

Continental river runoff enhances atmospheric aerosol formation over the Arctic Ocean

Corresponding Author: Dr Manuel Dall'Osto

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Decision Letter:

**** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors ****

Dear Dr Dall'Osto,

Please accept our sincere apologies for the unusual delay in sending a decision on your manuscript. Your manuscript titled "Continental river runoff enhances atmospheric aerosol formation over the Arctic Ocean" has now been seen by 3 reviewers, and we include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. Please highlight all changes in the manuscript text file.

Please submit your point-by-point responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and the completed checklist:

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

Please do not hesitate to contact us if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Keiichiro Hara, PhD
Editorial Board Member

Communications Earth & Environment
orcid.org/0000-0001-7440-7776

Alice Drinkwater, PhD
Associate Editor
Communications Earth & Environment
Consulting Editor
Communications Sustainability

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

For Manuscripts that fall into the following fields:

- Behavioural and social science
- Ecological, evolutionary & environmental sciences
- Life sciences

An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript

You can download the form here:

<https://www.nature.com/documents/nr-reporting-summary.pdf>

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist:

<https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf> Communications Earth & Environment formatting checklist

and also in our style and formatting guide <https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf> Communications Earth & Environment formatting guide.

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (<http://www.copdess.org/enabling-fair-data-project/>). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>.

In particular, the Data availability statement should include:

- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at <http://www.nature.com/sdata/policies/repositories>.

If a community resource is unavailable, data can be submitted to generalist repositories such as <https://figshare.com/> or <http://datadryad.org/> Dryad Digital Repository. Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at <http://www.nature.com/authors/policies/availability.html>.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

New particle formation and growth in the Arctic is important, being related to the production of particles that can act as cloud

condensation nuclei and thus affect climate. Moreover, the Arctic is changing rapidly with major changes in the hydrological cycle, including sea ice melting driven by climate warming amplification. It is important to understand the relative importance of biogenic vs. anthropogenic sources in this region.

I have a generally high opinion of this work and I think it should get published. New particle formation and growth in the Arctic have been previously observed by a number of groups but it is not known which chemicals promote nucleation and subsequent condensation. The paper builds upon observations from Mungall et al. who documented the release of VOCs from ocean water with riverine input. Multiphase chemistry may drive these emissions. The authors should add a new reference on this topic (Schneider et al., ACS Earth and Space Chemistry, 2024; 10.1021/acsearthspacechem.4c00163).

Mungall et al. hypothesized that these emissions may be related to particle growth but did not have direct evidence to support this claim. This paper attempts to test this hypothesis by analyzing time dependent particle size distributions from northern Russia in a statistical manner to determine which environmental factors, including exposure to waters with high DOM and riverine input, are associated with new particle formation and growth.

It is found that parameters such as temperature, boundary layer height, and precipitation are most strongly correlated. Temperature could drive VOC volatilization from the ocean. Precipitation is likely important because it drives the condensation sink down. (The authors should point this out, if they do not already.) Is the correlation with boundary layer height direct or inverse? In particular, it would make sense if it was inverse, i.e., emissions into a smaller boundary layer would drive up concentrations, leading to enhanced nucleation and growth rates. If the correlation is direct, then is it because higher temperatures have higher boundary layer heights??? The correlations with the DOM (from satellite measurements) in the near coastal regions, close to the observation site, are noticeably weaker but positive I assume.

Correlation analysis is not the same as causation analysis; this drives my only suggestions. First, why only look at particle size distribution data from this Russian site? To generalize the conclusions, why not use data from other Arctic sites? Arriving at the same correlations from these sites also would significantly strengthen the conclusions. Second, the DOM that is claimed to affect the particle growth and nucleation is necessarily close to shore. So, I was trying to think what else might correlate with coast environments, so that the DOM correlations are just a coincidence? Different local meteorological conditions are very likely associated with coastal environments. Could the authors re-do their analyses by disregarding their DOM measurements, and attempt to correlate with just "coastal air contact time" or "RH in the coastal region" or "Precipitation in the coastal region"? If the correlations turn out to be weaker than with satellite DOM instead, this would strengthen the overall conclusions of the paper.

Reviewer #2 (Remarks to the Author):

This manuscript by James Brean et al. combine ocean satellite data, air mass trajectories and meteorological conditions and find that the influx of riverine and sediment-derived terrigenous dissolved organic matter is key factors contributing to the frequent occurrence of NPF during the Arctic summer.

Although marine local sources seem to contribute significantly,

Arctic summer NPF, the interaction between the ocean and the atmosphere is not well understood, so a long-term dataset is valuable in providing scientific insight.

The approach from case studies to cluster analysis and machine learning, as well as the characteristics of NPF at each Pan-Arctic site and for each air mass classification, are also interesting.

My main concern is the assessment using tDOM and its certainty as an indicator.

1) The study discusses the difference between high and low tDOM. However, how much difference can be expected if tDOM is assumed not to be involved, likely typically processes (i.e. only on the atmospheric side and without water-mediated reactions) and the contribution from the ocean is added?

This question may be relevant to Extended Data Figures 6–7.

Could you provide more detail on this matter, as well as the constraints in lines 125–135 and the explanation in lines 214–221?

2) The contribution to CCN is mentioned in the introduction and the final sentence of the abstract. However, this study mainly focuses on the relationship and analysis of NPF and tDOM, and the analysis itself is not directly linked to CCN research. It would be better to rephrase the last sentence, 'Riverine input and subsequent NPF are the major source of Arctic CCN', i.e. 'Consideration of riverine input and subsequent NPF may contribute significantly for Arctic CCN research, especially in summer', etc.

3) Technical comments:

L105: 'On the 15th September 2021' -> 'September 2012'.

Reviewer #3 (Remarks to the Author):

See attached PDF file.

** Visit Nature Portfolio's author and referees' website at <http://www.nature.com/authors> for information about policies, services and author benefits**

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the [Platform Support Helpdesk](http://platformsupport.nature.com/).

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Dr Dall'Osto,

Your manuscript titled "Continental river runoff enhances atmospheric aerosol formation over the Arctic Ocean" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment.

We therefore invite you to revise your paper one last time to edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Portfolio, please visit <https://www.nature.com/commsenv/open-access>

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Alice Drinkwater, PhD
Associate Editor
Communications Earth & Environment
Consulting Editor
Communications Sustainability

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have nicely addressed by comments. The paper is ready to be published.

Reviewer #2 (Remarks to the Author):

Thank you for addressing my comments.
I think the manuscript has been appropriately revised, incorporating the reviewers' suggestions.

** Visit Nature Portfolio's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons

license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer's report on the manuscript by Brean et al. "Continental river runoff enhances atmospheric aerosol formation over the Arctic Ocean", Communications Earth & Environment, Manuscript ID: COMMSENV-25-1564-T

The manuscript presents an investigation into whether terrigenous dissolved organic matter (tDOM) in continental river runoff makes an important contribution to atmospheric aerosol formation in the Arctic, through analysis performed on particle number-size distribution data collected from Tiksi site in Russian Arctic. Although this study does not particularly shed light on the actual mechanism how (and what kind of) tDOM in river discharge to the Arctic ocean surface water contributes to atmospheric new particle formation (due to the lack of chemical information on the dissolved organic matter, the associated VOCs, and the composition of atmospheric aerosol particles formed), it does provide a strong indication that tDOM in river runoff can be an important source of marine organic aerosol precursors contributing to new particle formation (NPF) in summertime Arctic. I do have a few comments regarding some details of methodology as listed below. I hope that this study will lead to investigations to further characterise the role of tDOM in Arctic coastal surface water contributing to secondary aerosol formation in the Arctic.

General comments

1. The authors employed several approaches to analyse the nine-year Tiksi particle number-size distribution (PNSD) data to establish evidence for the impact of tDOM in river runoff on NPF in the Lena Delta area, including 1) an NPF case illustration, 2) manual classification of marine NPF events, 3) cluster analysis on the PNSD data, and 4) use of a machine learning tool to elucidate factors (including air mass exposure to high tDOM sea surface water) controlling the observed number concentrations of particles in the size range (15 – 100 nm) most relevant to NPF. While the analysis presented in 1) and 2) provided rather convincing evidence to support the author's hypothesis posed for the study (line 82 – 84), the cluster analysis did not seem to yield anything particularly useful to support the argument. The evidence from the machine learning exercise is also not particularly strong (see specific comments below).
2. The use of 24-hour back trajectories: given that the Arctic air mass has relatively long residence time within the Arctic particularly at low altitudes, e.g., average of 7 days in winter and 14 days in summer (see Stohl, 2006; Willis et al., 2018), longer back trajectories may be more appropriate for this analysis. Did the authors consider using longer back trajectories (e.g., 3- or 5-day trajectories)? How does the analysis depend on the length of trajectories?

