modes and pressure ranges. Aitken modenumber concentrations show the largest deviations, particularly near the surface and upper troposphere, while biases in MAMnet captures overall modal number patterns well, but errors remain in the Aitken mode,~~ primary carbon and coarse dust ~~modes are localized to mid-tropospheric levels.~~

Figure 4 shows the zonal mean profiles of modal aerosol number concentration, comparing GEOS+MAM ~~MAM7~~ outputs (left column), MAMnet predictions (center column). Consistent with Fig. 3, the Aitken mode exhibits the largest biases, characterized by underestimation above 400 hPa and overestimation in the lower troposphere, mostly below 700 hPa. Unlike other aerosol modes, the vertical distribution of Aitken mode particles is unique, with higher concentrations found in the upper troposphere and lower stratosphere ($p < 400$ hPa) compared to lower altitudes, where the other modes exhibit the highest concentrations near the surface. The underestimation in the upper troposphere and overestimation in the lower troposphere may be influenced by a “dilution” effect, as Aitken particles contribute relatively little mass compared to other modes. [This may be exacerbated in regions with active particle nucleation processes and combustion emissions which tend to disproportionately enhance number over mass concentration.](#)

The accumulation mode (ACC), coarse sea salt (CSS), and fine sea salt (FSS) modes show generally strong agreement between true and predicted values, with minimal biases across most pressure levels, with MLB values close to zero. However, localized biases are apparent for FSS and PCM near the surface, suggesting slight overestimation in these regions. The coarse dust mode (CDU) and fine dust mode (FDU) exhibit minimal errors overall, with MLB values near zero across most pressure levels. MAMnet accurately predicts the aerosol number concentration for most modes, with systematic biases for the Aitken and primary carbon modes, particularly near the tropopause and in the lower troposphere, suggest that smaller ~~and less massive~~ particles are more difficult to predict accurately due to their unique vertical distribution and sparse representation in the data.

MAMnet accurately reproduces the spatial patterns of the aerosol mass, with accumulation mode mass variables such as SU_ACC, SS_ACC, and SOA_ACC showing ~~near-perfect~~ [high](#) correlations ($R > 0.9$) across the entire pressure range (Fig.

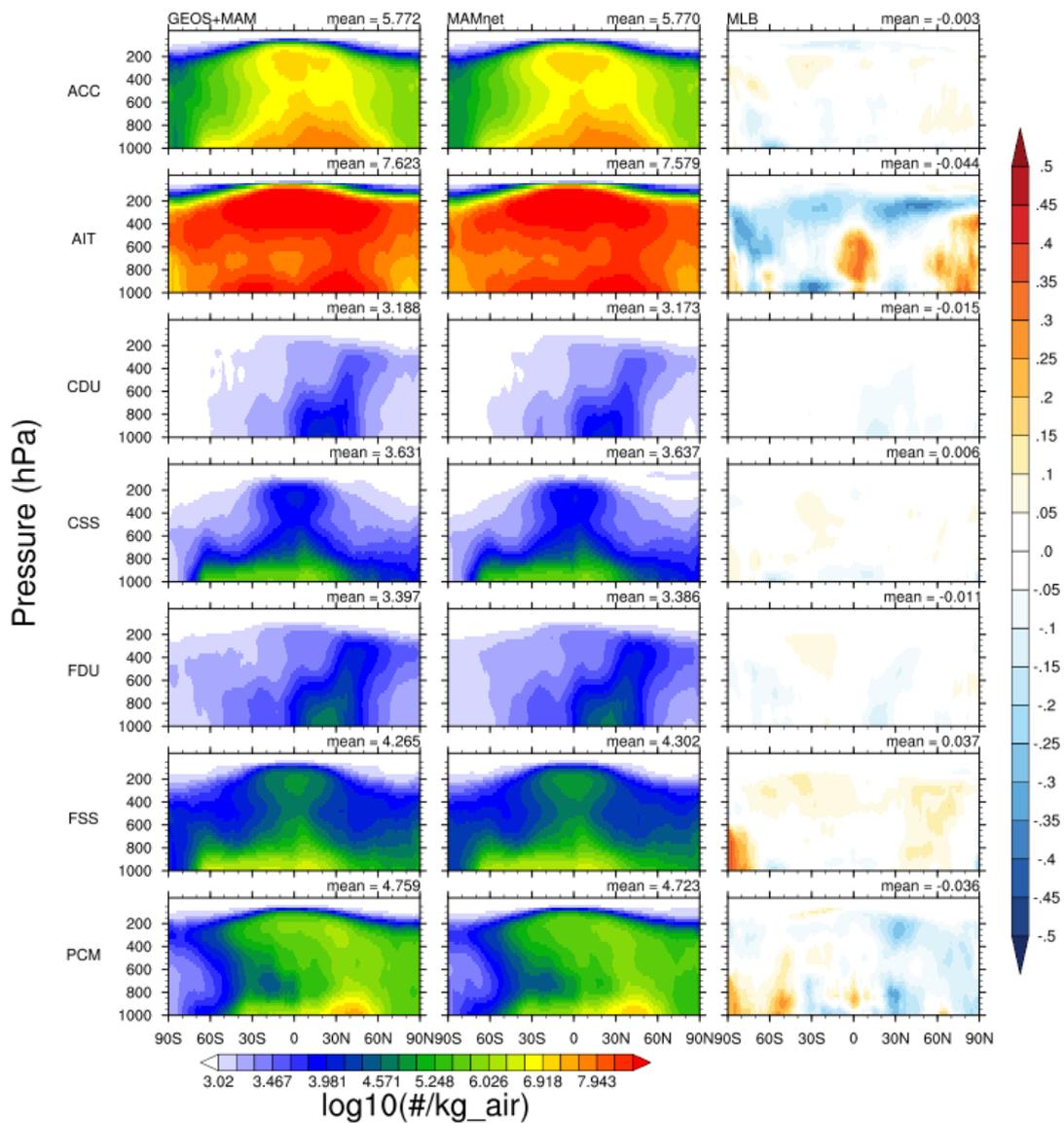


Figure 4. Zonal profiles for modal aerosol number concentration. Left: GEOS+MAM-MAM7 reserved test set. Middle: MAMnet prediction. Right: Mean Log-Bias. From top: Accumulation (ACC), Aitken (AIT), coarse dust (CDU), coarse sea salt (CSS), fine dust (FDU), fine sea salt (FSS), primary carbon matter (PCM).

3) This is also the case for most other variables with only DU_FDU and AMM_FSS showing slight reduction in correlation near 1000 hPa, indicating slightly worse performance in the lower atmosphere. Biases shown in Fig. 3 (bottom) indicate that all but six mass tracers (SOA_ACC, SU_AIT, SOA_AIT, SU_CSS, SS_A_CSS, AMM_CSS) are systematically overpredicted for $p < 200$ hPa where mass values are very small ($\sim 10^{-20}$ kg kg⁻¹). These errors tend to be exacerbated in logarithmic
330 space but remain negligible in absolute terms. Negative biases are also notable for SU_FDU and AMM_FDU, which become increasingly negative towards the surface. This is explained by the low mass of ~~sulfates~~ sulfate in the fine dust mode leading to a “dilution” of sulfate in fine dust mode relative to other aerosol modes. Overall, the model demonstrates robust predictive skill for most aerosol types, with minor discrepancies concentrated near the surface and in sparse aerosol regimes.

Differences in the zonal distribution between aerosol modes can contribute to errors, especially due to the uneven representation of smaller particles, such as those in the Aitken and organic modes, in the training dataset, typically referred to as “class
335 imbalance” (Japkowicz and Stephen, 2002; Buda et al., 2018). For instance, the mass of sulfate particles in the accumulation mode is often at least ten times greater than in the Aitken mode, whereas the opposite is true for the number concentration. As a result, the variability in sulfate mass is primarily driven by the accumulation mode, causing the Aitken mode to be underrepresented in the neural network’s input data. Despite this, the residual differences between predicted and true values for Aitken
340 mode aerosol number concentrations are small compared to other modes, and the mean global error remains well below an order of magnitude, highlighting the neural network’s accuracy.

Figures 5 and 6 show the MAMnet-derived D_{pg} in ~~remarkable~~ agreement with GEOS+MAM. ~~MLBs between the two datasets are typically below 0.01, indicating consistency in the predicted modal~~ MAM7. D_{pg} is buffered against variation in mass and number concentration (Eq. 1). This buffering effect holds only if mass and number vary coherently.
345 In practice, MAMnet predicts number concentration as a single output per mode, while mass is distributed across multiple species per mode. Unlike in the physical model, where mass and number concentrations, and suggesting that MAMnet conserves mass. This is are dynamically linked through the governing equations, MAMnet treats them as independent outputs. Therefore, accurate reproduction of D_{pg} by MAMnet is not guaranteed and must be learned implicitly. MLB for D_{pg} when tested against the test set is typically below 0.01, consistent with the high correlation coefficients for DPG ($R > 0.9$), shown
350 in Fig. 3, ~~highlighting strong agreement with the test dataset across most pressure levels.~~ The fact that MAMnet is able to maintain low bias in D_{pg} without being explicitly constrained to do so suggests that the network has successfully learned a physically consistent relationship between mass and number.

