Authors' Response to Reviews of

Deep learning representation of the aerosol size distribution

Donifan Barahona, Katherine Breen, Karoline Block, Anton Darmenov
Geoscientific Model Development, 2025

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

RC: ***Use of deep learning model to approximate aerosol size distribution from bulk mass inputs is interesting and operationally valuable. Integration with MERRA-2 opens opportunities for reanalysis and assimilation improvements.***

AR: We appreciate the positive assessment. Please find below detailed responses to each comment.

1. Line-by-line comments

RC: ***L24-27: You may also flag that some modal schemes like GLOMAP in UK Met Office Unified Model assume a lognormal shape for each mode with prescribed geometric standard deviation and each mode is internally mixed. (in L87-88 you do mention something similar for another model)***

AR: Thanks for bringing this to our attention. We now make reference to the GLOMAP model as well.

RC: ***L35: Consider specifying orders of magnitude or cite a study quantifying what is really "better" for representing ASD in models.***

AR: Thanks for the comment. We have added the following paragraph detailing the role of the ASD in simulating aerosol processes:

"The ASD and mixing state is 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]."

RC: ***L56-57: Clarify what is meant by "meteorological state"—mention that it includes only temperature and air density up front, since this is unexpectedly minimal and a key methodological decision.***

AR: Thanks for the comment. We realize that our input set selection requires additional justification. The paragraph has been extended as:

"MAMnet uses a minimal set of inputs, total aerosol mass, air density, and temperature, to predict the ASD. This design choice 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 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, Block et al., 2024], suggesting that such relationships can be effectively learned by a neural network."

RC: ***L77-79: Clarify whether the model includes any simplified representation of aerosol growth, aging, or wet removal in GOCART (even if parametrized) because "transport and evolution" maybe construed for many physical phenomenon.***

AR: This statement has been clarified as: "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. "

RC: ***L87-88: Suggest clarifying whether the geometric mean diameter is prognosed or computed diagnostically.***

AR: The statement has been rewritten as: "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"

RC: ***L96-97: What years were simulated? Why only two time points per day? This sparsity might miss diurnal features. How were the 25 output files selected—what does "one file" correspond to (single timestamp across globe?)? The use of only 25 files for training seems low given the mention of >100M samples later. Please clarify.***

AR: Thanks for bringing this up. Yes, out of the thousands of files of the run we selected 25 timestamps, at random for training, with no duplicates (without replacement). During training the loss is calculated on data not used to update the parameters of the network, termed validation loss, taken as 10 additional files. Each file represents global instantaneous output from the GEOS+MAM7 model. Since the validation loss still guides optimization choices it is considered part of the training step. The testing data is completely independent, and we have used 5 additional files taken from the year 2006, which was not used at all during training.

It is important to clarify that MAMnet is not designed to emulate the underlying aerosol processes, but rather to learn the statistical relationship between the ASD and the total mass of each species. In this scheme each grid cell represents a training example. Thus just a few files already contain a very large number of samples. For our final training we used $N_s = 72 * 180 * 360 * 25 \approx 1.12 \times 10^8$ samples. Given this large number, it is assumed as representative of all possible mass-number combinations produced by MAM, which is what the model needs to train.

Sampling at 12-hour intervals allowed us to use more data for training while still capturing differences between day and night. We agree that higher-frequency sampling could better resolve the diurnal cycle, but this comes at the cost of fewer training time steps due to memory limitations. However, this is not expected to be critical, as the relationship between mass and number likely exhibits weaker diurnal variability than the aerosol mass itself, which would be already represented by the bulk model.

This has been clarified in the work in sections 2.2.1 as follows:

"Each grid cell in the GEOS+MAM7 output is treated as an independent training example, resulting in a large volume of data: with $N_{\mathrm{time}} = 25$ timestamps, $N_{\mathrm{lev}} = 72$ vertical levels, $N_{\mathrm{lat}} = 181$ latitudes, and $N_{\mathrm{lon}} = 360$ longitudes, the training set contains over 100 million samples. 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. To balance data volume and temporal representativeness, we sample at 12-hour intervals, which allows the network to capture differences between day and night while maximizing the number of training samples under memory constraints. Although this approach may omit spatial or vertical correlations, tests using a full-column inputs showed no significant gain in accuracy (not shown). Additionally, 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."

