

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

We appreciate the constructive comments. We have clarified these points in the revised manuscript, as detailed below.

1. Reviewer 1

RC: *Response to Major Comment 1: The fact that calculating the ASD for each species in MAM as a comparison test requires strong assumptions (really just 1 assumption, that the ASD is fixed for each species globally) is the whole point of doing it. This calculation is to test how much better the ML method does compared to the simplest assumption. If the ML method is not much better than the simplest assumption, then the ML method is not creating much value (but you can really highlight how much value it's adding if the ML estimates make the estimates much closer to ML than the simplest calculation).*

*This calculation *is* straightforward. For each species, calculate the globally averaged fraction of that species' mass that appears in each of the 7 modes. Let's say that globally averaged, 5% of sulfate is in mode 1, 10% in mode 2, etc. You would then apply these fractions to the bulk model (doing it for all species) to split the bulk mass across the 7 modes. You would also calculate the globally averaged diameter for each of your modes and apply this to the bulk model globally.*

Yes, it is a big assumption, but it's the "baseline" calculation to see if MAMnet adds information above a 1st-order estimate. And it is a straightforward calculation that could have been done in a day's work during the responses. The authors seemed to have overthought the calculation, and the article will be weaker without this baseline comparison.

AR: We appreciate the reviewer's comment and detailed explanation. We agree that some applications may not require a fully time and space varying ASD, and that using a characteristic average ASD might be sufficient. For example, one could argue that aerosol–radiation interactions over long timescales (e.g., a century) are not strongly affected by short-term variations in the ASD. In contrast, processes like scavenging and aerosol activation are highly sensitive to the shape of the ASD and can significantly impact cloud evolution.

It is clear that MAMnet provides added value over a fixed, constant ASD assumption, making the representation more realistic, as it is well-known that the ASD varies significantly in time and space. Whether this added complexity improves the representation of downstream processes depends on the specific application.

To address the reviewer's concern, we have added the following paragraph to the Conclusions section:

MAMnet is designed to improve the approximation of the ASD within retrieval algorithms and atmospheric models. Whether this added sophistication is necessary depends on the downstream application. In many cases, a simple prescribed ASD may be sufficient.

RC: *Responses to specific comments: Again, a bulk model does not "inherently" need to treat particles as*

externally mixed. Many models assume this (e.g., GOCART), but you can easily assume internal mixing. If your model is predicting 5 ug m-3 of sulfate and 5 ug m-3 of organics, one can easily assume that the particles are internally mixed as a 50-50 mixture of the two species. This can be a minor technical correction in the manuscript

AR: Thanks for the comment. We agree there may be some ambiguity in defining externally and internally mixed aerosol. In this point and in our response we followed the definition adopted by definition of [Riemer et al., 2019]. However, to avoid confusion however we have rewritten the statements as:

The bulk mass approach predicts the transport and evolution of aerosols by tracking the mass concentration of individual chemical species... Each particle is assumed to consist of a single chemical component or their surrogate.

... Because they predict the number concentration and mass independently, modal schemes can better resolve the composition of aerosol species, particularly when several subpopulations are used.

RC: *I was not suggesting that I thought that MAMnet emulated aerosol processes (I'm not sure what I said that gave that impression), but rather that using SO₂, NO_x, and solar radiation as inputs would give ML information on how the size distribution should look different (because certain processes were more likely to be happening). I think there would be a lot of power in this approach of including non-aerosol inputs to help predict aerosol size distributions moving forward. The authors could add this in as future work, but I'm not tied to this change.*

AR: Thank you for the comment, and we apologize for the misunderstanding. We agree that including additional input variables, particularly aerosol-related ones, as the reviewer suggests, could likely improve the network's accuracy. However, we intentionally chose a minimal setup that can be more easily applied in satellite retrievals and assimilation models.

That said, we now acknowledge in the Conclusions section that there is room for improvement:

Including additional input variables, such as gaseous species or solar radiation, may make MAMnet predictions more physically interpretable.

2. Reviewer 2

We thank the reviewer for the additional assessment.