The global distribution of D_{pg} for the different aerosol modes closely matches the spatial patterns of the modal number concentrations. Larger residuals are observed near the surface in the tropics and the Southern Hemisphere, particularly for fine
355 dust (FDU). This discrepancy arises primarily due to the very low number concentrations of fine dust over the oceans, making it challenging for the neural network to accurately predict values close to zero. However, such inaccuracies are unlikely to affect climate simulations, as regions with negligible aerosol concentrations typically do not contribute significantly to atmospheric processes. Residuals for D_{pg} in the coarse dust (CDU) mode are also slightly larger compared to other modes. This is evident in the zonal mean profiles (Fig. 6), where biases are most prominent in the tropics and near the Arctic. Additionally, MAMnet
360 tends to underestimate D_{pg} in the Southern Hemisphere around 30S, particularly in the free troposphere for fine dust. This

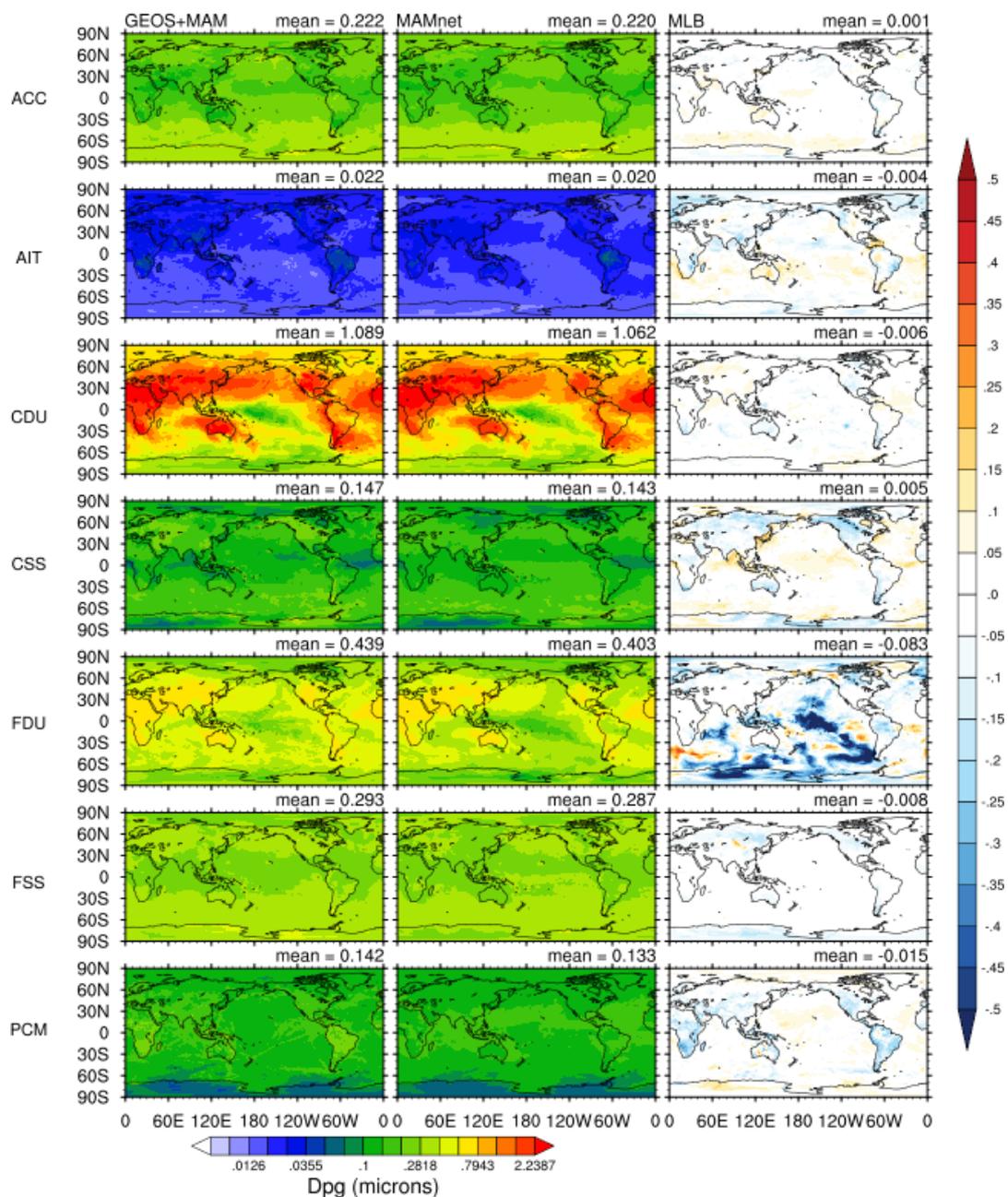


Figure 5. Modal geometric diameter, D_{pg} at 950 hPa. Left: GEOS+MAM-MAM7 reserved test set. Middle: MAMnet prediction. Right: Residual ($\log_{10}(\text{MAMnet}) - \log_{10}(\text{GEOS+MAM})$). From top: Accumulation (ACC), Aitken (AIT), coarse dust (CDU), coarse sea salt (CSS), fine dust (FDU), fine sea salt (FSS), primary carbon matter (PCM).

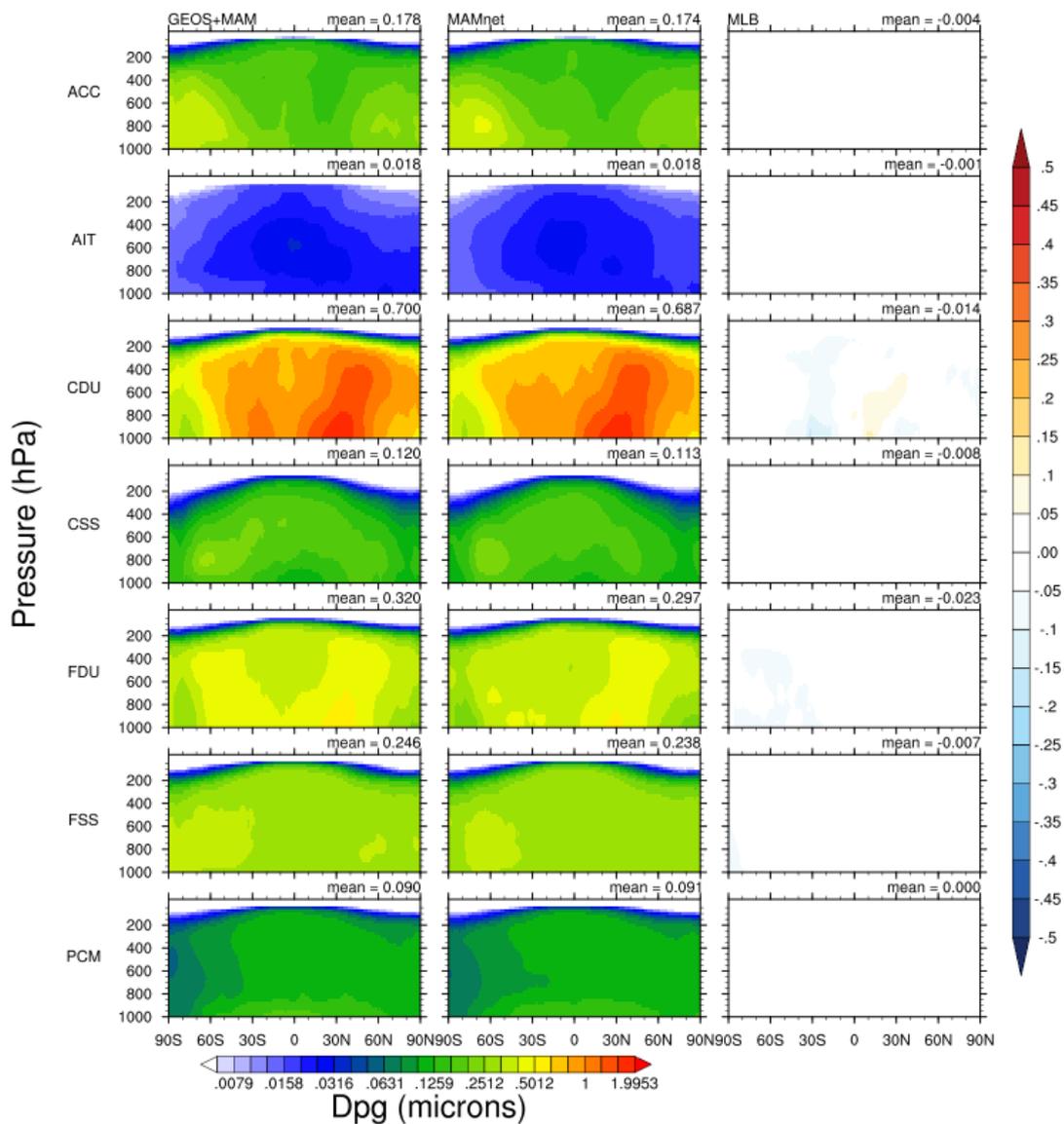


Figure 6. As in Figure 4, but for geometric mean diameter (D_{pg}) per mode.

underestimation likely results from class imbalance, as dust concentrations in this region are very low, making it difficult for the neural network to learn accurate predictions. Overall MAMnet shows remarkable skill at reproducing the ASD from GEOS+MAM, with no evidence of systematic biases. It is also possible that MAMnet has learned associations biased toward aerosol-rich environments, which are more prevalent in the Northern Hemisphere.

365 3.2 Explainable machine learning analysis

Shapley values (Winter, 2002), originally developed in cooperative game theory, are now widely used to interpret predictions from neural networks (Kwon et al., 2023; Jeggle et al., 2023; Jia et al., 2023; Ma and Stinis, 2020; Lundberg and Lee, 2017). A Shapley value quantifies the contribution of a single input feature to a specific model prediction by comparing the prediction for a given sample to the average prediction across all samples. This contribution is averaged over all possible combinations of the remaining input features, referred to as coalitions. Because the number of such combinations grows rapidly with the number of features, we approximate Shapley values using 1,000 randomly selected coalitions per calculation, facilitated by the SHAP python library using the kernel explainer method (Lundberg et al., 2020). In this study, Shapley values are used to assess the influence of each input feature on the predicted aerosol number concentrations for each mode.

Figure 7 shows a summary plot of Shapley values calculated for each input feature relative to predicted targets for aerosol modal number concentration (left) and mass (right). Each row represents a specific feature. The x-axis represents SHAP values, indicating the impact (positive or negative) of each feature on the model's prediction, so that the features with larger SHAP values contribute more significantly to the model output. Red dots represent high feature values, while blue dots indicate low feature values.