Specific comments

- line 82: Tiksi, given its location, is probably not as likely in a CCN limited regime (as defined in Mauritsen et al., 2011) as other Arctic sites (such as Alert, Villum, and Ny-Ålesund).
- On the manual classification of marine NPF events, 24 out of the 169 summertime NPF events observed at Tiksi were deemed to be primarily marine NPF events and without significant influence of marginal sea ice. What is the breakdown for other categories out of the 169 NPF events (e.g., land influenced, marginal sea ice influenced)?

- Line 143-144: it is unclear how the different frequency of occurrence for NPF events linked to high- vs. low-tDOM air mass is determined. If the 24 events were split into 12 high-tDOM and 12 low-tDOM, one would expect same frequency of occurrence for both cases.
- Line 146-147: N_{100} would reflect both the initial particle formation stage and condensational growth stage. Did the author consider N_{50} for example, which may reflect NPF better?
- Line 154: Figure 3d seems to show mode diameter <30 nm for low-tDOM cases?
- Line 155-158: it's interesting to notice that the "bananas shape" seem to show up more clearly on the low-tDOM plots (Fig. 3e and extended data Fig. 4). There also seem to be an indication for particles stop growing beyond 30 nm in some of the high-tDOM cases (particularly noticeable on extended data Fig. 4).
- On cluster analysis, how were the 9 clusters generated from the k-means cluster analysis grouped into the 4 categories discussed in this section (which is not explained in the method section)? The marine NPF category does not distinguish between cases influenced by tDOM enriched sea surface water and cases influenced by marginal sea ice, for example. It is not clear from this cluster analysis how much the tDOM in the river runoff may have impacted the NPF events observed at Tiksi.
- Line 202-204: Figure 4b seems to indicate air temperature being the dominating predictor in the machine-learning model to predict N_{15-100} at Tiksi, followed by three other meteorological predictors (boundary layer height, surface pressure, and precipitation). The airmass tDOM exposure predictor falls within a group of other predictors, such as solar radiation, time over marginal ice zone, etc., with considerably lower SHAP values than the first four meteorological predictors.
- On machine learning exercise, did the authors consider withholding the airmass history predictors (e.g., exposure to tDOM and time spent over the marginal ice zone) in training to see if resulting machine-learning model had inferior predictability than the one including either or both airmass history predictors? This may provide stronger support for the hypothesis posed for this study.
- Line 226: this seems to be an awkward statement – if PNSD is given, one would not need a model to predict N_{15-100} .
- Line 233-234: it is not clear how the cluster analysis presented in this study yielded support for this statement (see my comment above regarding the cluster analysis).
- Figure 5: are the data points shown for other Arctic sites taken directly from Brean et al. (2023) or reevaluated for the present study based on the data included in the said previous study (e.g., J_{15} here vs. J_{10} shown in Brean et al., 2023)? Regarding the data points related to Tiksi, are the "Terrestrial" and "Sea ice" constitute the 169-24 cases (what is the breakdown)?
- Line 254-256: this is also implicated from some of the findings from existing studies, e.g., Mungall et al. (2017), where it was suggested that OVOC in sea surface microlayer observed in the Nare Strait (which contributed to marine organic particle formation) may be from the organic matter-laden river waters flown into the Strait from the Arctic basin.
- Line 295-296: Leaitch et al., (2016) also showed that particles smaller than 50 nm can act as CCN in summertime Arctic marine boundary layer.

- Line 303-304: I don't think the term "Arctic emission inventory" is an appropriate term to use in this context. You could simply say that continental runoff may be an important source for marine organic aerosols in the Arctic which have not been accounted for (or represented) in current atmospheric models.
- Line 313-315: this maybe somewhat an overstatement. Not sure if "aerosol vertical flux" is an appropriate term to use to describe NPF process.

References

Brean, J., Beddows, D. C. S., Harrison, R. M., Song, C., Tunved, P., Ström, J., Krejci, R., Freud, E., Massling, A., Skov, H., Asmi, E., Lupi, A., and Dall'Osto, M.: Collective geographical ecoregions and precursor sources driving Arctic new particle formation, *Atmos Chem Phys*, 23, 2183–2198, <https://doi.org/10.5194/acp-23-2183-2023>, 2023.

Leaitch, W. R., Korolev, A., Aliabadi, A. A., Burkart, J., Willis, M. D., Abbatt, J. P. D., Bozem, H., Hoor, P., Köllner, F., Schneider, J., Herber, A., Konrad, C., and Brauner, R.: Effects of 20–100 nm particles on liquid clouds in the clean summertime Arctic, *Atmos Chem Phys*, 16, 11107–11124, <https://doi.org/10.5194/acp-16-11107-2016>, 2016.

Mauritsen, T., Sedlar, J., Tjernström, M., Leck, C., Martin, M., Shupe, M., Sjogren, S., Sierau, B., Persson, P. O. G., Brooks, I. M., and Swietlicki, E.: An Arctic CCN-limited cloud-aerosol regime, *Atmos Chem Phys*, 11, 165–173, <https://doi.org/10.5194/acp-11-165-2011>, 2011.

Mungall, E. L., Abbatt, J. P. D., Wentzell, J. J. B., Lee, A. K. Y., Thomas, J. L., Blais, M., Gosselin, M., Miller, L. A., Papakyriakou, T., Willis, M. D., and Liggio, J.: Microlayer source of oxygenated volatile organic compounds in the summertime marine Arctic boundary layer, *Proc. Natl. Acad. Sci.*, 114, 6203–6208, <https://doi.org/10.1073/pnas.1620571114>, 2017.

Stohl, A.: Characteristics of atmospheric transport into the Arctic troposphere, *J. Geophys. Res. Atmospheres*, 111, <https://doi.org/10.1029/2005JD006888>, 2006.

Willis, M. D., Leaitch, W. R., and Abbatt, J. P. D.: Processes Controlling the Composition and Abundance of Arctic Aerosol, *Rev. Geophys.*, 56, 621–671, <https://doi.org/10.1029/2018RG000602>, 2018.

Response to Reviewers

We thank both reviewers for their careful reading of our manuscript and for their constructive suggestions. Below, review comments are in black, and we provide our point-by-point responses in blue. Sections from the revised manuscript are in green.

General changes

All references to “extended data” figures have been changed to supplementary figures. E.g., “Extended data figure 2” has become “Figure S2”.

Reviewer 1

I have a generally high opinion of this work and I think it should get published. New particle formation and growth in the Arctic have been previously observed by a number of groups but it is not known which chemicals promote nucleation and subsequent condensation. The paper builds upon observations from Mungall et al. who documented the release of VOCs from ocean water with riverine input. Multiphase chemistry may drive these emissions. The authors should add a new reference on this topic (Schneider et al., ACS Earth and Space Chemistry, 2024; 10.1021/acsearthspacechem.4c00163).

Thank you for the kind comments. We agree that Scheider et al. is a great paper and we incorporate it as follows (Schneider et al. is our *reference 12*). We also realise upon reading Schneider et al. that we have omitted some important references, and include refs 52 and 53 from Schneider et al. as our refs 35 and 36. New text is in bold below.

“Mungall et al. (2017)¹⁰ reported a microlayer source of oxygenated volatile organic compounds in the summertime marine Arctic boundary layer, possibly attributed to the surface waters carrying terrigenous dissolved organic matter (tDOM) to Nares Strait in the Canadian Archipelago. Abiotic processing of the sea-surface microlayer—including photochemical or heterogeneous oxidation reactions^{11,12}—is therefore a potential source of vapours which drive NPF events. This source may be very different from those in pelagic and sympagic marine or terrestrial settings.”

“A comprehensive understanding of Arctic climate change necessitates an appreciation of all NPF precursor sources. In sunlit ocean waters, the strong absorption of high-energy UV radiation by chromophoric dissolved organic matter leads to the photochemical oxidation and degradation of DOM³⁴, generating secondary organic precursors **and small oxygenated organics**^{35,36}. Experiments simulating this, **alongside direct ozone exposure**, produce large fluxes of organic compounds, especially nonanal^{11,12} which is known to undergo autoxidation and produce highly oxygenated molecules³⁷. **Crucially, a study using Arctic seawater show that the total flux of these VOCs correlates strongly with total organic carbon**¹². **This laboratory finding is consistent with field data from Mungall et al., who also observed a strong correlation between a marine source of oxygenated-VOCs and the concentration of dissolved organic carbon (DOC)**¹⁰.”

Mungall et al. hypothesized that these emissions may be related to particle growth but did not have direct evidence to support this claim. This paper attempts to test this hypothesis by analyzing time dependent particle size distributions from northern Russia in a statistical manner to determine which environmental factors, including exposure to waters with high DOM and riverine input, are associated with new particle formation and growth.

