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: ***This manuscript uses a global aerosol microphysics model to train a neural network model to estimate aerosol size distributions and mixing state from bulk aerosol masses. This is overall a useful contribution. However, I feel like the overview of microphysics methods and understanding needs to be improved, and I'd like the authors to evaluate their results against the typical way of estimating the size distribution and CCN from bulk masses. Once these and specific issues have been addressed, I am supportive of this manuscript being published. Editorial note: Often, the figures are quite far from where they are discussed in the manuscript, which requires a lot of scrolling or flipping. In most cases, it seems like it would have been straightforward to have arranged the figures to be closer to their discussion.***

AR: We appreciate the positive assessment. We have reorganized the figures to facilitate reading.

1. Major comment

RC: ***I feel that the paper is missing the 1 test of ML size distributions that I'd want to see. The easiest (and usual) way to get size distributions (and CCN etc.) from a bulk model is simply to assume a fixed size distribution for each species. In your case, it would be good to just use the global average distributions from MAM as a global conversion from bulk to a size distribution. I'd like to know how much better MAMnet is compared to this simplest approximation, which to me is the test of the value of the ML.***

AR: Thank you for raising this important point. We agree that comparing MAMnet to a simpler baseline like using a fixed global-average ASD is in principle a useful way to assess the added value of the machine learning approach. However, implementing such a comparison is not straightforward. There is no unique or well-defined way to extract a global mean ASD and associated mixing state from the output of MAM, as both properties vary significantly with location, time, and aerosol source. Defining a representative "mean" ASD would require strong assumptions about how to average across modes, species, and atmospheric conditions, which could introduce biases of their own. Moreover, mixing state is inherently a modal property in MAM, and collapsing it into a fixed global approximation would overlook important regional distinctions.

While we do not present a direct global comparison using this method, our results in Section 3.1.2 and Figure 8 provide indirect insight into this issue. In particular, Figure 8 shows that datasets such as GiOcean, MERRA-2, and CAMS, despite being based on similar aerosol mass constraints, produce substantially different CCN concentrations. These differences, which exceed those reported for aerosol mass concentrations [Gueymard and Yang, 2020], likely arise from the use of fixed or simplified ASD assumptions in bulk schemes. This underscores the sensitivity of CCN and other number-based diagnostics to the specification of the ASD.

To clarify this pointy we have added the following paragraph to the introduction:

"In computationally intensive applications and satellite retrievals it is desirable to maintain the efficiency and simplicity of the bulk schemes. 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 [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."

2. Specific comments

RC: ***L8: This needs more information than "two-moment". You track two moments for 7 different modes, so this is a two-moment *modal* scheme. There are just plain two-moment (or X-moment) schemes (without assuming a modal shape, e.g., the MATRIX scheme in the GISS climate model) or two-moment sectional schemes (two moments in each size section; e.g., Adams and Seinfeld, 2002 referenced in the manuscript), so I wasn't sure what you were referring to when originally reading the abstract.***

AR: We have rewritten the abstract to better reflect the nature of the ASD as multi-modal, as follows: "In this work, we develop a neural network model, MAMnet, to predict the ASD and mixing state for seven lognormal modes based on the bulk aerosol mass and the meteorological state."

RC:

1. ***L22: Bulk models don't need to have bins (in the model you describe later, most species don't have bins). The key is that even if there are bins, there are no microphysics calculations (nucleation, condensation, coagulation, etc.) that would let the size distribution evolve.***
2. ***L23: Bulk models do not need to assume external mixing. You can assume that all at any given size (would need to assume a size distribution), all species are mixed into the same particles (internal mixing). Just because GOCART assumes external mixing doesn't mean that you need to assume external mixing with a bulk model.***

AR: Thanks for the comment. Externally mixed in our context does not refer to pure species, but rather to the fact that each particle is assumed to be composed of a single chemical component, which can still be a mixture of different materials. We have rewritten the paragraph clarifying the definition of bulk aerosol schemes as follows:

"The bulk mass approach 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]. 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."

RC:

1. ***L32: two moments of the ASD *for each mode*.***
2. ***L34: Again, bulk schemes can "handle" internal mixing (you just assume it). What modal schemes can do (assuming they are simulating multiple modes) is to have an explicit calculation of which***

particles are internally vs. externally mixed that varies in space and time. You also haven't established that models can have multiple modes yet.