Features such as sulfate (SU), sea salt (SS), dust (DU), temperature (T), and air density (AIRD) are consistently ranked as dominant contributors, with their relative importance varying across aerosol modes. Fine-mode outputs, such as ACC and AIT, are strongly influenced by sulfate and temperature, where higher feature values positively impact predictions. In contrast, coarse-mode outputs like CDU and FDU are heavily driven by dust, with significant positive contributions observed for high dust concentrations. Intuitively, this makes sense as SU, SS, and DU are the largest components of accumulation mode aerosols. However the relation is non-linear as high SU values correspond to a strong positive impact on aerosol number, particularly for dust (CDU) and sea salt (CSS) modes, whereas low values lead to neutral or negative contributions. The SHAP values further highlight the critical role of air density and temperature in for CSS and FSS, which may be related to the aerosol activation processes.

For aerosol mass, the SHAP plots show significant influence from DU, SS, BC, SU. The interplay between feature importance and values is evident as, high sea salt concentrations (SS) are positively correlated with increased mass in CSS, while low values lead to neutral or negative contributions. One significant characteristic of the mass SHAP plots is the broader range of SHAP values compared to number concentration, indicating greater variability in the importance of input features for predicting mass. For instance, black carbon (BC) has a consistently positive influence on ACC and AIT mass predictions, but its impact is less pronounced for other modes. Additionally, temperature at low values negatively impacts aerosol mass, while at higher values, it positively influences mass. In some cases however a SHAP value for a feature may have no obviously interpretable significance to the target prediction, or may be so strongly correlated with another feature that the individual contribution is negligible with respect to the feedbacks between a pair of features or more (Aas et al., 2020).

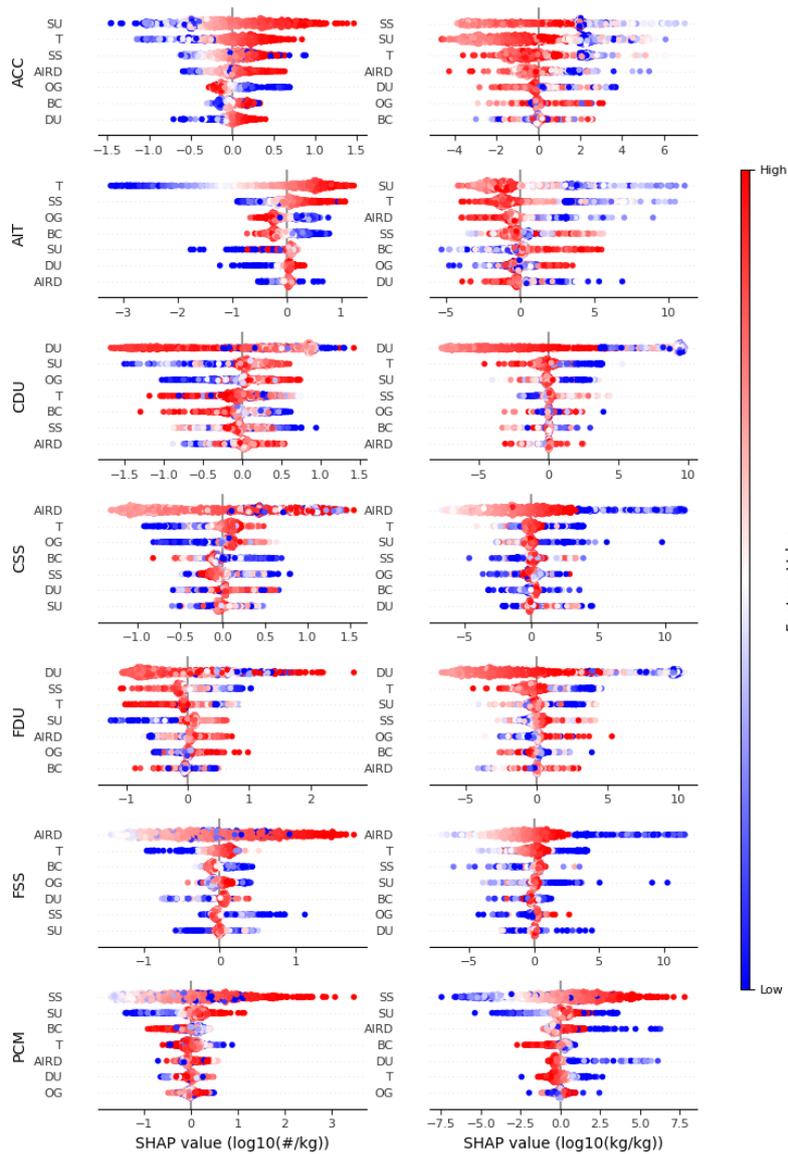


Figure 7. SHAP analysis for aerosol number concentration (left) and total mass (right) across different modes in the troposphere, based on 1,000 randomly selected samples from the test set. Modes are displayed from top to bottom: Accumulation (ACC), Aitken (AIT), coarse dust (CDU), coarse sea salt (CSS), fine dust (FDU), fine sea salt (FSS), and primary carbon matter (PCM). The color gradient (red for high values, blue for low values) indicates the relative value of each feature, with features ordered top-to-bottom by their importance to the prediction (most sensitive at the top). The x-axis represents SHAP values, quantifying how much each feature contributes to deviations from the mean prediction.

3.3 Evaluation against observations

The ability of a neural network to generalize to new data is a key measure of its effectiveness and reliability in real-world applications. While the neural network may perform well reproducing simulated data, it is important to test whether MAMnet is able to reproduce patterns observed in nature. To accomplish this, we take advantage of the ability of MAMnet to work with reanalysis data, that is, using as input the assimilated fields of MERRA-2.

~~We first tested whether MAMnet had learned the physical relationships underlying the ASD or simply memorized the training data. To investigate this, we evaluated whether MAMnet, when driven with MERRA-2 inputs, would reproduce the biases of the GEOS+MAM simulation or instead align with the MERRA-2 fields. Figure 2 compares the total aerosol mass column from MERRA-2 (left), MAMnet driven by MERRA-2 inputs (center), and includes aerosol mass fields that are constrained by satellite observations through data assimilation (Buchard et al., 2017; Sun et al., 2019; Ukhov et al., 2020; Gueymard and Yang, 2020; Sun et al., 2020), and thus provides a more realistic input compared to free-running model simulations. Although GEOS+MAM (right). Since aerosol concentrations in GEOS+MAM are not assimilated, they are expected to differ from MERRA-2, which incorporates observational constraints. This discrepancy is evident in Fig. 2, where GEOS+MAM tends to underestimate black carbon (BC) and sea salt (SS) over the ocean. If MAMnet had merely memorized the GEOS+MAM outputs, these same biases would persist when MERRA-2 inputs were used. Instead, MAMnet accurately reproduces the MERRA-2 aerosol concentrations when driven by MERRA-2 inputs. It is important to note that in this analysis, the aerosol outputs were mapped from the seven modes produced by MAMnet. This result highlights the internal consistency of the model and demonstrates that MAMnet generalizes to new, unseen data. MAM7, which was used to train MAMnet, does not assimilate aerosols and cannot be directly compared to observations at specific sites, it provides physically consistent mass and number concentrations from which the network learns the relationship between these quantities. When applied to MERRA-2, MAMnet combines this learned relationship with more observation-constrained aerosol mass fields, allowing us to evaluate how well it maintains physical consistency in a more realistic setting. This comparison does not validate MAMnet independently of its training data but serves to assess its performance when driven by the best available mass estimates.~~

3.3.1 Comparison against ground observations

Figure 8 compares the cumulative ASD predicted by MAMnet against surface observations from different European sites. These are mostly coastal sites with composition typical of clean and polluted continental origin, that is, mostly of sulfates, dust, organics and sea salt (Asmi et al., 2011). Altitude ranges from a few meters to about 3 km providing a good overview of the lower troposphere. Although representing a limited set, the range of aerosol compositions, sources, and altitudes offers a meaningful assessment of the model's ability to generalize to different atmospheric states. To carry out the comparison, MAMnet was run using collocated aerosol concentrations and meteorological fields obtained from MERRA-2 at each site, and using Eq. 3.

Except for high altitude sites (Fig. 8, bottom row), MAMnet tends to predict slightly lower median values compared to observations, with this discrepancy becoming more pronounced as particle size increases (N100 and N250). This is particularly

430 noticeable at locations such as PAL, PLA, OBK, MHD, FKL, and JFJ, where the model underestimates values consistently. The pattern is reversed at high-altitude sites (PDD, SCH, ZSF, JFJ, BEO, and CMN), where median N100 and N250 are generally overestimated by MAMnet, although the observations themselves display significant variability. Some locations like SMR, WAL, CBW, and SCH exhibit better agreement, with overlapping medians and interquartile ranges. Additionally, the spread of values for MAMnet is typically narrower than for observations, indicating that the model underestimates variability.

435 Observations also show more outliers, whereas MAMnet predictions are more constrained. There are a few exceptions where MAMnet slightly overestimates values, such as VHL (N250) and SMR (N100). Overall, systematic bias exists in MAMnet-predicted particle concentrations, particularly for larger size bins, capturing less variability compared to observations.

MERRA-2 data, which provide MAMnet's input, may not resolve local emissions, terrain, or small-scale meteorology, such as boundary layer height and humidity. This can lead to biases, particularly in the larger particle size categories that depend on aerosol growth processes. Moreover, the training data likely lacks sufficient diversity, particularly for remote or high-altitude sites, which are probably underrepresented in the training set. Additionally, aerosol evolution involve complex, nonlinear interactions that are not explicitly modeled by MAMnet. These factors likely contribute to the model's challenges in capturing the magnitude and variability of aerosol concentrations observed in the real world. Additionally, it is important to note that retrievals of ASD are inherently complex, and experimental errors can be significant, particularly for larger particle sizes (Asmi et al., 2011). Nevertheless, the consistent results across many sites indicate that MAMnet is capable of reasonably capturing the ASD on regional scales when driven by reanalysis data.