It is found that parameters such as temperature, boundary layer height, and precipitation are most strongly correlated. Temperature could drive VOC volatilization from the ocean. Precipitation is likely important because it drives the condensation sink down. (The authors should point this out, if they do not already.) Is the correlation with boundary layer height direct or inverse? In particular, it would make sense if it was inverse, i.e., emissions into a smaller boundary layer would drive up concentrations, leading to enhanced nucleation and growth rates. If the correlation is direct, then is it because higher temperatures have higher boundary layer heights??? The correlations with the DOM (from satellite measurements) in the near coastal regions, close to the observation site, are noticeably weaker but positive I assume.

Thank you again. We provide the SHAP values for each of the variables in Figure S11. Temperature indeed likely drives volatilisation, and also would promote autoxidation if these molecules are indeed organic and undergo autoxidation. We agree that precipitation drives down the condensation sink, but the SHAP values didn't produce a clear picture of this relationship. The relationship with boundary layer is inverse: a lower boundary layer leads to higher predicted N_{15-100} and visa versa. We attribute this therefore to dilution of larger particles. We address these comments by incorporating the following lines (new text in bold):

“A positive relationship is also seen for temperature (Figure 4e), **likely reflecting its dual role in driving the volatilisation of VOCs from the ocean surface²⁸ and in promoting the autoxidation rates of organic precursors²⁹.** The model also finds an inverse relationship with boundary layer height, where shallower boundary layers strongly predict higher N_{15-100} (Fig. S11). This is consistent with the physical mechanism of dilution; precursor vapours are concentrated within a smaller volume, enhancing particle formation and growth. The relationship with precipitation is not clear, with an initial small positive influence, likely because wet scavenging reduces the condensation sink by removing larger aerosol particles, which creates more favourable conditions for NPF, but a decrease in predicted N_{15-100} at high precipitation volumes. The exact effect of precipitation is likely not clear from the model, as particle production typically follows precipitation, rather than being concurrent with it. Similarly, solar radiation influences oxidation processes by driving hydroxyl radical production (Figure 4f), while wind speed impacts the rate of precursor emission from the sea surface microlayer²⁸ (Fig. S11). “

Correlation analysis is not the same as causation analysis; this drives my only suggestions. First, why only look at particle size distribution data from this Russian site? To generalize the conclusions, why not use data from other Arctic sites? Arriving at the same correlations from these sites also would significantly strengthen the conclusions.

We thank the reviewer for this important point. In this work we set out to investigate the influence of tDOM from major river systems. The selection of the Tiksi station was a deliberate choice. Its location makes it uniquely exposed to airmasses passing over the tDOM-rich shelf seas fed by Siberian rivers. Conversely, other long-term Arctic monitoring stations (e.g., in Svalbard, Alaska, Greenland) are situated in environments largely disconnected from these major terrestrial inputs.

While an analysis including those sites would be valuable for studying other NPF drivers. Our single-site focus is therefore necessary to understand the role of tDOM. We agree that a future pan-Arctic analysis would be a valuable follow-up study. We emphasise the uniqueness of Tiksi now in the following sentence:

“**The atmospheric monitoring site at Tiksi is positioned within the vast Eurasian Arctic bioregion that receives the majority of the Arctic's river discharge (Figure 1). Its location near the Lena River delta, the eleventh-longest river globally and the longest entirely**

within Russia, makes it uniquely representative for studying atmospheric processes influenced by these freshwater inputs”

Second, the DOM that is claimed to affect the particle growth and nucleation is necessarily close to shore. So, I was trying to think what else might correlate with coast environments, so that the DOM correlations are just a coincidence? Different local meteorological conditions are very likely associated with coastal environments. Could the authors re-do their analyses by disregarding their DOM measurements, and attempt to correlate with just “coastal air contact time” or “RH in the coastal region” or “Precipitation in the coastal region”? If the correlations turn out to be weaker than with satellite DOM instead, this would strengthen the overall conclusions of the paper.

Thank you for the comment. Firstly, we must emphasise that the elevated tDOM concentrations in the Eurasian Arctic extend >100 kilometres from the coastline (e.g. Figure 1), indicating a broad chemical footprint that extends far beyond the coastline across a large swathe of the Eurasian Arctic ocean.

We performed some extra analyses to investigate. Firstly, we calculated the distance of every trajectory datapoint to the nearest coastline (*Figure R1*). After removing every trajectory that spent >33% of it's time over land, we plotted the mean ag320 for each trajectory against the distance to the coast, and found no negative relationship, although we do note that high ag320 values never occurred when the trajectories spent most of their time far from the coastline (*Figure R2*).

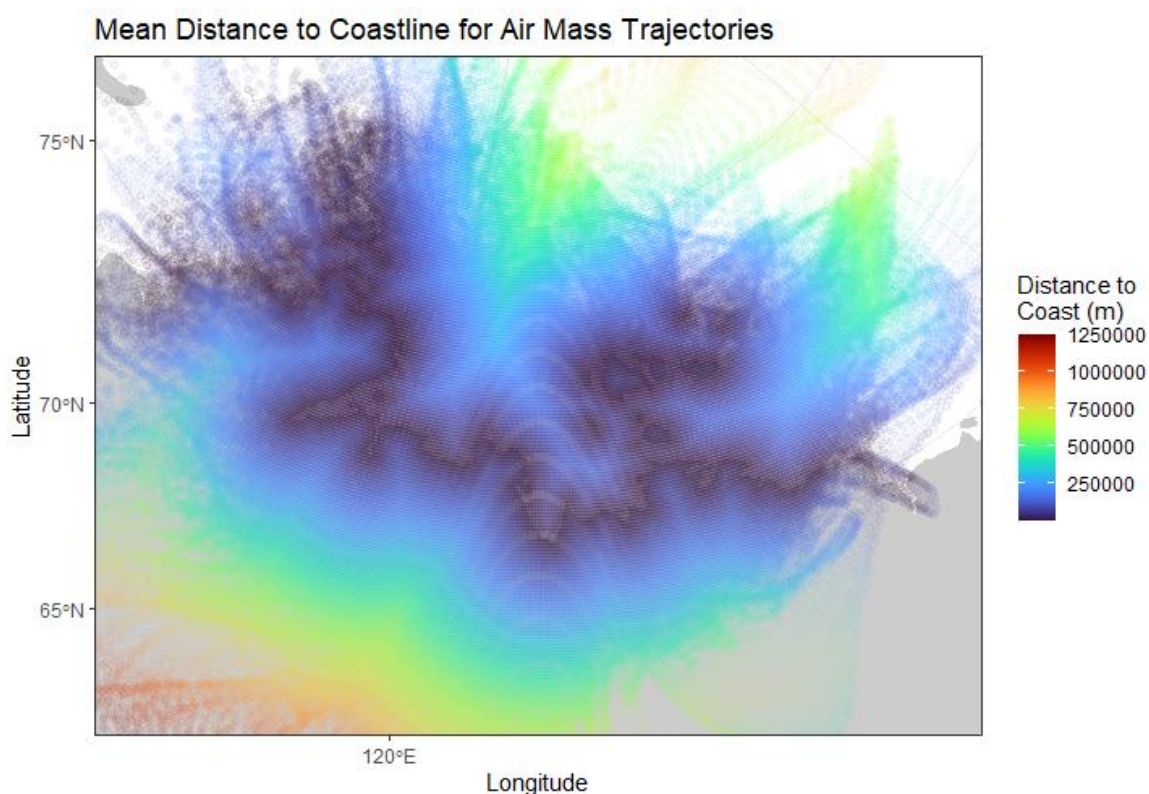


Figure R1: Distance of each back trajectory datapoint from the nearest coastline, inc. terrestrial airmasses.

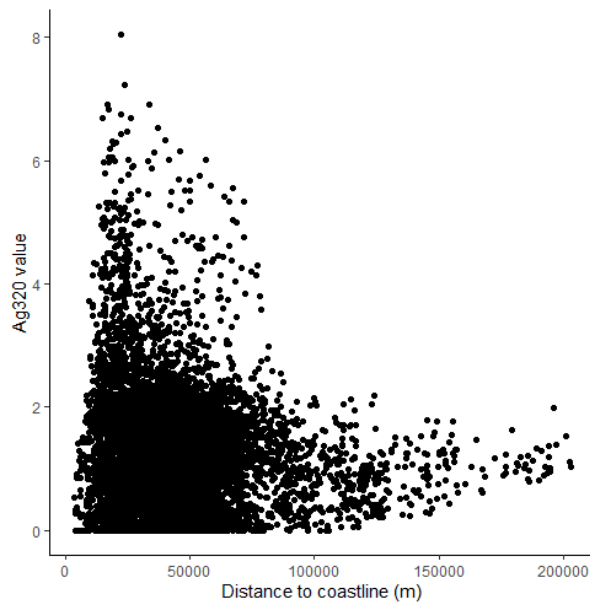


Figure R2: The airmass exposure to AG320 (a.k.a. tDOM) versus it's distance from the coastline for every trajectory that spent >33% of it's time over the ocean.

Subsequently, we built five random forest models (*Figure R3*).

