

Pernov et al. presents a data-driven approach to model particulate methanesulfonic acid (MSAp) concentrations in the Arctic using Random Forest (RF) and Additive Models (AM). The work is interesting because it combines several data sources and uses careful feature engineering to capture different environmental condition. I would suggest the paper to be published after improvement of the following points by the authors to make their discussion deeper:

We thank the reviewer for their comments and suggestions. We have addressed each comment below with review comments in black, author response in blue, and additions to the original text in red. We have indented the author's response for clarity. Line numbers given in the author's response refer to lines in the revised manuscript.

1. The paper mentions the atmospheric oxidation of dimethyl sulfide (DMS) to form MSAp but does not go into detail about how gas-phase and aqueous-phase oxidation work. I would suggest the authors to add a brief discussion on these processes and explain how they change under different Arctic conditions.

The reviewer is correct that we could have added more details about the mechanisms of gas and aqueous phase oxidation of DMS into MSA_p. Therefore, we have added the following lines about the mechanisms, intermediates, and unique aspects relevant to the Arctic.

Lines 66-87: DMS oxidation to MSA_p involves a variety of gas- and aqueous-phase reactions, the latter occurring in cloud droplets or on deliquesced particles (Barnes et al., 2006a; Chen et al., 2018; Fung et al., 2022; Hoffmann et al., 2016). DMS is first oxidized through two major branches. One is the abstraction pathway by reactions with OH, NO₃, and Cl radicals in the gas-phase to yield methylthiomethylperoxy radical (MTMP: CH₃SCH₂OO) (Berndt et al., 2019; Hoffmann et al., 2016). MTMP can undergo isomerization to form hydroperoxymethylthioformate (HPMTF) (Berndt et al., 2019; Veres et al., 2020), or oxidation by NO or RO₂ to produce CH₃SO₂, which can then form SO₂, sulfuric acid (H₂SO₄), or MSA with strongly temperature-dependent yields (Berndt et al., 2023; Chen et al., 2023; Shen et al., 2022). The other DMS oxidation branch is the addition pathway through reactions with OH, BrO, Cl, as well as O₃ to yield dimethyl sulfoxide (DMSO), which occurs mainly through gas-phase but partly in the aqueous-phase through reaction with O₃ (Hoffmann et al., 2016). DMSO is a semi-volatile species which reacts with OH in both gas- and aqueous-phase to form methanesulfinic acid (MSIA). MSIA then reacts with OH or O₃ in the aqueous-phase to produce MSA_p (Chen et al., 2018; von Glasow and Crutzen, 2004; Hoffmann et al., 2016; Wollesen de Jonge et al., 2021), although it can also undergo gas-phase oxidation by OH to yield CH₃SO₂ and thus contributing to the temperature-dependent pathway to produce gaseous MSA (Chen et al., 2023;

Shen et al., 2022). The produced gas-phase MSA can condense on cloud droplets or existing particles to form MSA_p (Hoffmann et al., 2016). Aqueous phase reactions are dominant formation mechanisms for MSA_p in a typical marine boundary layer condition (Baccarini et al., 2021; Chen et al., 2018; Hoffmann et al., 2016; Kecorius et al., 2023). In the Arctic, cold temperatures (Barnes et al., 2006a; Chen et al., 2023; Shen et al., 2022) and elevated halogen levels will favor MSA/MSA_p formation relative to SO_2 (Chalif et al., 2024; Jongebloed et al., 2023; Sørensen et al., 1996) especially in the springtime, through both gas- and aqueous-phase pathways. The presence of ice containing clouds may limit the aqueous-phase production since reactions between DMS and its intermediates and oxidants are mainly able to occur at the surface (Chen et al., 2018). For a full description of DMS oxidation mechanisms and pathways see Barnes et al. (2006a).

2. The paper uses RF and AM to analyze the data, but I would suggest the authors to discuss more clearly the strengths of these methods. For example, RF is useful for capturing non-linear relationships and provides feature importance, while AM offers a simple way to understand how each variable affects MSA_p . A clear discussion on these points would help readers see why these methods were chosen.

We have added a new Section (2.6.3) to clarify and discuss further the respective strengths and limitations of both RF and AM approaches. Also, Supplementary Text 2 and Figure S1 provide summarized results and explanations about other approaches we tested. As explained in the first paragraph of Section 2.6, these were not retained mainly because they performed worse in terms of prediction error.