440

445

3.3.2 Comparison against global CCN datasets

~~Comparison of column-integrated bulk aerosol (from top: sulfates, sea salt, dust, black carbon, organic matter) represented by (from left) MERRA-2, the trained MAMnet model applied to MERRA-2 inputs, and the reserved GEOS+MAM test data.~~

450 Figure 9 illustrates the global mean distribution of cloud condensation nuclei (CCN) at 0.2% supersaturation at 900 hPa, derived from MAMnet driven by MERRA-2 (shown as MAMnet-MERRA2), GiOcean (Song et al., 2025), the CAMS aerosol reanalysis (Block et al., 2024), and CALIOP satellite retrievals (Choudhury and Tesche, 2022). Data were averaged over the period 2006 – 2021. All datasets reveal similar spatial patterns, with lower CCN concentrations over oceans, particularly in polar regions, and higher concentrations over central and eastern Asia, Europe, and the Americas. However, large differences in absolute values are evident with MAMnet-MERRA2 consistently showing the lowest N_{CCN} , and CALIOP-derived the highest.

455 These discrepancies likely arise from differing assumptions in estimation methods. GiOcean and CAMS estimate CCN based on aerosol mass, prescribing the ASD, and assuming externally-mixed aerosols, which may double-count CCN as organics and sulfates are typically internally mixed (Adachi and Buseck, 2008; Kirpes et al., 2018). MAMnet-MERRA2 avoids this issue but underpredicts N_{CCN} over oceanic regions, likely due to low sea salt concentrations in MERRA-2, stemming from uncertainties in the aerosol assimilation system (Buchard et al., 2017). In contrast, CALIOP may overestimate N_{CCN} from the assumption of CCN as all soluble aerosols above 50 nm, some of which may not activate as CCN at 0.2% supersaturation.

460

Figure 10 compares global mean vertical profiles of N_{CCN} from the datasets in Figure 9 and from in-situ observations (Watson-Parris et al., 2019). Vertical distributions vary significantly. GiOcean, CALIOP, and in-situ profiles exhibit similar

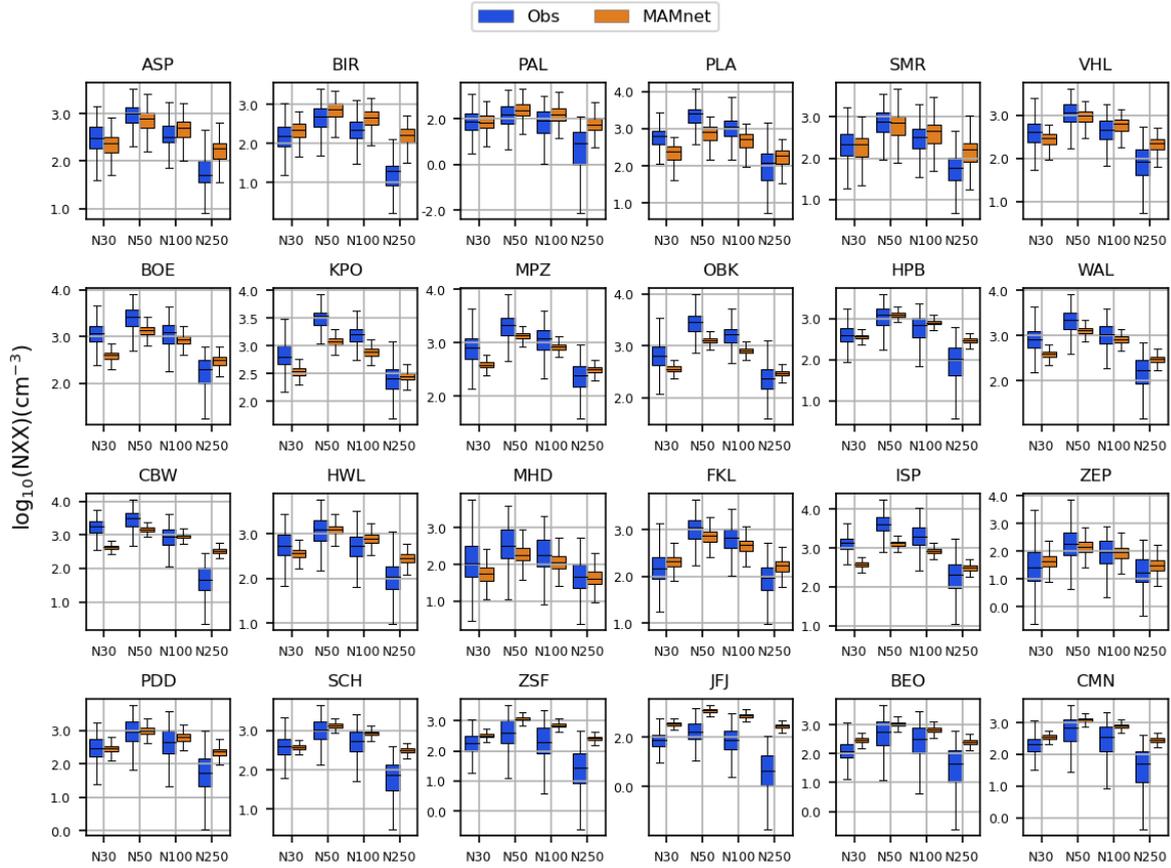


Figure 8. Cumulative size distribution comparison of the trained MAMnet model applied to MERRA-2 inputs against surface measurements (Asmi et al., 2011). The sites in the bottom row (PDD, SCH, ZSF, JFJ, BEO, CMN) are characterized as high altitude sites, with altitudes between 1200 and 3600 m.a.s.l.

shapes with peak concentrations around 950 hPa, while the MAMnet-MERRA2 and CAMS profiles show a monotonic decrease with altitude. The peak in N_{CCN} at 950 hPa may result from more efficient aerosol scavenging near the surface, better represented by two-moment cloud microphysics in GiOcean (Song et al., 2025; Barahona et al., 2014). In contrast, MAMnet-MERRA2 and CAMS rely on single-moment cloud microphysics, which may explain the smoother decrease in N_{CCN} with height. The reliance on single-moment microphysics may also explain the more gradual decrease in N_{CCN} with height in CAMS and MAMnet-MERRA2 than in the other data sets, noticeable over the ocean. In the free troposphere, MAMnet-MERRA2 tends aligns more closely with GiOcean, CALIOP and the in situ data.

3.4 Explainable machine learning analysis

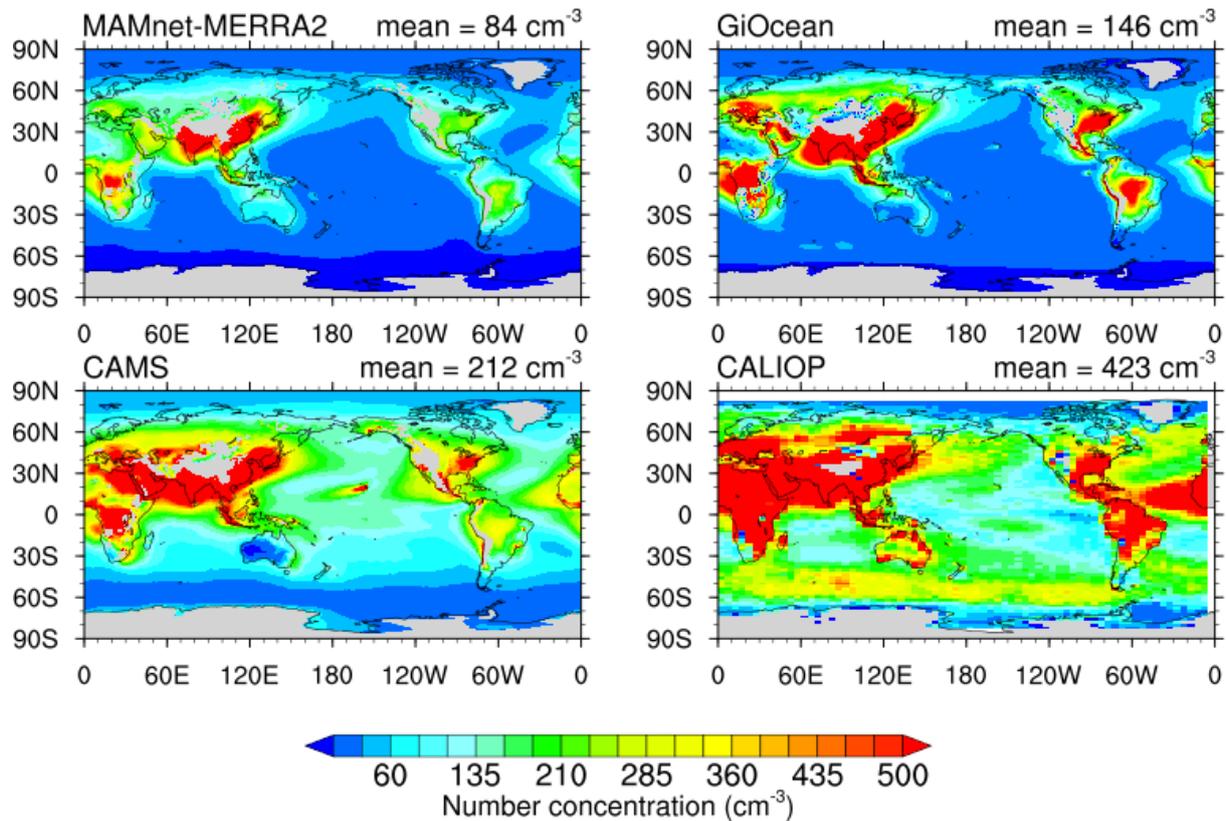


Figure 9. MAMnet-derived CCN at 0.2% supersaturation at 900 hPa using the MERRA2 reanalysis (MAMnet-MERRA2) against global CCN datasets. Also shown are results from the GiOcean reanalysis (Song et al., 2025), CALIOP- (Choudhury and Tesche, 2022), and CAMS- (Block et al., 2024) derived CCN.

Shapley values were used to elucidate the relative influence of input features on the model's decision-making process (Winter, 2002). Figure 7 is a summary plot of Shapley values calculated for each input feature relative to predicted targets for aerosol modal number concentration (left) and mass (right). Each row represents a specific feature. The x-axis represents SHAP values, indicating the impact (positive or negative) of each feature on the model's prediction, so that the features with larger SHAP values contribute more significantly to the model output. Red dots represent high feature values, while blue dots indicate low feature values.