- “Baseline_paper” is the model as in the paper
- “Baseline_plus_distance” is the model in the paper with the “distance from coast” added
- “Plus_distance_no_AG_ice” is the model in the paper with “distance from coast” added and exposure to tDOM (a.k.a. AG320) and the marginal ice zone removed.
- “Baseline_no_AG_ice” is the model in the paper with exposure to tDOM (a.k.a. AG320) and the marginal ice zone removed.

We constructed and ran each model 100 times to account for the slight differences in performance for different starting points (random forests utilise bootstrapping and therefore there is small run-to-run variation) and present the Root Mean Squared Error (RMSE) for each run below (*Figure R3*). RMSE quantifies the average magnitude of the prediction error in the same units as the target variable and is therefore an interpretable measure of the model's accuracy. The model run in the paper had an RMSE of 66.6 cm^{-3} . Adding in the airmass exposure to the coastline reduces the RMSE negligibly ($\sim 1 \text{ cm}^{-3}$, from 66.6 cm^{-3} to 65.2 cm^{-3}), while *removing* the exposure to tDOM and the MIZ increases the model's RMSE by 7 cm^{-3} , from 66.6 cm^{-3} to 73.8 cm^{-3} . Re-adding the airmass exposure to the coastline then reduces the RMSE by $\sim 2 \text{ cm}^{-3}$.

This confirms that exposure to tDOM and the MIZ are the key predictors. Adding coastal proximity had a small effect on model performance (reducing RMSE by $\sim 2\%$), whereas removing tDOM and MIZ exposure significantly weakened it. This confirms the minimal influence of the coastline compared to the primary importance of tDOM and MIZ.

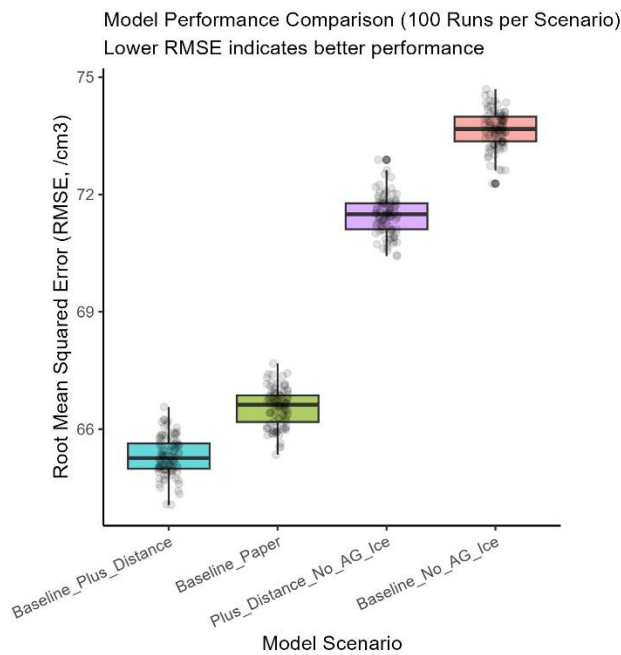


Figure R3: Four different random forest runs. “Baseline_paper” is the model as in the paper; “Baseline_plus_distance” is the model in the paper with the “distance from coast” added; “Plus_distance_no_AG_ice” is the model in the paper with “distance from coast” added and exposure to tDOM (a.k.a. AG320) and the marginal ice zone removed.; “Baseline_no_AG_ice” is the model in the paper with exposure to tDOM (a.k.a. AG320) and the marginal ice zone removed.

We include the following new paragraph in our text:

“To validate our choice of predictors, we performed an ablation study. First, we tested the model without the air mass history variables. Removing the exposure to tDOM and the marginal ice zone weakened the model, increasing the Root Mean Squared Error (RMSE) from a baseline of 66.6 cm⁻³ to 73.8 cm⁻³. This impact is comparable to removing temperature, a known key driver, confirming the predictive power of air mass history. Conversely, we also investigated the potential influence of terrestrial sources by adding air mass distance from the nearest coastline as a predictor. This yielded only a negligible improvement, reducing the RMSE by just ~2%. These tests confirm that air mass history over tDOM and the marginal ice zone are the key predictors for N₁₅₋₁₀₀ in our model and are not a proxy for proximity to coastlines.”

Reviewer 2

This manuscript by James Brean et al. combine ocean satellite data, air mass trajectories and meteorological conditions and find that the influx of riverine and sediment-derived terrigenous dissolved organic matter is key factors contributing to the frequent occurrence of NPF during the Arctic summer.

Although marine local sources seem to contribute significantly,

Arctic summer NPF, the interaction between the ocean and the atmosphere is not well understood, so a long-term dataset is valuable in providing scientific insight.

The approach from case studies to cluster analysis and machine learning, as well as the characteristics of NPF at each Pan-Arctic site and for each air mass classification, are also interesting.

My main concern is the assessment using tDOM and its certainty as an indicator.

1) The study discusses the difference between high and low tDOM. However, how much difference can be expected if tDOM is assumed not to be involved, likely typically processes (i.e. only on the atmospheric side and without water-mediated reactions) and the contribution from the ocean is added?

This question may be relevant to Extended Data Figures 6–7.

Could you provide more detail on this matter, as well as the constraints in lines 125–135 and the explanation in lines 214–221?

The reviewer asks how much difference can be expected in NPF if tDOM is assumed not to be involved. Given the nature of atmospheric observation, this is not possible. Due to the location of Tiksi there is little time when the adjacent ocean is extremely low in tDOM. The closest we can provide is a quantitative comparison of NPF against a baseline scenario where tDOM is minimally involved. This comparison is the central focus of our manual classification analysis, presented in Figure 3. We have now clarified in the text that we consider the airmasses exposed to low-tDOM to represent a baseline for typical open-ocean NPF as follows (new text in bold):

“Next, these 24 events were split into two categories of 12 NPF events based on the concentration of tDOM in surface waters (Figure 3a):

- 12 low-tDOM events, with air masses encountering relatively low tDOM concentrations (mean $a_g(320) < 1.25 \text{ m}^{-1}$)
- 12 high-tDOM events, where air masses were exposed to relatively high tDOM concentrations (mean $a_g(320) > 1.25 \text{ m}^{-1}$)

Both had similarly low black carbon concentrations, **indicating pristine Arctic conditions**. These high tDOM airmasses were exposed to regions of lower salinity, supporting that these regions are influenced by river discharge flooding over the Arctic Ocean (Figure 3b). **Consequently, the low-tDOM events are treated as a baseline representing more open-ocean conditions, against which the influence of riverine tDOM is compared**. Over the nine-year record, NPF event frequency was 6.6% for low-tDOM air masses compared to 10.8% for high-tDOM air masses.”

Our results show that, relative to this baseline, high-tDOM airmasses lead to a 300% enhancement in the particle formation rate (J_{15}) and a 60% enhancement in the growth rate (GR). This enhancement occurs despite both high- and low-tDOM event categories experiencing similar levels of solar radiation, strengthening the conclusion that tDOM-derived precursors, rather than variations in photochemistry, are driving the observed increase in NPF intensity.

To further quantify the importance of tDOM, we tested our random forest (RF) model. We constructed four types of model. The “full_model” is the model presented in our paper (Figure 4). We also present a model without the air mass history (tDOM and MIZ exposure) “no_tDOM_or_ice”, a model without temperature “no_temp”, and a model without air mass history nor temperature “no_AG_Ice_or_temp”. We constructed and ran each model 100 times to account for the slight differences in performance for different starting points (random forests

utilise bootstrapping and therefore there is small run-to-run variation), and present the Root Mean Squared Error (RMSE) for each run below (Figure R4).

The model in the paper (Figure 4) has a Root Mean Squared Error of 65.2 cm^{-3} . Removing the tDOM and MIZ variables worsens the model's predictive power, increasing the RMSE to 71.4 cm^{-3} . For comparison, removing temperature increases the RMSE equivalently to 71.1 cm^{-3} . Removing tDOM, MIZ, and temperature variables then reduces the model's accuracy to 79.1 cm^{-3} . RMSE quantifies the average magnitude of the prediction error in the same units as the target variable, and is therefore an interpretable measure of the model's accuracy

These results are fully consistent: the manual classification demonstrates that tDOM has an amplifying effect on NPF events when they occur. The RF model confirms that tDOM exposure is a statistically significant predictor of particle concentrations, in conjunction with meteorological factors like temperature.