RC: ***L98-100: Add sentence on whether aerosols evolve freely in these simulations or are constrained by observations. Can we understand to what model levels were these "horizontal winds" nudged and to what extend they affect aerosols number concentration?***

AR: Thank you for the comment. In the GEOS+MAM7 simulations used for training, aerosol mass and number evolve freely and is not directly constrained by observations. However, horizontal winds are nudged toward the MERRA-2 reanalysis fields at each model grid point every six hours, influencing the transport of aerosol species but not the aerosol concentrations.

While the nudging of winds may affect aerosol number concentrations indirectly, quantifying this effect would require a dedicated set of sensitivity experiments that are beyond the scope of this study. Here, our focus is on the design and evaluation of the neural network model, assuming the aerosol mass fields as given.

We have clarified this in the manuscript.

RC: ***L104-105: Better to specify: Were log10-transformed values standardized after transformation or before? Are temperature and air density standardized globally or per level?***

AR: The mass input variables were first $\log_{10}$-transformed, 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. This procedure has now been clarified in the manuscript.

RC: ***L110-115: clarify was this Dpg compared only during evaluation, or was it ever used in the loss function? Please state your loss function as some physics informed neural net models have tried modifying it as well.***

AR: The modal aerosol dry diameter was not directly included as a target of MAMnet, and it is not part of the loss function. We used minium square error as loss function with no additional constraints. This is now clarified in the Appendix (former section 2.2.2).

RC: ***L116-120: Clarify whether the flattened fields are shuffled across time and space, or whether there's structure preserved (e.g., batches by time or region). Were any vertical or horizontal correlations exploited or lost?***

AR: The training samples are randomly shuffled in both time and space prior to training. We acknowledge that this approach omits explicit spatial and vertical correlations. However, we do not expect these correlations to significantly impact model performance, as MAMnet is designed to learn the relationship between aerosol mass and number concentration for each mode. This relationship depends primarily on the relative abundance of species within a given grid cell, which the network captures without requiring spatial context. We have clarified this point in the manuscript.

RC: ***L125-134: Were other architectures considered (e.g., transformers, residual connections)? If not, briefly justify.***

AR: Thank you for the suggestion. In designing MAMnet, our priority was to ensure simplicity and computational efficiency, as the neural network is intended for potential use in online applications within large-scale atmospheric models and retrieval systems. More complex architectures such as visual transformers or convolutional networks with residual connections are designed to exploit local spatial correlations within structured data. Since MAMnet treats each grid cell as an independent sample such architectures are not expected to result on improved accuracy. For this reason, we focused on standard fully connected networks, which provide an effective balance between simplicity and computational cost.

RC: ***L139-140: The earlier statement (line 97-98) says 25 files used for training, but here it says "5 for training, 2 for validation." I think I am missing something here?***

AR: Thanks for bringing this up. This line refers specifically to the training data used for hyperparameter optimization. This is conducted by performing about 1500 optimization trials using random configurations of the hyperparameters, and a smaller data set over a fixed number of epochs. In this case only 5 output files were used with 2 for validation.

This explanation is now added to the section, which has been moved to the Appendix for clarity.

RC: ***L163-175: Briefly discuss how errors in Dpg propagate to aerosol number concentration errors for ccn?***

AR: We have added the following explanation to the section: "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 underestimation in N_{CCN}."

RC: ***L244: The underestimation of Dpg in SH is attributed to low data availability? Could it also be due to extrapolation error—MAMnet may have learned associations biased toward NH-dominant training. Can we not test this applying class reweighting in the loss function?***

AR: This is a valid point. It is possible that MAMnet has learned associations biased toward aerosol-rich environments, which are more prevalent in the Northern Hemisphere. The underestimation of D_{pg} in the Southern Hemisphere is most pronounced for fine dust, which is relatively scarce in that region and thus underrepresented in the training data. While applying class weighting in the loss function could help emphasize low-concentration regimes, it may also reduce performance in regions with higher aerosol concentrations, where errors may have a larger climatic impact. We have now acknowledged this limitation in the revised manuscript.