AR: We have rewritten the paragraph to clarify the definition of modal aerosol schemes, as follows:

"In contrast to bulk methods, modal aerosol schemes estimate both the number concentration and mass of atmospheric aerosol, approximating the ASD as the combination of overlapping populations, termed modes, each typically following a log-normal distribution [e.g., Whitby and McMurry, 1997, Wilson et al., 2001, Stier et al., 2005, Liu et al., 2012]. Because they predict the number concentration and mass independently, modal schemes can resolve the mixing state of aerosol species particularly when several subpopulations are used [Riemer et al., 2019]. This leads to an improved 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]."

RC: ***L46 and throughout. Like the other reviewer said, it's common with MAM to put the number of modes after (in this case MAM7). However, I believe there may be multiple MAM configurations with 7 modes in the literature (e.g., I believe that one has a nucleation mode, and this one does not).***

AR: Thanks for pointing this out. We have clarified throughout the paper that the MAM7 model is used. The particular configuration used in our work is detailed in Table 2.

RC: ***L45: There is other work to parameterize bulk aerosol mass to the size distribution, such as `https://doi.org/10.1029/2021GL094133` and `https://doi.org/10.5194/acp-23-5023-2023`.***

AR: Thanks for pointing these out, they are now discussed in the introduction.

RC:

1. ***L97: "From these simulation*s*"***
2. ***L110: "a a"***
3. ***L116: My brain wants to read "During training data" as one phrase. Please add a comma after "training".***

AR: Corrected.

RC: ***L168-177: Please elaborate more on how these methods estimate CCN. Do you expect them to be better than your model estimates? I would guess a lot of assumptions go into these products, so I wouldn't necessarily expect them to be a useful evaluation. (Note: Figures 8 and 9 show huge disagreements between the datasets.)***

AR: Thank you for this comment. The datasets shown in Figures 8 and 9 (MAMNet-MERRA2, CALIOP, CAMS, and GIOcean) estimate CCN using different methods, each based on bulk aerosol mass and simplified assumptions about the ASD and hygroscopicity. For example, GiOcean uses the GOCART bulk aerosol scheme and assumes prescribed lognormal size distributions for each species, to make an online estimation of N_{CCN}. CAMS is also based on a bulk scheme but uses a different assumption for the ASD and performs an offline calculation of N_{CCN}. CALIOP derives N_{CCN} from pre-calculated conversion factors. These products do not resolve modal or particle-resolved microphysics and rely on fixed tuned parameters to infer N_{CCN}. We also included the GASSP CCN dataset, which compiles aerosol and CCN measurements from 37 field campaigns.

We agree that these assumptions introduce uncertainties and that none of these datasets can be considered a definitive reference for CCN. Rather than treating these datasets as ground truth, we use them to demonstrate that MAMnet produces CCN estimates that are consistent with the range of values reported in the literature. They illustrate the spread in CCN estimates arising from different assumptions about ASD and composition.

This is now clarified in the section.

RC: ***L185: [-1, 1] would be the range for "within an order of magnitude of the target value". [-0.5, 0.5] is an order of magnitude window around the target value (or within a factor of about 3.2).***

AR: Corrected.

RC: ***Figure 2: (1) The colormap is a strange choice for a diverging colorbar in the lower panel. It would be better to make it white in the middle. (2) Would be easier to interpret if it were rotated such that height was the y axis. (3) Have the surface (1000 hPa) be at the origin rather than the model top.***

AR: Thanks for the suggestions. The figure has been updated.

RC: ***Figure 2 and discussion around line 200: I suspect that the challenge in MAMnet predicting the Aitken mode may stem from the difficulty of predicting when/where nucleation is occurring. If other inputs that may help predict nucleation, like solar radiation and SO2, were included, it might do a better job with the Aitken mode. Also possible is that fresh fossil-fuel combustion emissions (vs. aged in the accumulation mode) into the Aitken mode might be hard to predict, and NOx as an input might help with this as the NOx lifetime is on a similar order as a typical aging timescale (12-24 hours) and they tend to be co-emitted.***

AR: This is an important point. However, 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. It acts as a surrogate for the output of the host model, rather than replicating physical processes such as nucleation or chemical aging.