RC: *1) Please ensure consistent naming of organic and ammonium-related tracers across text, tables, and figures. For example, Table 1 defines SU as "Sulfates, includes ammonium" and defines OG as "Primary and secondary organics", yet later discussion/figures refer to SOA and AMM (e.g., "SOA_ACC", "AMM_FSS", "AMM_CSS"). If these are internal GEOS/MAM variable names, please add a short mapping (e.g., "SOA<->OG", "AMM?") to avoid confusion for readers. If ammonium is present, what about nitrate? And several other acronyms that are clearer but not strictly defined.*

AR: Thanks for pointing this out. This is a semantic mismatch between the original MAM7 variables and the MERRA2 aerosol composition. To clarify it we have modified the definitions as follow:

1. MAM7 refers to organics as secondary organic matter (SOM). MERRA-2 uses Organics to refer to both primary and secondary organic matter (OG). Internally GEOS uses SOA. To avoid confusion with previous work we have decided to use “SOM” and clarified in the text how it relates to previous work. We have made ammonium explicit in the notation to remain consistent.
2. Nitrates are not considered in our version of MAM7 neither in MERRA-2. We now acknowledge this limitation.

RC: *2) The manuscript notes that some mass tracers are very small at high altitude ($\sim 10^{-20}$ kg/kg) and that errors are “exacerbated in logarithmic space”. Please add one or two explicit sentences describing the handling of near-zero/very small values (e.g., clipping floor prior to log transforms, how “missing/zero” is treated) and clarify which figures/metrics are computed in linear vs. log space. This may improve interpretability of mean log bias patterns in clean or remote regimes?*

AR: This is explained in Section 2.2.1 and has now been expanded as follows. Mass and number are floored using lower limits of, $10^{-20} \mu\text{g kg}^{-1}$ and 10^{-4}mg^{-1} , respectively, prior to training. Values below these thresholds are held constant and therefore do not contribute to the gradient. During testing, points below the thresholds are masked. All metrics are computed in logarithmic space.

RC: *3) Several surface sites are used for evaluation include high-altitude sites. A short reminder in the methods/caption about how model fields are collocated would prevent misinterpretation of near-surface biases at elevated stations. It would be interesting to discuss N30 in section 3.3.1, relating to Figure 8. The ML model lands within the observed variability most of the time but there is a consistent low bias at low altitude. Most likely this is also the case in GEOS-MAM?*

AR: Thank you for raising this point. We train and evaluate MAMnet on model levels, which always begin at the surface. For near-surface comparisons, we use the average of the two lowest model layers, corresponding to roughly 200 m altitude.

We also recognized the importance of assessing whether the biases seen in MAMnet are inherited from the training data. To address this, we have added a new figure (now Fig. 9) in which GEOS+MAM7 data are collocated at each observational site. This comparison is not exact, as the time periods differ and the free-running model does not include aerosol assimilation. Nevertheless, it allows us to assess whether GEOS+MAM7 exhibits systematic biases that could explain those seen in MAMnet. The new figure shows that part of the bias is indeed inherited from GEOS+MAM7. Using MERRA-2 as input does not substantially increase the bias, but it does markedly reduce the variability. These results are now described in Section 3.3.2.

RC: *4) Lines 453: “All datasets reveal similar spatial patterns”. I would argue that the CALIOP based CCN does not have a similar pattern. The zonal mean over ocean in the CALIOP dataset would be minimum at the Equator, while all the others have zonal mean over ocean that minimize near Antarctica.*

AR: The statement has been modified to: “All datasets display on average lower CCN concentrations over oceans, particularly in polar regions, and higher concentrations over land particularly in central and eastern Asia, Europe, and the Americas.”

RC: *5) please add and comment on lines for GEOS-MAM training or validation data. Since GEOS-MAM has double moment microphysics, we might expect this to look like GiOcean/CALIOP which presumably reflects reality better than MERRA2. If I’m right, this begs the question of whether in future work it might be possible to tweak MAMnet to retain the correct shape? Maybe by allowing air density to play a stronger role?*

AR: It is difficult to determine which CCN dataset best represents reality, as all of them rely on strong assumptions, including GiOcean. We don't know what the correct shape looks like at this point. We therefore prefer to refrain from making such a judgment at this stage. Further research is needed before these datasets can be robustly evaluated and compared. This is acknowledged in the section.

RC: *6) Lines 357-358: suggest rephrase “regions with negligible aerosol concentrations typically do not contribute significantly to atmospheric processes.”, taken out of context this doesn't seem correct.*

AR: We agree this is speculative. To avoid confusion the statement has been removed.

References

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. .