RC: ***L340-344: Under what conditions does this input feature set (bulk mass, T,) suffice? Where do predictions degrade (e.g., strong vertical motions, boundary layer transitions)? Why were other physically relevant predictors (RH, precipitation, cloud fraction, wind) excluded? Does it limits its use in complex meteorological regimes. Without input features tied to wet/dry removal, nucleation, or chemical aging, can the model really be used in weather forecasting or satellite retrievals across diverse regions even when we find high correlations?***

AR: Thank you for the comment. As noted above, MAMnet is designed to map bulk aerosol mass across species into a 7-modal ASD, rather than to emulate the full range of aerosol processes in MAM7. This is a simpler task, as the input mass fields already reflect the integrated effects of meteorology, cloud processes, and emissions. Using only bulk mass, temperature, and air density helps maintain model independence and makes MAMnet suitable for use in both forecast models and satellite retrievals, where other process-based inputs may not be available.

We trained and evaluated MAMnet globally under a wide range of meteorological conditions and aerosol regimes. The large number of training samples ensures broad coverage of realistic atmospheric states. While the model does not include explicit predictors for wet/dry removal or nucleation, our results based on independent testing across years suggest that MAMet learned the nonlinear relationship between mass and number.

RC: ***L345-350: MAMnet is trained only on MAM model outputs, so how does the model avoid learning MAM's own biases? Can we say that evaluation against MERRA-2 is not necessarily independent since the training data is nudged to MERRA-2 meteorology?***

AR: This is a good point. 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 we demonstrate that it is able to learn such a relationship independently of the particular set of simulations used for training. That is, MAMnet learns a physically consistent relationship between the mass of different species and the ASD, applicable across different atmospheric conditions.

This has been clarified in the paragraph.

RC: ***L351-357: Can MAMnet conserve total aerosol mass by design, or does this emerge of calculation? This is never proven numerically—just implied via Dpg.***

AR: This is an important point, which we aimed to illustrate in Figure 6. There, we show that when MERRA-2 fields are used as input and all species predicted by MAMnet are combined (as in Figure 1), the resulting bias is very small. While we originally presented this as evidence that MAMnet does not inherit the biases of GEOS+MAM7, it also demonstrates that the model conserves mass, as it accurately maintains the total mass of each species. Additional support for this comes from the the modal geometric mean diameter, $D_{pg,i}$, remains very close to MAM7. Since $D_{pg,i}$ is not predicted by MAMnet, but instead calculated from its output, it indicates that both mass and number concentrations evolve in a physically consistent manner.

We have moved Figure 6 and the corresponding discussion before Figures 2 and 3 at the start of the new Section 3.1 to clarify this point.

RC: ***L358-370: There's no attribution of error—how much is due to MAMnet, and how much due to MERRA-2 inputs?***

AR: This is correct. The errors against observations reflect the combined effects of MERRA-2 inputs and MAMnet's predictions. Isolating their individual contributions is challenging. However, since MAMnet performs well on independent test data, it is likely the error arises from the input fields or from limitations in the aerosol representation in MAM7. We now acknowledge this as a limitation in the manuscript.

2. General comments

RC: ***Unclear what one "file" represents—single timestep? Single day? Entire global field? It is also unclear whether any temporal or spatial overlap exists between train/test sets.***

AR: Each file corresponds at an instantaneous global field. This is now clarified in the paper. There is no overlap between test and train sets as they are sampled from different years. This is now clarified in the work.

RC: ***No analysis on extrapolation over different time periods (e.g., pre-2000). The network is trained on a***

5-year window using meteorology from MERRA-2 (likely post-2000). How would the model perform in periods with different emissions (e.g., 1980s)? Alternatively, discuss potential limitations in extrapolating to past or future climate states.

AR: We appreciate the reviewer's comment. MAMnet was tested using data from a different year not used during training, demonstrating it is able to generalize beyond ist training data. To clarify, MAMnet is not intended to fully emulate MAM7. Rather, its purpose is to map the simulated aerosol mass across species into a 7-modal aerosol ASD. 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 it many cases it can be approximated to some degree using prescribed formulations for the ASD. Therefore it is likely that it can be learned only from the relative abundances of the aerosol species and their vertical variation (which is encapsulated in the density and temperature). We believe such to be the case as MAMnet performs well using data from a different year of simulation, not used during training. Hence, we don't anticipate a major effect of the time period. This explanation has been added to Section 2.2.

RC: ***How is the SHAP analysis computed over such a high-dimensional sample space (using any explainer method)? Was it computed on the flattened single-level dataset? How do you deal with feature correlation?***

AR: The following clarifying paragraph has been added to the paper:

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 usng the kernel explainer [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.