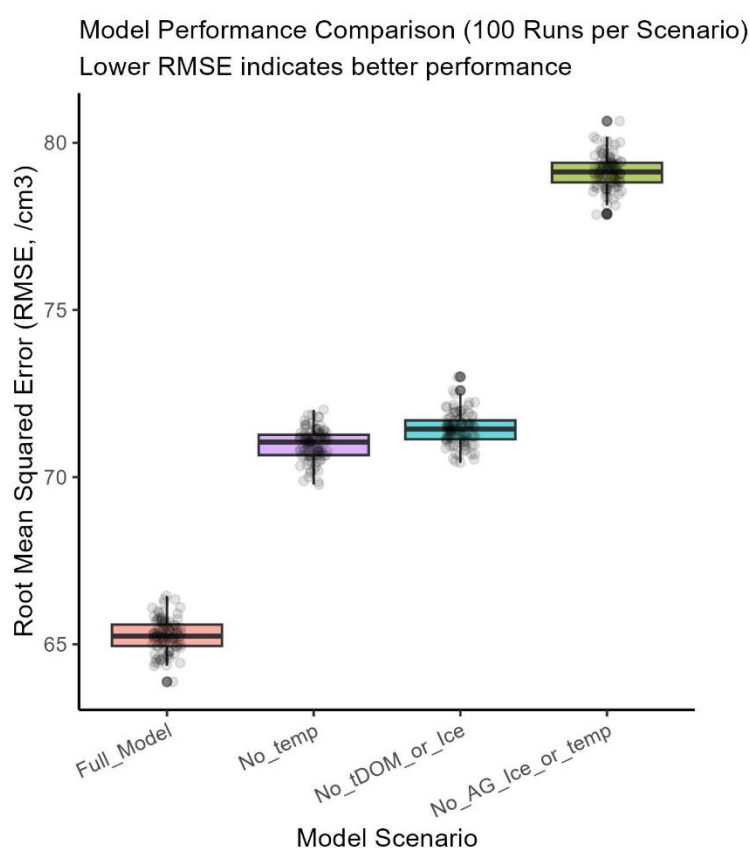


Figure R4: Four different random forest runs. “Full_model” is the model as in the paper; “No temp” is the model with temperature removed; “No_tDOM_or_Ice” is the model with air mass exposure to tDOM and the marginal ice zone removed; “No_AG_Ice_or_temp” is the model with exposure to tDOM and the marginal ice zone, as well as temperature removed.

2) The contribution to CCN is mentioned in the introduction and the final sentence of the abstract. However, this study mainly focuses on the relationship and analysis of NPF and tDOM, and the analysis itself is not directly linked to CCN research.

It would be better to rephrase the last sentence, 'Riverine input and subsequent NPF are the major source of Arctic CCN', i.e. 'Consideration of riverine input and subsequent NPF may contribute significantly for Arctic CCN research, especially in summer', etc.

Very good point! We amend the final sentence of our abstract as follows:

“Our findings indicate that riverine input and subsequent NPF is a major source of Arctic aerosol which has been largely overlooked.”

Although we do maintain that many of these particles once they reach ~50-100 nm will be CCN active.

3) Technical comments:

L105: 'On the 15th September 2021' -> 'September 2012'.

Thank you! This has been amended

Reviewer 3

General comments

The authors employed several approaches to analyse the nine-year Tiksi particle number-size distribution (PNSD) data to establish evidence for the impact of tDOM in river runoff on NPF in the Lena Delta area, including 1) an NPF case illustration, 2) manual classification of marine NPF events, 3) cluster analysis on the PNSD data, and 4) use of a machine learning tool to elucidate factors (including air mass exposure to high tDOM sea surface water) controlling the observed number concentrations of particles in the size range (15 – 100 nm) most relevant to NPF. While the analysis presented in 1) and 2) provided rather convincing evidence to support the author's hypothesis posed for the study (line 82 – 84), the cluster analysis did not seem to yield anything particularly useful to support the argument. The evidence from the machine learning exercise is also not particularly strong (see specific comments below).

Thank you for your constructive feedback. We agree that the manual classification provides the core evidence for our hypothesis. We have included the cluster analysis not as a primary line of evidence, but as a crucial independent and objective validation of our findings. Whereas our manual classification relies on pre-defined thresholds (e.g., for time over land or ice), the k-means clustering is an unsupervised method. It identifies inherent structures within the nine-year dataset without any prior assumptions about what constitutes an 'NPF event type'. The fact that this purely statistical approach independently identified a 'marine NPF' cluster with the exact same key characteristics as our manually selected high-tDOM events (i.e., the smallest particle mode (Figure S5), the highest exposure to tDOM (Figure S6), and the lowest salinity (Figure S7)) is a significant result. This convergence between two distinct methodologies, one rule-based and one data-driven, demonstrates that the association between tDOM and NPF is a robust feature of the data, not simply an artefact of our manual selection criteria. It substantially reduces the possibility of confirmation bias and strengthens the overall conclusion.

We acknowledge that the manuscript could state this role more explicitly. We will revise the text to clarify that the cluster analysis serves as an objective validation of the manual classification, thereby enhancing the scientific merit of our approach. We add these arguments to our cluster analysis section as follows:

To provide an independent and objective test of our hypothesis, we further classified our PNSD through k-means cluster analysis to identify different types of Arctic PNSDs in our data. **Unlike the manual classification, which relies on pre-defined thresholds, k-means is an unsupervised method that identifies inherent structures within the dataset without prior assumptions.** The nine K-means clusters obtained (see methods) are summarized in four main categories (Fig. S5): two categories attributable to NPF (one marine and one from land), one associated with pristine bimodal PNSDs, and a last cluster dominated by a coarse mode

of particles. The marine NPF category accounted for 21% of the PNSDs measured over the nine years, with peaks during summer months of June-October (37-45%); they are characterised by low black carbon levels, and occur under air masses that spend much of their time over water enriched in tDOM (Fig. S6).

This marine NPF category is also exposed to the lowest salinity, consistent with our manually assigned NPF events, and the highest concentrations of marginal sea ice zones composed of sea ice flows (Fig. S7). **This convergence between a rule-based selection and an unsupervised statistical method provides robust support for our conclusions, demonstrating that the link between tDOM and NPF is a feature of the data, not an artefact of our selection criteria.** Whilst we cannot apportion the relative importance of melting sea ice and riverine influenced marine regions on NPF, our K-means clustering point to an association between both regions and NPF activity.

The use of 24-hour back trajectories: given that the Arctic air mass has relatively long residence time within the Arctic particularly at low altitudes, e.g., average of 7 days in winter and 14 days in summer (see Stohl, 2006; Willis et al., 2018), longer back trajectories may be more appropriate for this analysis. Did the authors consider using longer back trajectories (e.g., 3- or 5-day trajectories)? How does the analysis depend on the length of trajectories?

Thank you for this insightful comment. While we agree that Arctic air masses have long residence times, our choice of a 24-hour back trajectory is dictated by the much shorter atmospheric lifetimes of the key gaseous precursors that drive NPF

The relevant precursors, such as dimethyl sulfide (DMS), SO₂ reactive biogenic organics, and iodine compounds, have lifetimes of only hours to, at most, a day or two in the Arctic atmosphere (e.g., (Carpenter and Nightingale, 2015; Kirkby et al., 2023) Once these gases are oxidised to form the condensable vapours responsible for NPF (methanesulfonic acid, sulfuric acid, iodic oxoacids, and OOMs, or HOMs), their lifetime before being lost to the aerosol phase is on the order of minutes with minimal re-evaporation, except in the case of MSA, where there is likely some re-evaporation. Ammonia and amines also have short lifetimes in the gas phase, and do not need to undergo oxidation before they partition into the particle phase, although ammonia will likely re-evaporate into the gas phase.

Consequently, any emissions from 3-5 days upwind would have already undergone chemical transformation and contributed to the pre-existing aerosol population, rather than to the new particle formation event being observed. A 24-hour trajectory is therefore the most appropriate length to identify the recent emission sources directly responsible for the NPF. We now include this argument in the following sentences in the manuscript (new text in bold):

“Each hour of measurement was linked to a 24-hour back trajectory calculated using HYSPLIT¹⁹. **This duration was chosen specifically because the key gaseous precursors for NPF (e.g., dimethylsulfide, reactive organics, iodine compounds) have short atmospheric lifetimes of hours to a day or two, making recent air mass history the most relevant for identifying emission sources**^{20,21}.”

Below we include the impact of extending trajectory lengths. We extend from 24 to 48 hours (Figure R5). The GRs and particle number counts in both cases are the same (greater for high tDOM), but the formation rates are the same for both cases. This is mostly from some NPF events falling from the “low-tDOM” category into the “high-tDOM category” and visa versa. We argue that 24 hours is, following our above arguments, the better trajectory length to use.