Lines 540-558 (new Section 2.6.3 titled Strengths and Limitations of RF and AM): Both RF and AM are non-linear regression models although they differ in a few key aspects. First, RF's output is the average of predictions from many decision trees which are based on random subsets of the training data and random subsets of features, while AM considers the entire training set at once (i.e., without random subsets). This randomization generally reduces the risk of overfitting (yielding a smaller prediction variance) compared to constructing a single, large decision tree. For AM, the risk of overfitting was minimized by keeping the number of splines bases low ($J=5$) and enforcing this stepwise variable selection scheme so that the number of parameters P stayed relatively low. In that sense, AM is generally a simpler model than RF. Second, the predictions from decision trees, and thus from RF, seen as a mathematical mapping from a feature space to a target variable space, are piecewise constant functions. By contrast, the predictions from AM are by design smooth, as they are computed as the sum of cubic splines (with

continuous second derivatives). In practice, this means that the predicted target surface from RF looks like jagged stairs, with jumps at feature splits, while for AM this looks like a smooth surface. Finally, the additive structure of AM in Eq. (4) is quite constrained: features have their respective effects, and these add up to a prediction. We considered pairwise interactions but no higher-level terms (e.g., three-way interactions). By comparison, RF inherently can include higher level interactions, as splits are being added sequentially (i.e., conditioned on previous splits) when growing a decision tree (up to a maximum depth, which we tuned as a hyperparameter). This higher complexity makes RF generally more flexible than AM. Although this flexibility comes at the cost of harder interpretability as one cannot easily visualize the estimated effect of a feature on the target, specifically because of such interactions likely being different from tree to tree within the ensemble. The additive constraint of AM is what makes the estimated partial effects directly interpretable, say, as a curve displayed in a plot.

3. While the authors note the low R^2 values, I would suggest they discuss this further. They could explain which important variables might be missing and how these missing elements or measurement issues might affect the results. I would also suggest discussing future improvements, like using more advanced methods or combining physical models with data-driven approaches.

We have added some discussion about possible reasons for the low evaluation metrics. A common theme for all models and feature groups tested is that they under/over-predict high/low observations, therefore, we have added an example of how the models could under/over-predict high/low observations based on inadequacies of the input features.

Lines 762-767: For instance, models underpredict high observations of MSA_p which could be due to large DMS emissions not being captured in the input features either due to being based on a climatology of seawater DMS (Lana et al., 2011; Nightingale et al., 2000) or occurring further back in time than 5 days. The models also overpredict low MSA_p observations, which could be due to extreme precipitation events not being captured by ERA5 (Loeb et al., 2022) or being smoothed out in the feature engineering procedure. Although the summation was used as an aggregation method for precipitation therefore smoothing over one day should not affect extreme events (Table 2).

We have also added some text adding to our discussion of potential improvements to data-driven modeling for Arctic MSA_p predictions in the Conclusions.

Lines 1234-1244: When sufficient long-term high-resolution data becomes available, leveraging the power of other data-driven approaches (e.g., neural networks) could be an option for advancing data-driven modeling of Arctic aerosols. However, when limited by the sample size, less complex tree-based models and statistical methods can often perform on par or better than neural networks (Grinsztajn et al., 2022). One option would be to combine data driven modeling with physically constrained loss mechanisms, known as physics-informed neural networks (PINN) (Cuomo et al., 2022). This avenue could help ensure the proper ingredients (precursors, oxidants, meteorology, etc.) are present at sufficient levels and with the correct temporal occurrence. The multi-input/output functionality of neural networks could also help elucidate the branching ratio of DMS oxidation mechanisms and the partitioning between gaseous and particulate phase MSA. However, satisfactory long-term high resolution, concurrent measurements of gas-phase DMS, gas-phase MSA and particulate phase MSA need to become available at appropriate locations dispersed around the Arctic region, which remains a challenge both logistically and monetarily.

We have added text to show that aerosol chemical and physical properties are missing from our input features and how this would affect MSA partitioning between gas and condensed phase in the Conclusion.

Lines 1251-1256: Another missing component from our feature list is aerosol chemical and physical properties (e.g., surface area, mixing state, hygroscopicity, and composition), which largely determines the reactive uptake of gaseous MSA (and its intermediates) onto preexisting aerosols (Dada et al., 2022; Yan et al., 2020). Acidic and effloresced aerosols are less likely to uptake gaseous MSA while alkaline and deliquesced aerosols are more likely, including these parameters in future data-driven approaches could help resolve the equilibrium partitioning between gas and condensed phase MSA, thus representing another sink term for MSA_p .