Features such as sulfate (SU), sea salt (SS), dust (DU), temperature (T), and air density (AIRD) are consistently ranked as dominant contributors, with their relative importance varying across aerosol modes. Fine-mode outputs, such as ACC and AIT, are strongly influenced by sulfate and temperature, where higher feature values positively impact predictions. In contrast, coarse-mode outputs like CDU and FDU are heavily driven by dust, with significant positive contributions observed for high dust concentrations. Intuitively, this makes sense because as SU, SS, and DU are the largest components of accumulation

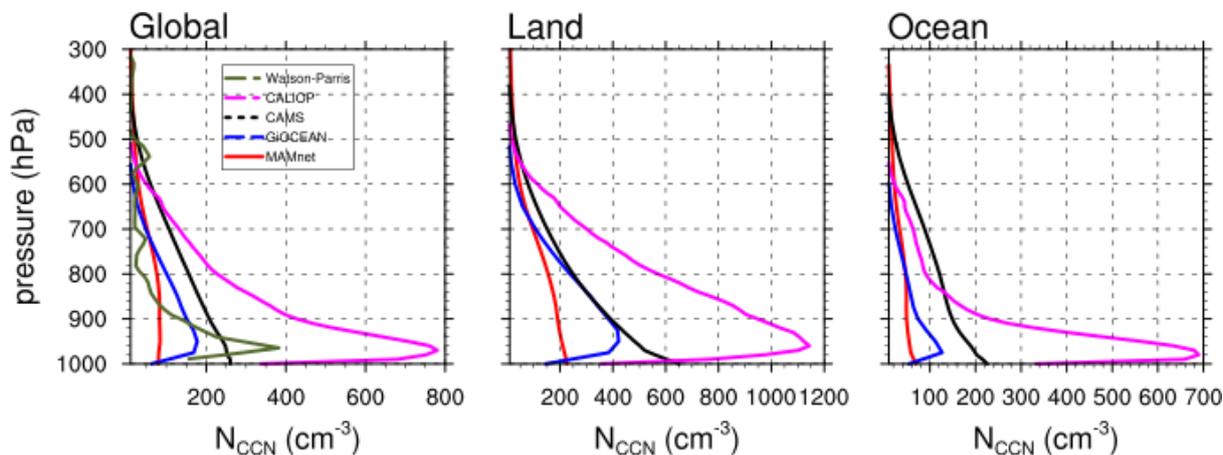


Figure 10. Annual mean profile of CCN concentration derived from MERRA2 using MAMnet (red). Also shown are CALIOP-derived CCN (magenta; Choudhury and Tesche, 2022), the GiOcean reanalysis (blue; Song et al., 2025), CCN derived from field campaign data around the globe (green; Watson-Parris et al., 2019), and CAMS-derived CCN (black; Block et al., 2024).

mode aerosols. However the relation is non-linear as high SU values correspond to a strong positive impact on aerosol number, particularly for dust (CDU) and sea salt (CSS) modes, whereas low values lead to neutral or negative contributions. The SHAP values further highlight the critical role of air density and temperature in for CSS and FSS, which may be related to the aerosol activation processes.

For aerosol mass, the SHAP plots show significant influence from DU, SS, BC, SU. The interplay between feature importance and values is evident as, high sea salt concentrations (SS) are positively correlated with increased mass in CSS, while low values lead to neutral or negative contributions. One significant characteristic of the mass SHAP plots is the broader range of SHAP values compared to number concentration, indicating greater variability in the importance of input features for predicting mass. For instance, black carbon (BC) has a consistently positive influence on ACC and AIT mass predictions, but its impact is less pronounced for other modes. Additionally, temperature at low values negatively impacts aerosol mass, while at higher values, it positively influences mass, possibly by promoting secondary aerosol formation. In some cases however a SHAP value for a feature may have no obviously interpretable significance to the target prediction, or may be so strongly correlated with another feature that the individual contribution is negligible with respect to the feedbacks between a pair of features or more (Aas et al., 2020). Overall, the SHAP analysis identifies key physical drivers that govern aerosol behavior across fine and coarse scales in MAMnet.

SHAP analysis for aerosol number concentration (left) and total mass (right) across different modes in the troposphere, based on 1,000 randomly selected samples from the test set. Modes are displayed from top to bottom: Accumulation (ACC), Aitken (AIT), coarse dust (CDU), coarse sea salt (CSS), fine dust (FDU), fine sea salt (FSS), and primary carbon matter (PCM). The color gradient (red for high values, blue for low values) indicates the relative value of each feature, with features ordered

top-to-bottom by their importance to the prediction (most sensitive at the top). The x-axis represents SHAP values, quantifying how much each feature contributes to deviations from the mean prediction.

4 Conclusions

505 This study develops a neural network, termed MAMnet, to predict the aerosol size distribution and mixing state using as input the bulk mass of different aerosol species, temperature and density. MAMnet is oriented towards allowing a better estimation of the ASD and the aerosol physicochemical properties in cases where computational cost considerations prevent the usage two and higher moment aerosol microphysics schemes, for instance, weather forecast, or where limited information is available to constraint the ASD as in remote sensing and data assimilation.

510 ~~MAMnet was designed to reproduce the output of the MAM model (Liu et al., 2012).~~ The neural network was optimized for performance taking into account the model architecture, training parameters and the rank of the data used as input. ~~We performed~~

MAMnet is designed to reproduce the mapping from aerosol mass to size distribution as generated by the MAM7 scheme within the GEOS system. As such, it inherits the assumptions and meteorological context of the training simulations. However
515 through a comprehensive evaluation of the NN model against simulated data and observations, ~~showing we demonstrated~~ that MAMnet is a robust and accurate model over a wide set of conditions. Importantly, MAMnet reproduces MERRA-2 aerosol concentrations when driven by MERRA-2 inputs, demonstrating that it has learned physical relationships rather than memorizing the training data, conserving the total aerosol mass. Explainable machine learning analysis showed that MAMnet identifies key physical drivers and the non-linear behavior governing the aerosol distribution across fine and coarse scales.

520 Comparison of MAMnet predictions against a reference dataset from GEOS+~~MAM simulations exhibited remarkable~~ MAM7 simulations resulted in good agreement, with log-mean residuals typically below 0.1 and spatial correlation typically exceeding 0.9 for all aerosol modes. The greatest discrepancies were observed near the surface and in regions with low aerosol concentrations i.e., fine dust over oceans and coarse dust in the Southern Hemisphere free troposphere. These discrepancies are primarily attributed to challenges in the prediction of concentrations near zero and class imbalance. Notably, biases
525 in number and mass concentrations do not significantly influence the prediction of geometric mean diameter, ~~which tends to compare exceedingly well against the GEOS+MAM simulations,~~ indicating that MAMnet captures the physical ~~relationship~~ consistency between aerosol mass and number, inherently conserving mass.

We took advantage of the fact that MAMnet can be driven by output from reanalysis data to evaluate its performance against observations. When driven using collocated MERRA-2 fields, MAMnet reasonably reproduced the measured aerosol size
530 distribution at different ground observation sites, representing a variety of aerosol composition, origin and meteorological conditions. The median values of the predicted number concentrations were generally consistent with observations. However the range of values predicted by MAMnet is in general smaller than observed, indicating that the model underestimates variability. It is likely that coarse reanalysis inputs, limited training data diversity, and the complexities of aerosol evolution, not modeled explicitly by MAMnet, contributed to the observed discrepancies.

535 CCN concentrations derived from MAMnet using the MERRA2 dataset were within the range of reported values but showed
discrepancy near the surface and in regions with high variability, which may originate from uncertainty in the MERRA-2
aerosol fields. As MAMnet reproduces well the training dataset it is likely that the biases against observations result from
biases in the input and complex physics not modelled by ~~MAM-MAM7~~. At this point it is difficult to explicitly attribute
sources of error in MAMnet and this is left for future research. Despite such biases, the comparison against observations
540 indicate that MAMnet is able to capture the aerosol size distribution on regional and global scales.

Strategies to address the remaining biases include applying physical constraints via transfer learning, as well as including
observational data during the training process (Barahona et al., 2024). Class imbalance can potentially be addressed by recon-
figuring the NN such that each mode would be predicted by a separate layer, or even individual NNs. The latter option is less
desirable because it would require developing, constraining, and maintaining multiple NNs as opposed to one. Future work
545 would focus on applying MAMnet to elucidate long-term trends in the ASD as well as on its implementation with GCMs (Ott
et al., 2020). The model MAMnet is designed to be resolution-agnostic, but the relationship between aerosol mass and size
distribution may vary with model resolution and it is suggested to be explored in future work. Nevertheless, the model developed
here provides a versatile foundation to improve the physical representation of aerosols in weather forecasting, remote sensing
and data assimilation, potentially enhancing our ~~understanding~~ understanding of their role in the climate system.

550 *Code and data availability.* The MERRA-2 Reanalysis is publicly available from <https://disc.gsfc.nasa.gov/> (GMAO, 2015). The MAMnet
model and training and test data can be downloaded at <https://zenodo.org/records/15190121> (Barahona and Breen, 2025). CAMS data was ob-
tained from Block (2023). Data from the GiOcean reanalysis was downloaded from [https://portal.nccs.nasa.gov/datashare/gmao/geos-s2s-3/
GiOCEAN_e1/](https://portal.nccs.nasa.gov/datashare/gmao/geos-s2s-3/GiOCEAN_e1/) (GMAO, 2025). CALIOP data was obtained from CALIPSO (2023).