During testing, MAMnet learns how the mass of each species is distributed across the seven modes. While this distribution is influenced by processes like nucleation and fresh emissions, these are not explicitly emulated. It is likely that events such as new particle formation or fresh fossil-fuel combustion, where number concentrations increase sharply with minimal changes in mass, exacerbate the biases in MAMnet, since they may lead to class imbalance. This is now acknowledged in the paper.

RC: ***L234-236: This sentence overstates things. Relative variability in Dpg is strongly buffered to relative variability in Mass/Number since it goes with the cube root of this ratio. For example, factor-of-2 error in M/N would only be a 26% error in Dpg. Is the MLB of 0.01 really that remarkable or surprising given that it's much more stable than M/N?***

AR: We agree that variability in D_{pg} is buffered relative to variability in the mass-to-number ratio, since it scales with the cube root of M/N. However, the consistency of D_{pg} remains a meaningful indicator of model performance. This buffering effect holds only if mass and number vary coherently. 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 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.

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. We now acknowledge in the manuscript that D_{pg} is relatively insensitive to errors in M/N, but still serves as a useful test of internal consistency and mass conservation in the model.

RC: ***L255: Do we expect the observational constraints (just polar-orbiter AOD in cloud-free regions, right?) on MERRA-2 to improve the relative balance of species masses? My understanding is it just scales the mass of all species in the column up/down until AOD is pushed closer towards the obs.***

AR: Thank you for the comment. You are correct that MERRA-2 assimilates satellite-retrieved AOD, primarily over cloud-free regions, adjusting the total aerosol mass in the column to better match observations. This is done through the Goddard Aerosol Assimilation System, which uses AOD observations from multiple sensors [Randles et al., 2017]. While this assimilation primarily constrains the total aerosol optical depth, it is carried out within a model framework (GEOS) that includes species-specific mass and optical properties, and thus indirectly affects the relative contributions of different aerosol species.

In addition, MERRA-2 benefits from the global meteorological data assimilation system, which ingests over six million observations every six hours [Gelaro et al., 2017]. This constrains winds, humidity and temperature, that influence aerosol transport and evolution, improving the realism of the aerosol spatial distribution. This bring the aerosol mass concentrations in MERRA-2 closer to observed values than what would be obtained from a free-running model [e.g., Buchard et al., 2017, Sun et al., 2019, Gueymard and Yang, 2020, Su et al., 2023].

Our goal in this section is not to evaluate the details of the MERRA-2 assimilation system, but rather to demonstrate that when MERRA-2 fields are used as inputs to MAMnet, and the predicted species are summed (as in Figure 1), the resulting bias is very small. While we initially presented this as evidence that MAMnet does not inherit the biases of GEOS+MAM7, it also serves as a validation of mass conservation and internal consistency in the model, as the total mass of each species is preserved.

To clarify this point earlier in the manuscript, we have moved Figure 6 and the associated discussion to the beginning of Section 3.1, before Figures 2 and 3.

RC: ***L271: How did you sample the model for the high-altitude sites? These sites are tricky since they are often at a much higher altitude than the gridbox mean altitude. Sometimes they are in the PBL, sometimes they aren't. I recommend just leaving them out.***

AR: Thank you for raising this point. For all sites, N_{XX} is calculated by interpolating the MERRA-2 fields in space and time to the site location and sampling the values on hybrid sigma-pressure levels [Gelaro et al., 2017], and using MAMnet to predict the ASD. We then compute the number concentration using Eq. 4, and report the average over the two lowermost model levels, which correspond to the surface layer in the model.

High-altitude stations are still treated as surface sites, but we acknowledge that the relatively coarse spatial resolution of MERRA-2 may not fully resolve the site's actual terrain height. As a result, there may be mismatches between the model's surface elevation and the true elevation of the site. This can lead to sampling biases, especially in complex terrain where vertical gradients are strong.

We retain these sites in our analysis to preserve spatial diversity. We have clarified this point in Section 2.1.