4. The paper explains how the models work but could do more to show how each feature relates to the physical processes in the Arctic. I would suggest the authors to compare the RF feature importance with the partial effects from the AM, so readers can see the real-world significance of the results.

In Sect 3.4, we detail the features selected by each model and for each group of variables. We also describe how each model selects features. We show that each model selects themes of variables (source related, chemical processing, and removal) for most stations. We also discuss the differences between the variable

groups (Group A and A+B) selected by the models. In the following sections, we extensively discuss the source related, chemical processing, and removal themes of features selected as well as other features such as physical meteorology related and vertically oriented features. In the main text, we describe the partial effects and their physical interpretations for Alert and Utqiagvik/Barrow and summarize the remaining stations. In the Supplemental Information, we go into detail on the partial effects for the other station/dataset combinations. Therefore, we feel we have adequately described how the features selected by the models relate to physical processes in the Arctic. The reviewer suggests comparing the RF feature importance, which is a single number or global importance value, to the relationship displayed by the partial effects plots (or local importance values) from the AM. We explored partial dependency plots (which is a local feature importance method) for RF and ultimately opted not to present the results since they were similar to the partial effects from the AM although less reliable. While the AM partial effects work in a similar manner to partial dependency plots from a RF model, by keeping all other features at their average while varying a single feature across its distribution range to examine the relationship with model output, we are more confident in the AM results due to differences in features selection and model design (increased interpretability compared to RF, see Sect. 2.6.3). RF struggles with explainability when features are strongly correlated due to the random feature sub-selection during tree construction, this makes consistently determining feature importance difficult. AM does not struggle with correlated features due to selecting them sequentially and by considering the entire dataset of features once selected. This is evident by RF selecting features of the same variable type but with different timestamps compared to AM (Table 4). This would make the interpretation of partial dependency plots more complex compared to the partial effects of AM, whose feature selection and partial effects are more straightforward and reliable. This is due to the additive constraint of AM, which makes the partial effects directly interpretable. The AM was tailored for this specific project while the RF served as a baseline model for comparison. Therefore, we opted to only present results for the partial effects of AM in the manuscript.

- Baccarini, A., Dommen, J., Lehtipalo, K., Henning, S., Modini, R. L., Gysel-Beer, M., Baltensperger, U., and Schmale, J.: Low-Volatility Vapors and New Particle Formation Over the Southern Ocean During the Antarctic Circumnavigation Expedition, *J. Geophys. Res. Atmospheres*, 126, e2021JD035126, <https://doi.org/10.1029/2021JD035126>, 2021.
- Barnes, I., Hjorth, J., and Mihalopoulos, N.: Dimethyl Sulfide and Dimethyl Sulfoxide and Their Oxidation in the Atmosphere, *Chem. Rev.*, 106, 940–975, <https://doi.org/10.1021/cr020529+>, 2006a.
- Berndt, T., Scholz, W., Mentler, B., Fischer, L., Hoffmann, E. H., Tilgner, A., Hyttinen, N., Prisle, N. L., Hansel, A., and Herrmann, H.: Fast Peroxy Radical Isomerization and OH Recycling in the Reaction of OH Radicals with Dimethyl Sulfide, *J. Phys. Chem. Lett.*, 10, 6478–6483, <https://doi.org/10.1021/acs.jpcclett.9b02567>, 2019.
- Berndt, T., Hoffmann, E. H., Tilgner, A., Stratmann, F., and Herrmann, H.: Direct sulfuric acid formation from the gas-phase oxidation of reduced-sulfur compounds, *Nat. Commun.*, 14, 4849, <https://doi.org/10.1038/s41467-023-40586-2>, 2023.
- Chalif, J. I., Jongebloed, U. A., Osterberg, E. C., Koffman, B. G., Alexander, B., Winski, D. A., Polashenski, D. J., Stamieszkin, K., Ferris, D. G., Kreutz, K. J., Wake, C. P., and Cole-Dai, J.: Pollution drives multidecadal decline in subarctic methanesulfonic acid, *Nat. Geosci.*, 17, 1016–1021, <https://doi.org/10.1038/s41561-024-01543-w>, 2024.
- Chen, J., Lane, J. R., Bates, K. H., and Kjaergaard, H. G.: Atmospheric Gas-Phase Formation of Methanesulfonic Acid, *Environ. Sci. Technol.*, 57, 21168–21177, <https://doi.org/10.1021/acs.est.3c07120>, 2023.
- Chen, Q., Sherwen, T., Evans, M., and Alexander, B.: DMS oxidation and sulfur aerosol formation in the marine troposphere: a focus on reactive halogen and multiphase chemistry, *Atmospheric Chem. Phys.*, 18, 13617–13637, <https://doi.org/10.5194/acp-18-13617-2018>, 2018.
- Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F.: Scientific Machine Learning Through Physics-Informed Neural Networks: Where we are and What's Next, *J. Sci. Comput.*, 92, 88, <https://doi.org/10.1007/s10915-022-01939-z>, 2022.
- Dada, L., Angot, H., Beck, I., Baccarini, A., Quéléver, L. L. J., Boyer, M., Laurila, T., Brasseur, Z., Jozef, G., de Boer, G., Shupe, M. D., Henning, S., Bucci, S., Dütsch, M., Stohl, A., Petäjä, T., Daellenbach, K. R., Jokinen, T., and Schmale, J.: A central arctic extreme aerosol event triggered by a warm air-mass intrusion, *Nat. Commun.*, 13, 5290, <https://doi.org/10.1038/s41467-022-32872-2>, 2022.
- Fung, K. M., Heald, C. L., Kroll, J. H., Wang, S., Jo, D. S., Gettelman, A., Lu, Z., Liu, X., Zaveri, R. A., Apel, E. C., Blake, D. R., Jimenez, J.-L., Campuzano-Jost, P., Veres, P. R., Bates, T. S., Shilling, J. E., and Zawadowicz, M.: Exploring dimethyl sulfide (DMS) oxidation and implications for global aerosol radiative forcing,