Appendix A: Network Training and Optimization

555 The MAMnet model was trained using the Keras library with Tensorflow backend (Chollet et al., 2015). Optimization was
carried out with the Adam algorithm (Kingma and Ba, 2014) using the minimum mean square error (MSE) as the loss function,
with no additional constraints. Hyperparameter optimization for MAMnet was performed using the Keras Tuner software
(O'Malley et al., 2019). Approximately 1500 optimization trials were performed using random configurations of the hyperparameters
in Table A1, using a subset of the training data as in Yu et al. (2024). All trials used the same subset of the training/validation
560 data (5 output files for training, 2 for validation). For each parameter set, a new model was built and trained for up to 100
epochs with the same early stopping criteria used during the training of MAMnet. For each trial, a custom metric was recorded
at the end of each epoch, the convergence loss ($\mathcal{L}_{\text{conv}}$), defined as the absolute difference between the training and validation
losses. Using this custom metric allowed us to select the model that generalizes the best over both the training and validation
data sets. The best set of hyperparameters was selected by choosing the configuration that minimized the MSE on the validation
565 set and had the lowest $\mathcal{L}_{\text{conv}}$.

Table A1. Parameters used during hyperparameter tuning for MAMnet. Optimal hyperparameters are shown in bold.

<u>Hyperparameter</u>	<u>Values Interrogated</u>
<u>Number of dense layers</u>	<u>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20</u>
<u>Number of nodes per layer</u>	<u>32, 64, 128, 256, 512</u>
<u>Batch normalization</u>	<u>True, False,</u>
<u>Dropout</u>	<u>True, False,</u>
<u>Dropout rate</u>	<u>0.1, 0.2, 0.3, 0.5</u>
<u>Initial learning rate</u>	<u>1e-3, 1e-4, 1e-5, 1e-6</u>
<u>Activation function</u>	<u>ReLU, ELU, Leaky ReLU</u>
<u>Activation α</u>	<u>0.1, 0.2, 0.3</u>
<u>Optimizer</u>	<u>Adam, SGD, RMSprop</u>
<u>Batch size</u>	<u>64 \times 72, 128 \times 72, 256 \times 72, 512 \times 72,</u>

Author contributions. D.B. conceived and directed the work. K.H.B. co-developed of the neural network model. A.D. implemented the MAM model within GEOS. K.B. Provided CCN data for comparison.

Competing interests. The authors declare no conflict of interest.

570 *Acknowledgements.* Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. Keras and Tensorflow libraries were obtained from <https://keras.io/>. Maps were created using the NCAR Command Language (Version 6.6.2) Software. (2019). Boulder, Colorado: UCAR/NCAR/CISL/TDD. <http://dx.doi.org/10.5065/D6WD3XH5>. The SHAP python package was used to conduct the explainable machine learning analysis as described in <https://shap.readthedocs.io/en/latest/>. This work was supported by the NASA Modeling, Analysis and Prediction program, Grant NNH20ZDA001N-MAP.

575 **References**

- Aas, K., Jullum, M., and Løland, A.: Explaining individual predictions when features are dependent: More accurate approximations to Shapley values, 2020.
- Adachi, K. and Buseck, P.: Internally mixed soot, sulfates, and organic matter in aerosol particles from Mexico City, *Atmospheric Chemistry and Physics*, 8, 6469–6481, <https://doi.org/10.5194/acp-8-6469-2008>, 2008.
- 580 Adams, P. J. and Seinfeld, J. H.: Predicting global aerosol size distributions in general circulation models, *Journal of Geophysical Research: Atmospheres*, 107, AAC–4, 2002.
- Amunson, C., Hanssen, J., Semb, A., and Steinnes, E.: Long-range atmospheric transport of trace elements to southern Norway, *Atmospheric Environment. Part A. General Topics*, 26, 1309–1324, 1992.
- Aquila, V., Hendricks, J., Lauer, A., Riemer, N., Vogel, H., Baumgardner, D., Minikin, A., Petzold, A., Schwarz, J., Spackman, J., et al.:
585 MADE-in: A new aerosol microphysics submodel for global simulation of insoluble particles and their mixing state, *Geoscientific Model Development*, 4, 325–355, 2011.
- Arfin, T., Pillai, A. M., Mathew, N., Tirpude, A., Bang, R., and Mondal, P.: An overview of atmospheric aerosol and their effects on human health, *Environmental Science and Pollution Research*, 30, 125 347–125 369, 2023.
- Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A.-M., Sellegri, K., Birmili, W., Weingartner, E., Baltensperger, U., Zdimal, V., Zikova, N.,
590 et al.: Number size distributions and seasonality of submicron particles in Europe 2008–2009, *Atmospheric Chemistry and Physics*, 11, 5505–5538, 2011.
- Barahona, D. and Breen, K.: MAMnet, <https://doi.org/10.5281/zenodo.15190120>, 2025.
- Barahona, D., Molod, A., Bacmeister, J., Nenes, A., Gettelman, A., Morrison, H., Phillips, V., and Eichmann, A.: Development of two-moment cloud microphysics for liquid and ice within the NASA Goddard Earth Observing System Model (GEOS-5), *Geosc. Model Dev.*,
595 7, 1733–1766, <https://doi.org/10.5194/gmd-7-1733-2014>, 2014.
- Barahona, D., Breen, K. H., Kalesse-Los, H., and Röttenbacher, J.: Deep Learning Parameterization of Vertical Wind Velocity Variability via Constrained Adversarial Training, *Artificial Intelligence for the Earth Systems*, 3, e230 025, <https://doi.org/10.1175/AIES-D-23-0025.1>, 2024.
- Bender, F. A.-M.: Aerosol forcing: Still uncertain, still relevant, *AGU advances*, 1, e2019AV000 128, 2020.
- 600 Bender, F. A.-M., Frey, L., McCoy, D. T., Grosvenor, D. P., and Mohrmann, J. K.: Assessment of aerosol–cloud–radiation correlations in satellite observations, climate models and reanalysis, *Climate Dynamics*, 52, 4371–4392, 2019.
- Bengio, Y., Goodfellow, I., and Courville, A.: *Deep learning*, vol. 1, MIT press Cambridge, MA, USA, 2017.
- Birmili, W., Berresheim, H., Plass-Dülmer, C., Elste, T., Gilge, S., Wiedensohler, A., and Uhrner, U.: The Hohenpeissenberg aerosol formation experiment (HAFEX): a long-term study including size-resolved aerosol, H₂ SO₄, OH, and monoterpenes measurements, *Atmo-
605 spheric Chemistry and Physics*, 3, 361–376, 2003.
- Birmili, W., Weinhold, K., Nordmann, S., Wiedensohler, A., Spindler, G., Müller, K., Herrmann, H., Gnauk, T., Pitz, M., Cyrys, J., et al.: Atmospheric aerosol measurements in the German ultrafine aerosol network (GUAN), *Gefahrst. Reinhalt. L.*, 69, 137–145, 2009.
- Block, K.: Cloud condensation nuclei (CCN) numbers derived from CAMS reanalysis EAC4 (Version 1), https://doi.org/10.26050/WDCC/QUAERERE_CCNCAMS_v1, 2023.
- 610 Block, K., Haghghatnasab, M., Partridge, D. G., Stier, P., and Quaas, J.: Cloud condensation nuclei concentrations derived from the CAMS reanalysis, *Earth System Science Data*, 16, 443–470, <https://doi.org/10.5194/essd-16-443-2024>, 2024.

- Brenowitz, N. D. and Bretherton, C. S.: Spatially extended tests of a neural network parametrization trained by coarse-graining, *Journal of Advances in Modeling Earth Systems*, 11, 2728–2744, 2019.
- Buchard, V., Randles, C., Da Silva, A., Darmenov, A., Colarco, P., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A., Ziemba, L., et al.:
615 The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies, *Journal of Climate*, 30, 6851–6872, 2017.
- Buda, M., Maki, A., and Mazurowski, M. A.: A systematic study of the class imbalance problem in convolutional neural networks, *Neural networks*, 106, 249–259, 2018.
- CALIPSO: Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observation Lidar Level 2 Aerosol Profile V4-20,
https://doi.org/10.5067/CALIPSO/CALIPSO/LID_L2_05KMAPRO-STANDARD-V4-20, 2023.
- 620 Červenkova, J. and Váňa, M.: Trend Assessment of deposition, throughfall and runoff water chemistry at the ICP-IM station Kosetice, Czech Republic, IAHS-AISH publication, 336, 103–108, 2010.
- Charron, A., Birmili, W., and Harrison, R. M.: Factors influencing new particle formation at the rural site, Harwell, United Kingdom, *Journal of Geophysical Research: Atmospheres*, 112, 2007.
- Chin, M., Rood, R. B., Lin, S.-J., Müller, J.-F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART:
625 Model description and global properties, *Journal of Geophysical Research: Atmospheres*, 105, 24 671–24 687, 2000.
- Chollet, F. et al.: Keras, <https://github.com/fchollet/keras>, 2015.
- Choudhury, G. and Tesche, M.: Estimating cloud condensation nuclei concentrations from CALIPSO lidar measurements, *Atmospheric Measurement Techniques*, 15, 639–654, 2022.
- Christensen, M. W., Jones, W. K., and Stier, P.: Aerosols enhance cloud lifetime and brightness along the stratus-to-cumulus transition,
630 *Proceedings of the National Academy of Sciences*, 117, 17 591–17 598, 2020.
- Chu, D., Kaufman, Y., Ichoku, C., Remer, L., Tanré, D., and Holben, B.: Validation of MODIS aerosol optical depth retrieval over land, *Geophysical research letters*, 29, MOD2–1, 2002.
- Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, *J. Geophys. Res.*, 115, D14 207–, <http://dx.doi.org/10.1029/2009JD012820>,
635 2010a.
- Colarco, P., da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, *Journal of Geophysical Research: Atmospheres*, 115, 2010b.
- Engler, C., Rose, D., Wehner, B., Wiedensohler, A., Brüggemann, E., Gnauk, T., Spindler, G., Tuch, T., and Birmili, W.: Size distributions of non-volatile particle residuals (D_p < 800 nm) at a rural site in Germany and relation to air mass origin, *Atmospheric Chemistry and
640 Physics*, 7, 5785–5802, 2007.
- Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., et al.: Changes in atmospheric constituents and in radiative forcing. Chapter 2, in: *Climate change 2007. The physical science basis*, 2007.
- Fountoukis, C. and Nenes, A.: Continued development of a cloud droplet formation parameterization for global climate models, *Journal of Geophysical Research: Atmospheres*, 110, 2005.
- 645 Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *Journal of Climate*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.