RC: ***Figure 6 and some other discussions: Is there a way to make MAMnet conserve the mass of the inputs? This seems like a critical thing to do. Also, I recommend ug kg-1 or sm-3 rather than kg kg-1 since people are used to thinking of aerosol masses in ug m-3.***

AR: While we don't explicitly include a mass conservation constraint in the model architecture, we show that MAMnet conserves mass. As mentioned above, we show in Figure 6 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.

This demonstrates that the model conserves mass, as it accurately maintains the total mass of each species. Additional support for this comes from the modal geometric mean diameter, $D_{pg,i}$, as it 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. In a online implementation MAMnet would be diagnostic to the online bulk model, hence would not directly influence mass conservation.

This clarification has been added to the section. There is also a typo in the figure as the units must be kg m^{-2}. This has been corrected.

RC: ***Figure 7: Is this any better or worse than how MAM itself does? I suspect they both have similar issues.***

AR: We agree that biases in the training data (GEOS+MAM7) can influence the learned mapping, and that the skill of the input data (MERRA-2) affects the final output. However, we do not present comparisons between GEOS+MAM7 and observations at specific sites, since such comparisons are not meaningful for a free-running model. We use GEOS+MAM7 as the training dataset because it provides internally consistent mass and number concentrations needed to learn the relationship between them. Instead, we emphasize that comparisons between MERRA-2+MAMnet and observations reflect the model's performance when driven by realistic aerosol fields. This also allows us to assess its performance in a more observationally constrained setting.

We now clarify this in the section:

"MERRA-2 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, Su et al., 2023], and thus provides a more realistic input compared to free-running model simulations. Although GEOS+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."

RC: ***Figures 8 and 9: Are these products that are used for comparison any good? They vary so much, and I'm guessing that there are a lot of assumptions that go into getting CCN from the products.***

AR: This comment has been clarified above. Rather than treating these datasets as ground truth, we use them to demonstrate that MAMnet produces CCN estimates that are consistent with the range of values reported in the literature.

RC: ***Figure 9: Please add a legend to the figure rather than stating the colors in the caption.***

AR: Corrected.

RC: ***L313: Please explain what a Shapley value is. What does a high or low feature value mean?***

AR: It means that inputs with larger SHAP values contribute more significantly to the model output. The following paragraph has been added to the section:

"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].”

RC: ***L348-349: Isn’t there a way to just force MAMnet to conserve the mass of the inputs?***

AR: As discussed above, we do not explicitly constrain MAMnet to conserve mass during training. However, we demonstrate that it effectively conserves mass in practice, as shown by the close agreement between the input mass and the reconstructed total mass across modes (e.g., Figure 6). In an online implementation, MAMnet would act as a diagnostic component within a bulk aerosol scheme and would not influence the model’s mass budget directly.

RC: ***L352-353: Are these better than the reference test (fixed size dist for each species) that I described above?***

AR: As discussed above, constructing a fixed ASD involves strong assumptions about how to average across space, time, modes, and species. These assumptions can vary widely and would significantly influence the results. Because of this, it is difficult to define a fair or representative baseline for comparison. Whether MAMnet performs better or worse could largely depend on how the fixed ASD is constructed. For this reason, we did not include such a reference test in our evaluation.

RC: ***L356-359: Like my earlier comment, “exceedingly well” is an overstatement. Dpm is buffered to errors in Mass/Number.***

AR: The statement has been removed.

References

Christian A Gueymard and Dazhi Yang. Worldwide validation of cams and merra-2 reanalysis aerosol optical depth products using 15 years of aeronet observations. *Atmospheric Environment*, 225:117216, 2020. .

John H Seinfeld and Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2016.

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. .

D. Barahona, A. Molod, J. Bacmeister, A. Nenes, A. Gettelman, H. Morrison, V. Phillips, and A. Eichmann. Development of two-moment cloud microphysics for liquid and ice within the NASA Goddard Earth Observing System Model (GEOS-5). *Geosc. Model Dev.*, 7(4):1733–1766, 2014. . URL `http://www.geosci-model-dev.net/7/1733/2014/`.

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/`.

ADLA Jones, DL Roberts, and A Slingo. A climate model study of indirect radiative forcing by anthropogenic sulphate aerosols. *Nature*, 370(6489):450–453, 1994.