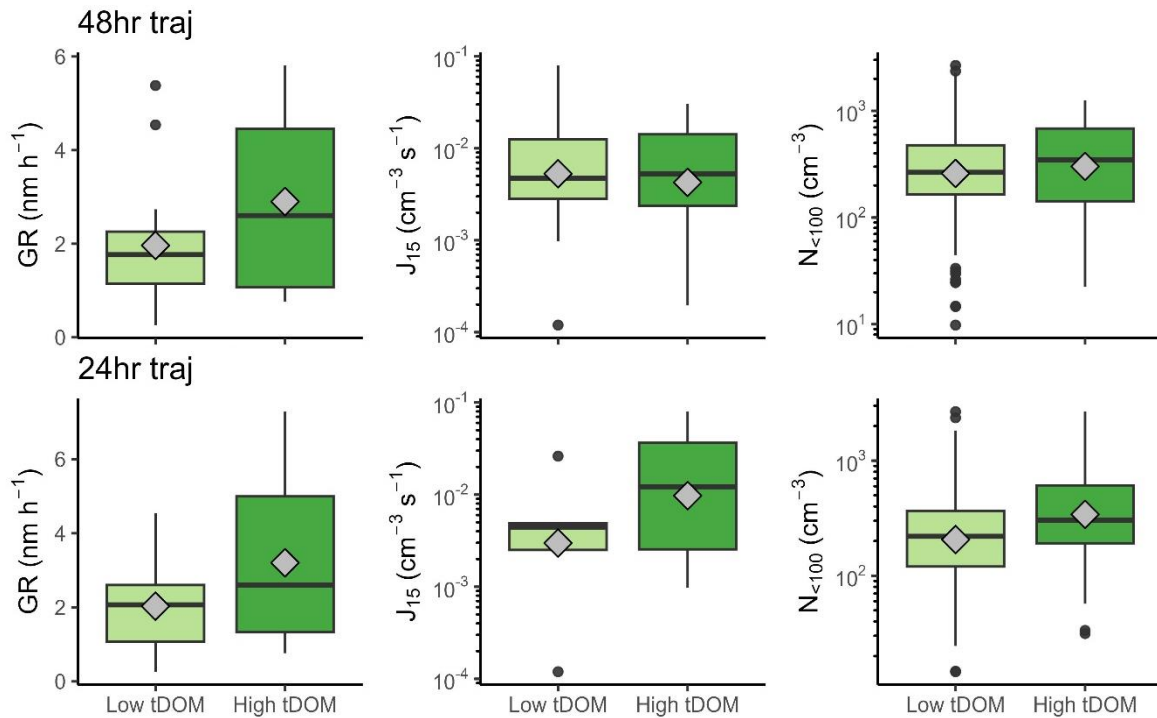


Figure R5: Growth rates, formation rates, and N_{100} as Figure 3, calculated for both 48 hour and 24 hour back trajectories.

Specific comments

line 82: Tiksi, given its location, is probably not as likely in a CCN limited regime (as defined in Mauritsen et al., 2011) as other Arctic sites (such as Alert, Villum, and Ny-Ålesund).

This is indeed hard to prove, however, we include some speculation in our supplementary data. We show in Figure S8 that when Tiksi experiences marine air masses the concentration of particles >50 nm (which we can regard as a proxy for CCN) is below 10 cm^{-3} 10% of the time.

The critical diameter is likely larger than 50 nm, and the concentration of particles is likely lower at higher altitudes. Therefore the counts of CCN are likely lower than the number of >50 nm particles, and we defend this statement.

On the manual classification of marine NPF events, 24 out of the 169 summertime NPF events observed at Tiksi were deemed to be primarily marine NPF events and without significant influence of marginal sea ice. What is the breakdown for other categories out of the 169 NPF events (e.g., land influenced, marginal sea ice influenced)?

Of the remaining NPF events, a fraction are sea ice influenced (3%), while the remainder are terrestrially influenced, as the MIZ is typically a great distance from Tiksi in the summertime. We include the following lines in our manuscript

“Of the excluded NPF events, only a few (3%) were excluded due to sea-ice influence, and so the majority of measured NPF events were terrestrially dominated.”

Line 143-144: it is unclear how the different frequency of occurrence for NPF events linked to high- vs. low-tDOM air mass is determined. If the 24 events were split into 12 high-tDOM and 12 low-tDOM, one would expect same frequency of occurrence for both cases.

Great point! This is poorly worded in the manuscript. We have amended our wording throughout this section as follows (new or edited text in bold, and mention of NPF frequency at the end):

“Extending our analysis, we identified 169 NPF events during the nine-year record of hourly PNSDs at Tiksi¹⁵. These NPF events were identified between June 1st through September 30th for each year, which roughly corresponds to the summer period following the Arctic haze⁹ and preceding the full ocean freeze-up near Tiksi (Fig. S2).

We applied a strict filtering protocol using 24-hour back trajectories for each hour of data. Since the marginal ice zone is a known source of NPF precursors^{22,23}, and the land is a source of both anthropogenic and biogenic NPF precursors²⁴. We discarded NPF events where the 24-hour back trajectory spent more than 40% of its time over land, minimising BC while retaining data coverage (Fig. S3). Events which spent more than 66% over marginal sea ice were discarded, balancing the need to limit marginal ice zone emissions while preserving enough open-ocean cases for analysis.

We retained 24 events dominated by open ocean conditions. The retained marine air masses contained, on average, 2.5 times less black carbon than the discarded terrestrial air masses (Fig. S3) and were therefore less influenced by anthropogenic and terrestrial emissions. **Of the excluded NPF events, only a few (3%) were excluded due to sea-ice influence, and so the majority of measured NPF events were terrestrially dominated.**

Next, these 24 events were split into two categories of 12 NPF events based on the concentration of tDOM in surface waters (Figure 3a):

- 12 low-tDOM events, with air masses encountering relatively low tDOM concentrations (mean $a_g(320) < 1.25 \text{ m}^{-1}$)
- 12 high-tDOM events, where air masses were exposed to relatively high tDOM concentrations (mean $a_g(320) > 1.25 \text{ m}^{-1}$).

Both had similarly low black carbon concentrations, **indicating pristine Arctic conditions.** These high tDOM airmasses were exposed to regions of lower salinity, supporting that these regions are influenced by river discharge flooding over the Arctic Ocean (Figure 3b). **Consequently, the low-tDOM events are treated as a baseline representing more open-ocean conditions, against which the influence of riverine tDOM is compared.**

In summertime marine PNSDs, secondary aerosol formation dominates the number concentration of particles smaller than 100 nm²⁵. In this dataset, we utilise the sum of particles between 15 and 100 nm, denoted as N_{15-100} . While primary sources such as sea spray and dust significantly contribute to the particle mass and larger bins of the PNSD, their impact on the sub-100 nm fraction is marginal^{2,25}. High N_{15-100} concentrations are therefore indicative of strong contribution of NPF.

Exposure of air masses to high-tDOM surface waters is associated with an increase of 60% in N_{15-100} (Figure 3c) relative to low-tDOM regions on NPF days. The mean PNSD on NPF days shows a mode of new particles centred at about 25 nm for the high-tDOM regions, whereas it is slightly larger (35 nm) for the low-tDOM (Figure 3d). The daily diurnal plots for the low-tDOM shows overall weaker NPF than the high tDOM regions (Figure 3e,f), with the higher mean diameter in the low-tDOM being skewed by a single strong event (concentration normalised contour plots in Fig. S4).

Finally, the particle formation rate at 15 nm is enhanced by 300% as a mean for the high-tDOM relative to the low-tDOM events, while the particle growth rate is enhanced by 60% as a mean (Figure 3g) despite slightly reduced solar radiation intensity for the high-tDOM events (Figure

3h). To place these findings in a broader context, we then calculated the overall NPF frequency across the entire nine-year record. Over this period, the NPF event frequency was 6.6% for all low-tDOM airmasses compared to 10.8% for all high-tDOM airmasses. Our manual classification data overall support the case study reported in Figure 2 suggesting that river runoff significantly enhances NPF activity in this Arctic region.”

Line 146-147: N_{100} would reflect both the initial particle formation stage and condensational growth stage. Did the author consider N_{50} for example, which may reflect NPF better?

We did indeed consider N_{50} . The results in this case are very similar, and we have included them both below. We argue that N_{100} better captures the total secondary particle population than N_{50} , and we opt to stick with N_{100} for the sake of consistency with our later analyses.

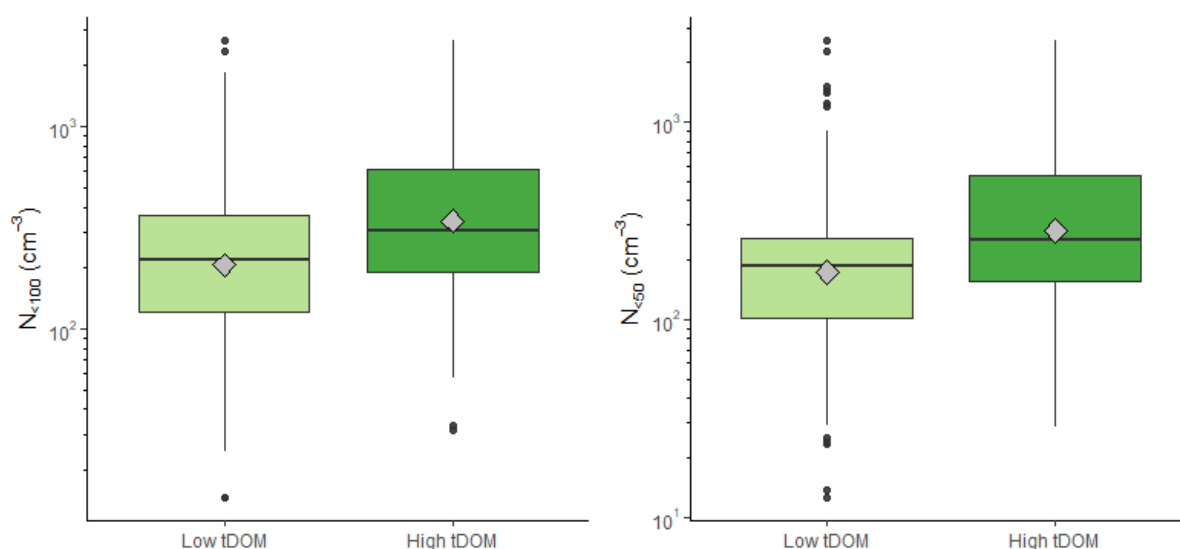


Figure R6: As Figure 3c, for both N_{100} and N_{50}

Line 154: Figure 3d seems to show mode diameter <30 nm for low-tDOM cases?