- 650 Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and distributions of dust aerosols simulated with the GOCART model, *Journal of Geophysical Research: Atmospheres*, 106, 20 255–20 273, 2001.
- GMAO: `inst33dasmCP: MERRA – 23DIAU State, MeteorologyInstantaneous3 – hourly(p – coord, 0.625x0.5L42)`, <https://doi.org/10.5067/WWQSXQ8IVFW8>, 2015.
- GMAO: GiOcean Coupled Reanalysis, https://portal.nccs.nasa.gov/datashare/gmao/geos-s2s-3/GiOCEAN_e1/, 2025.
- Gong, X., Wex, H., Müller, T., Henning, S., Voigtländer, J., Wiedensohler, A., and Stratmann, F.: An unsupervised machine-learning-based classification of aerosol microphysical properties over 10 years at Cabo Verde, *Atmospheric Chemistry and Physics Discussions*, pp. 1–27, 2021.
- 655 Gruening, C., Adam, M., Cavalli, F., Cavalli, P., Dell’Acqua, A., Martins Dos Santos, S., Pagliari, V., Roux, D., and Putaud, J.: JRC Ispra EMEP–GAW Regional Station for Atmos. Res, Tech. Rep. JRC55382, European Commission, <http://publications.jrc.ec.europa.eu/repository/handle/111111111/538>, 2009.
- Gueymard, C. A. and Yang, D.: Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of 660 AERONET observations, *Atmospheric Environment*, 225, 117 216, <https://doi.org/10.1016/j.atmosenv.2019.117216>, 2020.
- Harder, P., Watson-Parris, D., Stier, P., Strassel, D., Gauger, N. R., and Keuper, J.: Physics-Informed Learning of Aerosol Microphysics, arXiv preprint arXiv:2207.11786, 2022.
- Hari, P., Nikinmaa, E., Pohja, T., Siivola, E., Bäck, J., Vesala, T., and Kulmala, M.: Station for measuring ecosystem-atmosphere relations: SMEAR, *Physical and physiological forest ecology*, pp. 471–487, 2013.
- 665 Herzog, M., Weisenstein, D. K., and Penner, J. E.: A dynamic aerosol module for global chemical transport models: Model description, *Journal of Geophysical Research: Atmospheres*, 109, 2004.
- Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Fleming, J., et al.: The CAMS reanalysis of atmospheric composition, *Atmospheric Chemistry and Physics*, 19, 3515–3556, 2019.
- Japkowicz, N. and Stephen, S.: The class imbalance problem: A systematic study, *Intelligent data analysis*, 6, 429–449, 2002.
- 670 Jeggli, K., Neubauer, D., Camps-Valls, G., and Lohmann, U.: Understanding cirrus clouds using explainable machine learning, *Environmental Data Science*, 2, e19, 2023.
- Jennings, S., O’Dowd, C., O’Connor, T., and McGovern, F.: Physical characteristics of the ambient aerosol at Mace Head, *Atmospheric Environment. Part A. General Topics*, 25, 557–562, 1991.
- Jia, Y., Andersen, H., and Cermak, J.: Analysis of cloud fraction adjustment to aerosols and its dependence on meteorological controls using 675 explainable machine learning, *EGUsphere*, 2023, 1–25, 2023.
- Jones, A., Roberts, D., and Slingo, A.: A climate model study of indirect radiative forcing by anthropogenic sulphate aerosols, *Nature*, 370, 450–453, 1994.
- Jurányi, Z., Gysel, M., Weingartner, E., Bukowiecki, N., Kammermann, L., and Baltensperger, U.: A 17 month climatology of the cloud condensation nuclei number concentration at the high alpine site Jungfraujoeh, *Journal of Geophysical Research: Atmospheres*, 116, 680 2011.
- Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
- Kirpes, R. M., Bondy, A. L., Bonanno, D., Moffet, R. C., Wang, B., Laskin, A., Ault, A. P., and Pratt, K. A.: Secondary sulfate is internally mixed with sea spray aerosol and organic aerosol in the winter Arctic, *Atmospheric Chemistry and Physics*, 18, 3937–3949, <https://doi.org/10.5194/acp-18-3937-2018>, 2018.

- 685 Kiss, G., Varga, B., Galambos, I., and Ganszky, I.: Characterization of water-soluble organic matter isolated from atmospheric fine aerosol, *Journal of Geophysical Research: Atmospheres*, 107, ICC–1, 2002.
- Kreidenweis, S., Koehler, K., DeMott, P., Prenni, A., Carrico, C., and Ervens, B.: Water activity and activation diameters from hygroscopicity data-Part I: Theory and application to inorganic salts, *Atmospheric Chemistry and Physics*, 5, 1357–1370, 2005.
- Kristensson, A., Dal Maso, M., Swietlicki, E., Hussein, T., Zhou, J., Kerminen, V.-M., and Kulmala, M.: Characterization of new particle
690 formation events at a background site in Southern Sweden: relation to air mass history, *Tellus B: Chemical and Physical Meteorology*, 60, 330–344, 2008.
- Kwon, Y., An, S. A., Song, H.-J., and Sung, K.: Particulate Matter Prediction and Shapley Value Interpretation in Korea through a Deep Learning Model, *SOLA*, 19, 225–231, 2023.
- Langner, J. and Rodhe, H.: A global three-dimensional model of the tropospheric sulfur cycle, *Journal of Atmospheric Chemistry*, 13, 225–
695 263, 1991.
- Lee, L., Pringle, K., Reddington, C., Mann, G., Stier, P., Spracklen, D., Pierce, J., and Karshaw, K.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, *Atmospheric Chemistry and Physics*, 13, 8879–8914, <https://doi.org/10.5194/acp-13-8879-2013>, 2013.
- Lihavainen, H., Kerminen, V.-M., Komppula, M., Hyvärinen, A.-P., Laakia, J., Saarikoski, S., Makkonen, U., Kivekäs, N., Hillamo, R.,
700 Kulmala, M., et al.: Measurements of the relation between aerosol properties and microphysics and chemistry of low level liquid water clouds in Northern Finland, *Atmospheric Chemistry and Physics*, 8, 6925–6938, 2008.
- Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., et al.: Toward a minimal representation of aerosols in climate models: Description and evaluation in the Community Atmosphere Model CAM5, *Geoscientific Model Development*, 5, 709–739, 2012.
- 705 Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, *Advances in neural information processing systems*, 30, 2017.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, *Nature machine intelligence*, 2, 56–67, <https://doi.org/10.1038/s42256-019-0138-9>, 2020.
- 710 Ma, P. L. and Stinis, P.: Developing a simulator-based satellite dataset for using machine learning techniques to derive aerosol-cloud-precipitation interactions in models and observations in a consistent framework, Tech. rep., Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), 2020.
- Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., and Johnson, C. E.: Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model,
715 *Geoscientific Model Development*, 3, 519–551, <https://doi.org/10.5194/gmd-3-519-2010>, 2010.
- Marinoni, A., Cristofanelli, P., Calzolari, F., Roccatò, F., Bonafè, U., and Bonasoni, P.: Continuous measurements of aerosol physical parameters at the Mt. Cimone GAW Station (2165 m asl, Italy), *Science of the total environment*, 391, 241–251, 2008.
- McCoy, D., Bender, F.-M., Mohrmann, J., Hartmann, D., Wood, R., and Grosvenor, D.: The global aerosol-cloud first indirect effect estimated using MODIS, MERRA, and AeroCom, *Journal of Geophysical Research: Atmospheres*, 122, 1779–1796, 2017.
- 720 Mihalopoulos, N., Stephanou, E., Kanakidou, M., Pilitsidis, S., and Bousquet, P.: Tropospheric aerosol ionic composition in the Eastern Mediterranean region, *Tellus B*, 49, 314–326, 1997.

- Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, *Geosc. Model Dev.*, 8, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>, 2015.
- 725 Molod, A., Hackert, E., Vikhliav, Y., Zhao, B., Barahona, D., Vernieres, G., Borovikov, A., Kovach, R. M., Marshak, J., Schubert, S., et al.: GEOS-S2S version 2: The GMAO high-resolution coupled model and assimilation system for seasonal prediction, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031767, <https://doi.org/10.1029/2019JD031767>, 2020.
- Nair, A. A., Yu, F., Campuzano-Jost, P., DeMott, P. J., Levin, E. J. T., Jimenez, J. L., Peischl, J., Pollack, I. B., Fredrickson, C. D., Beyersdorf, A. J., Nault, B. A., Park, M., Yum, S. S., Palm, B. B., Xu, L., Bourgeois, I., Anderson, B. E., Nenes, A., Ziemba, L. D., Moore, R. H., Lee, T., Park, T., Thompson, C. R., Flocke, F., Huey, L. G., Kim, M. J., and Peng, Q.: Machine Learning Uncovers Aerosol Size Information
730 From Chemistry and Meteorology to Quantify Potential Cloud-Forming Particles, *Geophysical Research Letters*, 48, e2021GL094133, <https://doi.org/https://doi.org/10.1029/2021GL094133>, e2021GL094133 2021GL094133, 2021.
- Nojarov, P., Ivanov, P., Kalapov, I., Penev, I., and Drenska, M.: Connection between ozone concentration and atmosphere circulation at peak Moussala, *Theoretical and applied climatology*, 98, 201–208, 2009.
- O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al.: Keras Tuner, <https://github.com/keras-team/keras-tuner>, 2019.
- 735 Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., and Baldi, P.: A Fortran-Keras deep learning bridge for scientific computing, *Scientific Programming*, 2020, 2020.
- Philippin, S., Laj, P., Putaud, J.-P., Wiedensohler, A., LEEUW, G. D., FJAERAA, A. M., PLATT, U., BALTENSPERGER, U., and FIEBIG, M.: EUSAAR-An unprecedented network of aerosol observation in Europe, *Journal of Aerosol Research (Earozoru Kenkyu)*, 24, 78–83, <https://doi.org/10.11203/jar.24.78>, 2009.
- 740 Pierce, J. R., Croft, B., Kodros, J. K., D'Andrea, S. D., and Martin, R. V.: The importance of interstitial particle scavenging by cloud droplets in shaping the remote aerosol size distribution and global aerosol-climate effects, *Atmospheric Chemistry and Physics*, 15, 6147–6158, <https://doi.org/10.5194/acp-15-6147-2015>, 2015.
- Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinzuka, Y., and Flynn, C. J.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation
745 Evaluation, *Journal of Climate*, 30, 6823–6850, <https://doi.org/10.1175/JCLI-D-16-0609.1>, 2017.
- Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, *Proceedings of the National Academy of Sciences*, 115, 9684–9689, 2018.
- Reddington, C., Carslaw, K., Stier, P., Schutgens, N., Coe, H., Liu, D., Allan, J., Pringle, K., Lee, L., Yoshioka, M., et al.: The Global Aerosol Synthesis and Science Project (GASSP): measurements and modeling to reduce uncertainty, *Bulletin of the American Meteorological
750 Society*, 98, 1857–1877, 2017.
- Remer, L. A., Kaufman, Y., Tanré, D., Mattoo, S., Chu, D., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R., Kleidman, R., et al.: The MODIS aerosol algorithm, products, and validation, *Journal of the atmospheric sciences*, 62, 947–973, <https://doi.org/10.3390/rs13010002>, 2005.
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An improved in situ and satellite SST analysis for climate, *J. Climate*, 15, 1609–1625, 2002.
- 755 Riemer, N., Ault, A., West, M., Craig, R., and Curtis, J.: Aerosol mixing state: Measurements, modeling, and impacts, *Reviews of Geophysics*, 57, 187–249, <https://doi.org/10.1029/2018RG000615>, 2019.
- Rienecker, M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., Stajner, I., and Nielsen, J.: The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0., vol. 27 of *Technical Report Series on Global Modeling and Data Assimilation*, NASA Goddard Space Flight Center, Greenbelt, MD, USA, 2008.