RC: ***Is MAMnet architecture resolution-agnostic? Though you use single-level training to make the model resolution-independent, how would MAMnet perform in coarser (*** 2.5***) or finer (***<1***) gridded input?***

AR: MAMnet is designed to be resolution-agnostic, as it operates on single grid cell inputs without relying on spatial context. However, we agree that the relationship between aerosol mass and the resulting size distribution may vary with model resolution, as it is typically the case for physical parameterizations. Changes in resolution affect not only aerosol mass fields but also emissions and meteorological inputs, making this a complex issue to assess. A thorough evaluation of resolution dependence is beyond the scope of this study, but we now acknowledge this limitation in the manuscript and suggest to be addressed in future work.

References

V Aquila, J Hendricks, A Lauer, N Riemer, H Vogel, D Baumgardner, A Minikin, Andreas Petzold, JP Schwarz, JR Spackman, et al. Made-in: A new aerosol microphysics submodel for global simulation of insoluble particles and their mixing state. *Geoscientific Model Development*, 4(2):325–355, 2011.

F A-M Bender, Lena Frey, Daniel T McCoy, Daniel P Grosvenor, and Johannes K Mohrmann. Assessment of aerosol–cloud–radiation correlations in satellite observations, climate models and reanalysis. *Climate Dynamics*, 52:4371–4392, 2019.

Annele Virtanen, Jorma Joutsensaari, Harri Kokkola, Daniel G Partridge, Sara Blichner, Øyvind Seland, Eemeli Holopainen, Emanuele Tovazzi, Antti Lipponen, Santtu Mikkonen, et al. High sensitivity of cloud formation to aerosol changes. *Nature Geoscience*, pages 1–7, 2025. .

Chunhong Zhou, Xiaojing Shen, Zirui Liu, Yangmei Zhang, and Jinyuan Xin. Simulating aerosol size distribution and mass concentration with simultaneous nucleation, condensation/coagulation, and deposition with the grapes–cuace. *Journal of Meteorological Research*, 32(2):265–278, 2018. .

J. R. Pierce, B. Croft, J. K. Kodros, S. D. D'Andrea, and R. V. Martin. 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(11):6147–6158, 2015. . URL `https://acp.copernicus.org/articles/15/6147/2015/`.

N Riemer, AP Ault, M West, RL Craig, and JH Curtis. Aerosol mixing state: Measurements, modeling, and impacts. *Reviews of Geophysics*, 57(2):187–249, 2019. .

Lorraine A Remer, YJ Kaufman, D Tanré, S Mattoo, DA Chu, J Vanderlei Martins, R-R Li, C Ichoku, RC Levy, RG Kleidman, et al. The modis aerosol algorithm, products, and validation. *Journal of the atmospheric sciences*, 62(4):947–973, 2005. .

Antje Inness, Melanie Ades, Anna Agustí-Panareda, Jérôme Barré, Anna Benedictow, Anne-Marlene Blechschmidt, Juan Jose Dominguez, Richard Engelen, Henk Eskes, Johannes Flemming, et al. The cams reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, 19(6):3515–3556, 2019.

K. Block, M. Haghighatnasab, D. G. Partridge, P. Stier, and J. Quaas. Cloud condensation nuclei concentrations derived from the cams reanalysis. *Earth System Science Data*, 16(1):443–470, 2024. . URL `https://essd.copernicus.org/articles/16/443/2024/`.

L.A. Lee, K.J. Pringle, C.L. Reddington, G.W. Mann, P. Stier, D.V. Spracklen, J.R. Pierce, and K.S. Karslaw. The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry and Physics*, 13:8879—-8914, 2013. .

Eyal Winter. The shapley value. *Handbook of game theory with economic applications*, 3:2025–2054, 2002.

Youngchae Kwon, Seung A An, Hyo-Jong Song, and Kwangjae Sung. Particulate matter prediction and shapley value interpretation in korea through a deep learning model. *SOLA*, 19:225–231, 2023.

Kai Jeggle, David Neubauer, Gustau Camps-Valls, and Ulrike Lohmann. Understanding cirrus clouds using explainable machine learning. *Environmental Data Science*, 2:e19, 2023.

Yichen Jia, Hendrik Andersen, and Jan Cermak. Analysis of cloud fraction adjustment to aerosols and its dependence on meteorological controls using explainable machine learning. *EGUsphere*, 2023:1–25, 2023.

Po Lun Ma and Panagiotis Stinis. 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. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), 2020.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020. .