J Langner and H Rodhe. A global three-dimensional model of the tropospheric sulfur cycle. *Journal of Atmospheric Chemistry*, 13:225–263, 1991.

Paul Ginoux, Mian Chin, Ina Tegen, Joseph M Prospero, Brent Holben, Oleg Dubovik, and Shian-Jiann Lin. Sources and distributions of dust aerosols simulated with the gocart model. *Journal of Geophysical Research: Atmospheres*, 106(D17):20255–20273, 2001.

Mian Chin, Richard B Rood, Shian-Jiann Lin, Jean-Francois Müller, and Anne M Thompson. Atmospheric sulfur cycle simulated in the global model gocart: Model description and global properties. *Journal of Geophysical Research: Atmospheres*, 105(D20):24671–24687, 2000.

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. .

Evan R Whitby and Peter H McMurry. Modal aerosol dynamics modeling. *Aerosol Science and Technology*, 27(6):673–688, 1997.

Julian Wilson, Cornelis Cuvelier, and Frank Raes. A modeling study of global mixed aerosol fields. *Journal of Geophysical Research: Atmospheres*, 106(D24):34081–34108, 2001.

P Stier, J Feichter, S Kinne, S Kloster, E Vignati, J Wilson, L Ganzeveld, I Tegen, Martin Werner, Yves Balkanski, et al. The aerosol-climate model echam5-ham. *Atmospheric Chemistry and Physics*, 5(4): 1125–1156, 2005.

Xiaohong Liu, Richard C Easter, Steven J Ghan, R Zaveri, P Rasch, Xiangjun Shi, J-F Lamarque, A Gettelman, H Morrison, Francis Vitt, et al. Toward a minimal representation of aerosols in climate models: Description and evaluation in the community atmosphere model cam5. *Geoscientific Model Development*, 5(3):709–739, 2012.

Peter J Adams and John H Seinfeld. Predicting global aerosol size distributions in general circulation models. *Journal of Geophysical Research: Atmospheres*, 107(D19):AAC–4, 2002.

Michael Herzog, Debra K Weisenstein, and Joyce E Penner. A dynamic aerosol module for global chemical transport models: Model description. *Journal of Geophysical Research: Atmospheres*, 109(D18), 2004.

Linyi Wei, Zheng Lu, Yong Wang, Xiaohong Liu, Weiyi Wang, Chenglai Wu, Xi Zhao, Stefan Rahimi, Wenwen Xia, and Yiquan Jiang. Black carbon-climate interactions regulate dust burdens over india revealed during covid-19. *Nature communications*, 13(1):1839, 2022.

C. A. Randles, A. M. da Silva, V. Buchard, P. R. Colarco, A. Darmenov, R. Govindaraju, A. Smirnov, B. Holben, R. Ferrare, J. Hair, Y. Shinozuka, and C. J. Flynn. The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation. *Journal of Climate*, 30(17):6823–6850, 07 2017. ISSN 0894-8755. . URL `https://doi.org/10.1175/JCLI-D-16-0609.1`.

R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, 30(14):5419–5454, 2017. .

V Buchard, CA Randles, AM Da Silva, A Darmenov, PR Colarco, R Govindaraju, R Ferrare, J Hair, AJ Beyersdorf, LD Ziemba, et al. The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies. *Journal of Climate*, 30(17):6851–6872, 2017.

Enwei Sun, Huizheng Che, Xiaofeng Xu, Zhenzhu Wang, Chunsong Lu, Ke Gui, Hujia Zhao, Yu Zheng, Yaqiang Wang, Hong Wang, et al. Variation in merra-2 aerosol optical depth over the yangtze river delta from 1980 to 2016. *Theoretical and Applied Climatology*, 136(1):363–375, 2019. .

Xin Su, Yuhang Huang, Lunche Wang, Mengdan Cao, and Lan Feng. Validation and diurnal variation evaluation of merra-2 multiple aerosol properties on a global scale. *Atmospheric Environment*, 311:120019, 2023. .

Alexander Ukhov, Suleiman Mostamandi, Arlindo Da Silva, Johannes Flemming, Yasser Alshehri, Illia Shevchenko, and Georgiy Stenchikov. Assessment of natural and anthropogenic 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, 2020. .

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. .