Right you are. We have amended this to say “<30 nm”.

Line 155-158: it’s interesting to notice that the “bananas shape” seem to show up more clearly on the low-tDOM plots (Fig. 3e and extended data Fig. 4). There also seem to be an indication for particles stop growing beyond 30 nm in some of the high-tDOM cases (particularly noticeable on extended data Fig. 4).

Indeed this is the case, and this is the problem with averaging bananas. The disappearance of the banana typically has more to do with the moving of airmasses than it does with the cessation of NPF, as the latter would be seen as a coagulation effect with a gradual reduction of particle count.

On cluster analysis, how were the 9 clusters generated from the k-means cluster analysis grouped into the 4 categories discussed in this section (which is not explained in the method section)? The marine NPF category does not distinguish between cases influenced by tDOM enriched sea surface water and cases influenced by marginal sea ice, for example. It is not clear from this cluster analysis how much the tDOM in the river runoff may have impacted the NPF events observed at Tiksi.

This two-step methodology is a standard and robust practice for analysing PNSD data, as it combines objective clustering with physical interpretation (Beddows et al., 2009; Brean et al., 2025; Dall'Osto et al., 2017; Dall'Osto et al., 2017).

- First, we chose a relatively high number of initial clusters (nine) to allow the k-means algorithm to capture a wide range of variations within the dataset without prematurely forcing different patterns into one group.
- Second, we manually grouped these nine mathematical clusters into four physically meaningful categories based on their aerosol dynamic characteristics. For example, several of the initial clusters, despite minor differences in mode diameter or concentration, all shared the defining characteristic of a strong nucleation mode, a clear diurnal growth pattern, and occurrence in the summertime. These were logically grouped into the 'Marine NPF' category.

We include examples of the pre-grouped k-means cluster analysis data below (*Figure R7, R8*):

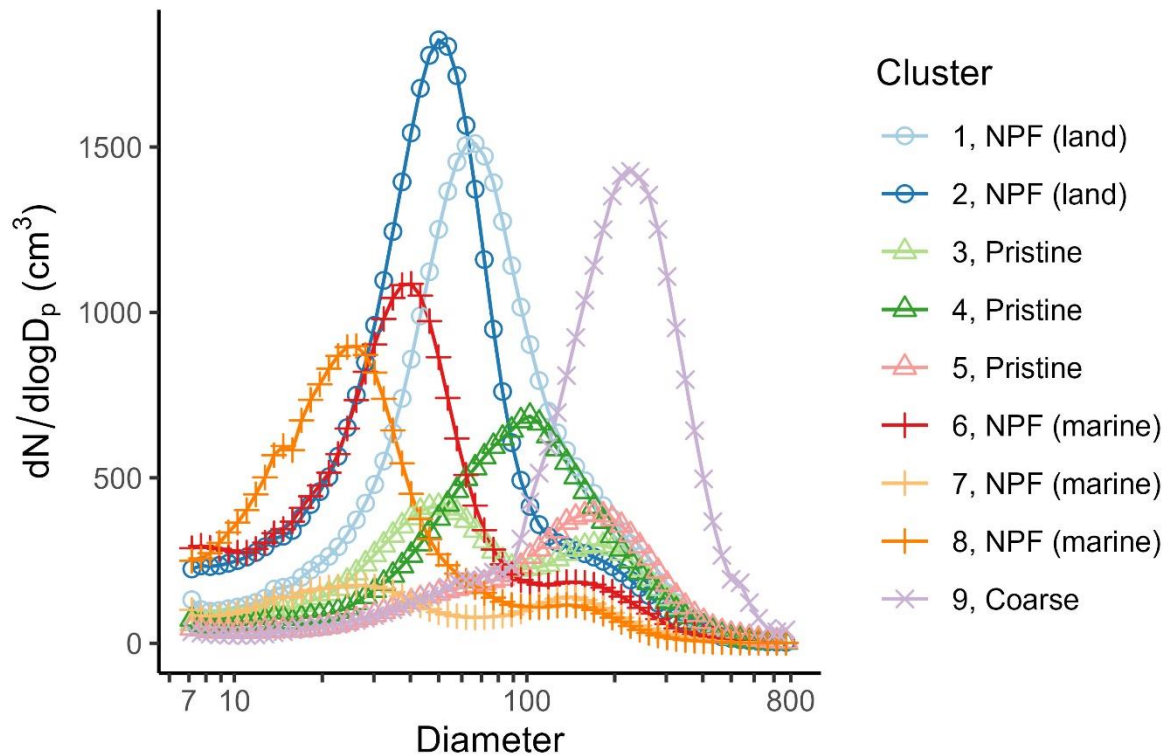


Figure R7: As Figure S6a, but for un-grouped clusters.

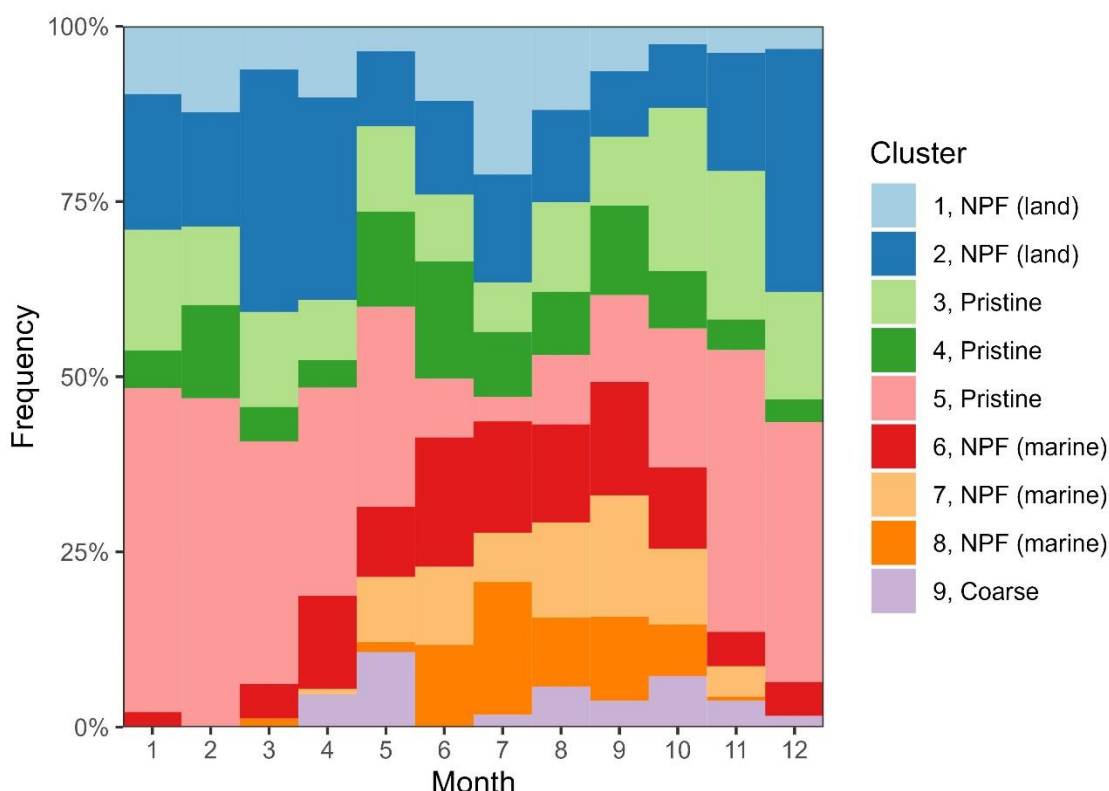


Figure R8: As Figure S6b, but for un-grouped clusters.