- 760 Russchenberg, H., Bosveld, F., Swart, D., ten BRINK, H., de LEEUW, G., Uijlenhoet, R., Arbesser-Rastburg, B., van der MAREL, H.,
LIGTHART, L., Boers, R., et al.: Ground-based atmospheric remote sensing in the Netherlands: European outlook, *IEICE Transactions
on Communications*, 88, 2252–2258, 2005.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric chemistry and physics: from air pollution to climate change*, John Wiley & Sons, 2016.
- Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R., et al.:
- 765 Improving our fundamental understanding of the role of aerosol- cloud interactions in the climate system, *Proceedings of the National
Academy of Sciences*, 113, 5781–5790, 2016.
- Silva, S. J., Ma, P.-L., Hardin, J. C., and Rothenberg, D.: Physically regularized machine learning emulators of aerosol activation, *Geoscientific
Model Development*, 14, 3067–3077, 2021.
- Song, C., McCoy, D., Molod, A., and Barahona, D.: Signatures of aerosol-cloud interactions in GiOcean: A coupled global reanalysis with
- 770 two-moment cloud microphysics, *EGUsphere*, 2025, 1–30, <https://doi.org/10.5194/egusphere-2024-4108>, 2025.
- Stier, P., Feichter, J., Kinne, S., Kloster, S., Vignati, E., Wilson, J., Ganzeveld, L., Tegen, I., Werner, M., Balkanski, Y., et al.: The aerosol-
climate model ECHAM5-HAM, *Atmospheric Chemistry and Physics*, 5, 1125–1156, 2005.
- Stier, P., van den Heever, S. C., Christensen, M. W., Gryspeerdt, E., Dagan, G., Saleeby, S. M., Bollasina, M., Donner, L., Emanuel, K.,
Ekman, A. M., et al.: Multifaceted aerosol effects on precipitation, *Nature Geoscience*, 17, 719–732, [https://doi.org/10.1038/s41561-024-](https://doi.org/10.1038/s41561-024-01482-6)
- 775 01482-6, 2024.
- Ström, J., Umegård, J., Tørseth, K., Tunved, P., Hansson, H.-C., Holmén, K., Wismann, V., Herber, A., and König-Langlo, G.: One year of
particle size distribution and aerosol chemical composition measurements at the Zeppelin Station, Svalbard, March 2000–March 2001,
Physics and Chemistry of the Earth, Parts A/B/C, 28, 1181–1190, 2003.
- Su, X., Huang, Y., Wang, L., Cao, M., and Feng, L.: Validation and diurnal variation evaluation of MERRA-2 multiple aerosol properties on
- 780 a global scale, *Atmospheric Environment*, 311, 120 019, <https://doi.org/10.1016/j.atmosenv.2023.120019>, 2023.
- Sun, E., Che, H., Xu, X., Wang, Z., Lu, C., Gui, K., Zhao, H., Zheng, Y., Wang, Y., Wang, H., et al.: Variation in MERRA-
2 aerosol optical depth over the Yangtze River Delta from 1980 to 2016, *Theoretical and Applied Climatology*, 136, 363–375,
<https://doi.org/10.1007/s00704-018-2490-9>, 2019.
- Takacs, L. L., Suárez, M. J., and Todling, R.: The stability of incremental analysis update, *Monthly weather review*, 146, 3259–3275, 2018.
- 785 Tunved, P., Ström, J., and Hansson, H.-C.: An investigation of processes controlling the evolution of the boundary layer aerosol size distri-
bution properties at the Swedish background station Aspöreten, *Atmospheric Chemistry and Physics*, 4, 2581–2592, 2004.
- Ukhov, A., Mostamandi, S., Da Silva, A., Flemming, J., Alshehri, Y., Shevchenko, I., and Stenchikov, G.: Assessment of natural and an-
thropogenic aerosol air pollution in the Middle East using MERRA-2, CAMS data assimilation products, and high-resolution WRF-
Chem model simulations, *Atmospheric Chemistry and Physics Discussions*, 2020, 1–42, [https://doi.org/doi.org/10.5194/acp-20-9281-](https://doi.org/doi.org/10.5194/acp-20-9281-2020)
- 790 2020, 2020.
- Ulevicius, V., Byčėnienė, S., Remeikis, V., Garbaras, A., Kecorius, S., Andriejauskienė, J., Jasinevičienė, D., and Mocnik, G.: Character-
ization of pollution events in the East Baltic region affected by regional biomass fire emissions, *Atmospheric Research*, 98, 190–200,
2010.
- Uno, I., Eguchi, K., Yumimoto, K., Takemura, T., Shimizu, A., Uematsu, M., Liu, Z., Wang, Z., Hara, Y., and Sugimoto, N.: Asian dust
- 795 transported one full circuit around the globe, *Nature Geoscience*, 2, 557–560, 2009.
- Venzac, H., Sellegri, K., Villani, P., Picard, D., and Laj, P.: Seasonal variation of aerosol size distributions in the free troposphere and residual
layer at the puy de Dôme station, France, *Atmospheric Chemistry and Physics*, 9, 1465–1478, 2009.

- 800 Virtanen, A., Joutsensaari, J., Kokkola, H., Partridge, D. G., Blichner, S., Seland, Ø., Holopainen, E., Tovazzi, E., Lipponen, A., Mikkonen, S., et al.: High sensitivity of cloud formation to aerosol changes, *Nature Geoscience*, pp. 1–7, <https://doi.org/10.1038/s41561-025-01662-y>, 2025.
- Watson-Parris, D., Schutgens, N., Reddington, C., Pringle, K. J., Liu, D., Allan, J. D., Coe, H., Carslaw, K. S., and Stier, P.: In situ constraints on the vertical distribution of global aerosol, *Atmospheric Chemistry and Physics*, 19, 11 765–11 790, 2019.
- Wei, L., Lu, Z., Wang, Y., Liu, X., Wang, W., Wu, C., Zhao, X., Rahimi, S., Xia, W., and Jiang, Y.: Black carbon-climate interactions regulate dust burdens over India revealed during COVID-19, *Nature communications*, 13, 1839, 2022.
- 805 Whitby, E. R. and McMurry, P. H.: Modal aerosol dynamics modeling, *Aerosol Science and Technology*, 27, 673–688, 1997.
- Wilson, J., Cuvelier, C., and Raes, F.: A modeling study of global mixed aerosol fields, *Journal of Geophysical Research: Atmospheres*, 106, 34 081–34 108, 2001.
- Winter, E.: The shapley value, *Handbook of game theory with economic applications*, 3, 2025–2054, 2002.
- Yu, S., Ma, P.-L., Singh, B., Silva, S., and Pritchard, M.: Two-step hyperparameter optimization method: Accelerating hyperparameter search
810 by using a fraction of a training dataset, *Artificial Intelligence for the Earth Systems*, 3, e230 013, 2024.
- Zhang, H., Sharma, G., Dhawan, S., Dhanraj, D., Li, Z., and Biswas, P.: Comparison of discrete, discrete-sectional, modal and moment models for aerosol dynamics simulations, *Aerosol Science and Technology*, 54, 739–760, <https://doi.org/10.1080/02786826.2020.1723787>, 2020.
- Zhou, C., Shen, X., Liu, Z., Zhang, Y., and Xin, J.: Simulating aerosol size distribution and mass concentration with simultaneous nucleation, condensation/coagulation, and deposition with the GRAPES–CUACE, *Journal of Meteorological Research*, 32, 265–278,
815 <https://doi.org/10.1007/s13351-018-7116-8>, 2018.
- Zhu, H., Martin, R. V., Croft, B., Zhai, S., Li, C., Bindle, L., Pierce, J. R., Chang, R. Y.-W., Anderson, B. E., Ziemba, L. D., Hair, J. W., Ferrare, R. A., Hostetler, C. A., Singh, I., Chatterjee, D., Jimenez, J. L., Campuzano-Jost, P., Nault, B. A., Dibb, J. E., Schwarz, J. S., and Weinheimer, A.: Parameterization of size of organic and secondary inorganic aerosol for efficient representation of global aerosol optical properties, *Atmospheric Chemistry and Physics*, 23, 5023–5042, <https://doi.org/10.5194/acp-23-5023-2023>, 2023.