Atmospheric Chem. Phys., 22, 1549–1573, <https://doi.org/10.5194/acp-22-1549-2022>, 2022.

von Glasow, R. and Crutzen, P. J.: Model study of multiphase DMS oxidation with a focus on halogens, *Atmospheric Chem. Phys.*, 4, 589–608, <https://doi.org/10.5194/acp-4-589-2004>, 2004.

Grinsztajn, L., Oyallon, E., and Varoquaux, G.: Why do tree-based models still outperform deep learning on typical tabular data?, in: *Advances in Neural Information Processing Systems*, 507–520, 2022.

Hoffmann, E. H., Tilgner, A., Schrödner, R., Bräuer, P., Wolke, R., and Herrmann, H.: An advanced modeling study on the impacts and atmospheric implications of multiphase dimethyl sulfide chemistry, *Proc. Natl. Acad. Sci.*, 113, 11776–11781, <https://doi.org/10.1073/pnas.1606320113>, 2016.

Jongebloed, U. A., Schauer, A. J., Cole-Dai, J., Larrick, C. G., Porter, W. C., Tashmim, L., Zhai, S., Salimi, S., Edouard, S. R., Geng, L., and Alexander, B.: Industrial-era decline in Arctic methanesulfonic acid is offset by increased biogenic sulfate aerosol, *Proc. Natl. Acad. Sci.*, 120, e2307587120, <https://doi.org/10.1073/pnas.2307587120>, 2023.

Kecorius, S., Hoffmann, E. H., Tilgner, A., Barrientos-Velasco, C., van Pinxteren, M., Zeppenfeld, S., Vogl, T., Madueño, L., Lovrić, M., Wiedensohler, A., Kulmala, M., Paasonen, P., and Herrmann, H.: Rapid growth of Aitken-mode particles during Arctic summer by fog chemical processing and its implication, *PNAS Nexus*, 2, pgad124, <https://doi.org/10.1093/pnasnexus/pgad124>, 2023.

Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., Dachs, J., Bopp, L., Saltzman, E. S., Stefels, J., Johnson, J. E., and Liss, P. S.: An updated climatology of surface dimethylsulfide concentrations and emission fluxes in the global ocean, *Glob. Biogeochem. Cycles*, 25, GB1004, <https://doi.org/10.1029/2010GB003850>, 2011.

Loeb, N. A., Crawford, A., Stroeve, J. C., and Hanesiak, J.: Extreme Precipitation in the Eastern Canadian Arctic and Greenland: An Evaluation of Atmospheric Reanalyses, *Front. Environ. Sci.*, 10, <https://doi.org/10.3389/fenvs.2022.866929>, 2022.

Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J., and Upstill-Goddard, R. C.: In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers, *Glob. Biogeochem. Cycles*, 14, 373–387, <https://doi.org/10.1029/1999GB900091>, 2000.

Shen, J., Scholz, W., He, X.-C., Zhou, P., Marie, G., Wang, M., Marten, R., Surdu, M., Rörup, B., Baalbaki, R., Amorim, A., Ataei, F., Bell, D. M., Bertozzi, B., Brasseur, Z., Caudillo, L., Chen, D., Chu, B., Dada, L., Duplissy, J., Finkenzeller, H., Granzin, M., Guida, R., Heinritzi, M., Hofbauer, V., Iyer, S., Kemppainen, D., Kong, W.,

- Krechmer, J. E., Kürten, A., Lamkaddam, H., Lee, C. P., Lopez, B., Mahfouz, N. G. A., Manninen, H. E., Massabò, D., Mauldin, R. L., Mentler, B., Müller, T., Pfeifer, J., Philippov, M., Piedehierro, A. A., Roldin, P., Schobesberger, S., Simon, M., Stolzenburg, D., Tham, Y. J., Tomé, A., Umo, N. S., Wang, D., Wang, Y., Weber, S. K., Welti, A., Wollesen de Jonge, R., Wu, Y., Zauner-Wieczorek, M., Züst, F., Baltensperger, U., Curtius, J., Flagan, R. C., Hansel, A., Möhler, O., Petäjä, T., Volkamer, R., Kulmala, M., Lehtipalo, K., Rissanen, M., Kirkby, J., El-Haddad, I., Bianchi, F., Sipilä, M., Donahue, N. M., and Worsnop, D. R.: High Gas-Phase Methanesulfonic Acid Production in the OH-Initiated Oxidation of Dimethyl Sulfide at Low Temperatures, *Environ. Sci. Technol.*, 56, 13931–13944, <https://doi.org/10.1021/acs.est.2c05154>, 2022.
- Sørensen, S., Falbe-Hansen, H., Mangoni, M., Hjorth, J., and Jensen, N. R.: Observation of DMSO and CH₃S(O)OH from the gas phase reaction between DMS and OH, *J. Atmospheric Chem.*, 24, 299–315, <https://doi.org/10.1007/BF00210288>, 1996.
- Veres, P. R., Neuman, J. A., Bertram, T. H., Assaf, E., Wolfe, G. M., Williamson, C. J., Weinzierl, B., Tilmes, S., Thompson, C. R., Thames, A. B., Schroder, J. C., Saiz-Lopez, A., Rollins, A. W., Roberts, J. M., Price, D., Peischl, J., Nault, B. A., Møller, K. H., Miller, D. O., Meinardi, S., Li, Q., Lamarque, J.-F., Kupc, A., Kjaergaard, H. G., Kinnison, D., Jimenez, J. L., Jernigan, C. M., Hornbrook, R. S., Hills, A., Dollner, M., Day, D. A., Cuevas, C. A., Campuzano-Jost, P., Burkholder, J., Bui, T. P., Brune, W. H., Brown, S. S., Brock, C. A., Bourgeois, I., Blake, D. R., Apel, E. C., and Ryerson, T. B.: Global airborne sampling reveals a previously unobserved dimethyl sulfide oxidation mechanism in the marine atmosphere, *Proc. Natl. Acad. Sci.*, 117, 4505–4510, <https://doi.org/10.1073/pnas.1919344117>, 2020.
- Wollesen de Jonge, R., Elm, J., Rosati, B., Christiansen, S., Hyttinen, N., Lüdemann, D., Bilde, M., and Roldin, P.: Secondary aerosol formation from dimethyl sulfide – improved mechanistic understanding based on smog chamber experiments and modelling, *Atmos Chem Phys*, 21, 9955–9976, <https://doi.org/10.5194/acp-21-9955-2021>, 2021.
- Yan, J., Jung, J., Zhang, M., Bianchi, F., Tham, Y. J., Xu, S., Lin, Q., Zhao, S., Li, L., and Chen, L.: Uptake selectivity of methanesulfonic acid (MSA) on fine particles over polynya regions of the Ross Sea, Antarctica, *Atmospheric Chem. Phys.*, 20, 3259–3271, <https://doi.org/10.5194/acp-20-3259-2020>, 2020.