Regarding tDOM and MIZ, you are correct that the k-means algorithm itself does not distinguish between these influences. This is a crucial feature of the analysis. The clustering was performed blindly, using only the PNSD shape data. The fact that the resulting 'Marine NPF' cluster, defined solely by its particle dynamics, was independently found to be strongly associated with the highest tDOM exposure and lowest salinity (Figs. S7, S8) serves as powerful, unbiased evidence supporting our manual classification. It demonstrates that the link to tDOM is a robust feature of the data, not an assumption we built into the analysis.

Line 202-204: Figure 4b seems to indicate air temperature being the dominating predictor in the machine-learning model to predict N_{15-100} at Tiksi, followed by three other meteorological predictors (boundary layer height, surface pressure, and precipitation). The air mass tDOM exposure predictor falls within a group of other predictors, such as solar radiation, time over marginal ice zone, etc., with considerably lower SHAP values than the first four meteorological predictors.

Our interpretation is that these variables play different roles. Meteorology (temperature, boundary layer height, etc.) acts as the primary controlling process, while air mass history (like exposure to tDOM) represents the precursor source.

Our hypothesis is that high-tDOM waters are a source of volatile organic compounds (VOCs). The emission of these VOCs from the ocean is not constant; it's strongly controlled by meteorological factors, especially temperature (which governs volatility) and wind speed.

Therefore, the Random Forest model is not showing that tDOM is unimportant. The high SHAP value for temperature reflects its direct role in driving the emissions from the tDOM-enriched source region. We update our arguments in this section as follows:

“A positive relationship is also seen for temperature (Figure 4e), likely reflecting its dual role in driving the volatilisation of VOCs from the ocean surface ²⁸ and in promoting the

autoxidation rates of organic precursors²⁹. The model also finds an inverse relationship with boundary layer height, where shallower boundary layers strongly predict higher N_{15-100} (Fig. S11). This is consistent with the physical mechanism of dilution; precursor vapours are concentrated within a smaller volume, enhancing particle formation and growth. The relationship with precipitation is not clear, with an initial small positive influence, likely because wet scavenging reduces the condensation sink by removing larger aerosol particles, which creates more favourable conditions for NPF, but a decrease in predicted N_{15-100} at high precipitation volumes. The exact effect of precipitation is likely not clear from the model, as particle production typically follows precipitation, rather than being concurrent with it. Similarly, solar radiation influences oxidation processes by driving hydroxyl radical production (Figure 4f), while wind speed impacts the rate of precursor emission from the sea surface microlayer²⁸ (Fig. S11)."

On machine learning exercise, did the authors consider withholding the airmass history predictors (e.g., exposure to tDOM and time spent over the marginal ice zone) in training to see if resulting machine-learning model had inferior predictability than the one including either or both airmass history predictors? This may provide stronger support for the hypothesis posed for this study.

Indeed we did. We constructed four types of model. The "full_model" is the model presented in our paper (Figure 4). We also present a model without the airmass history (tDOM and MIZ exposure) "no_tDOM_or_ice", a model without temperature "no_temp", and a model without airmass history nor temperature "no_AG_Ice_or_temp". We constructed and ran each model 100 times to account for the slight differences in performance for different starting points (random forests utilise bootstrapping and therefore there is small run-to-run variation), and present the Root Mean Squared Error (RMSE) for each run below (Figure R9).

The model in the paper (Figure 4) has a Root Mean Squared Error of 65.2 cm^{-3} . Removing the tDOM and MIZ variables worsens the model's predictive power, increasing the RMSE to 71.4 cm^{-3} . For comparison, removing temperature increases the RMSE equivalently to 71.1 cm^{-3} . Removing tDOM, MIZ, and temperature variables then reduces the model's accuracy to 79.1 cm^{-3} . RMSE quantifies the average magnitude of the prediction error in the same units as the target variable, and is therefore an interpretable measure of the model's accuracy

These results are fully consistent: the manual classification demonstrates that tDOM has an amplifying effect on NPF events when they occur. The RF model confirms that tDOM exposure is a statistically significant predictor of particle concentrations, in conjunction with meteorological factors like temperature. We include the following new paragraph in our text:

"To validate our choice of predictors, we performed an ablation study. First, we tested the model without the airmass history variables. Removing the exposure to tDOM and the marginal ice zone weakened the model, increasing the Root Mean Squared Error (RMSE) from a baseline of 66.6 cm^{-3} to 73.8 cm^{-3} . This impact is comparable to removing temperature, a known key driver, confirming the predictive power of airmass history. Conversely, we also investigated the potential influence of terrestrial sources by adding airmass distance from the nearest coastline as a predictor. This yielded only a negligible improvement, reducing the RMSE by just ~2%. These tests confirm that airmass history over tDOM and the marginal ice zone are the key predictors for N_{15-100} in our model and are not a proxy for proximity to coastlines."

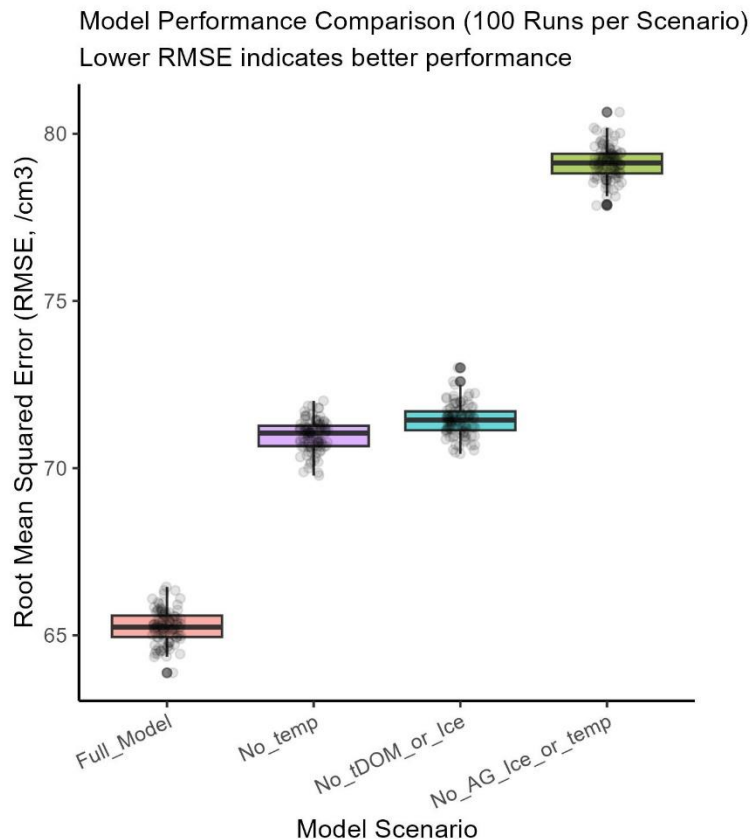


Figure R9: Four different random forest runs. “Full_model” is the model as in the paper; “No temp” is the model with temperature removed; “No_tDOM_or_Ice” is the model with air mass exposure to tDOM and the marginal ice zone removed; “No_AG_Ice_or_temp” is the model with exposure to tDOM and the marginal ice zone, as well as temperature removed.

On the basis of this we include the following line in our final manuscript:

Line 226: this seems to be an awkward statement – if PNSD is given, one would not need a model to predict N_{15-100} .

Good point—the point here is that we did not include e.g. the condensation sink. We now state

“Notably, our model accurately predicts N_{15-100} using only air mass history and meteorology, **without requiring any measurement data from Tiksi itself.**”

Line 233-234: it is not clear how the cluster analysis presented in this study yielded support for this statement (see my comment above regarding the cluster analysis).

See above enhancements to our cluster analysis segments.

Figure 5: are the data points shown for other Arctic sites taken directly from Brean et al. (2023) or reevaluated for the present study based on the data included in the said previous study (e.g., J_{15} here vs. J_{10} shown in Brean et al., 2023)? Regarding the data points related to Tiksi, are the “Terrestrial” and “Sea ice” constitute the 169-24 cases (what is the breakdown)?

Great point. We have amended our wording as follows (new text in bold)

“Figure 5: New Particle Formation the Arctic: Particle growth rates (GRs) versus particle formation rates at 15 nm (J_{15}) for six sites across the Arctic. Tiksi marine (low-tDOM) and Tiksi (high-tDOM) are the categories from Figure 3. Tiksi (Sea ice) and Tiksi (Terrestrial) are our

airmasses discarded from that figure due to sea ice and terrestrial exposure, **comprising 3 and 166 cases, respectively**. Points display means, error bars display standard errors. Data are adapted from Brean et al., 2023 comprising a total of 689 NPF events across all sites, **with formation rate calculation re-performed for J₁₅**”

Line 254-256: this is also implicated from some of the findings from existing studies, e.g., Mungall et al. (2017), where it was suggested that OVOC in sea surface microlayer observed in the Nare Strait (which contributed to marine organic particle formation) may be from the organic matter-laden river waters flow into the Strait from the Arctic basin.

In our introduction we had stated

“Mungall et al. (2017)¹⁰ reported a microlayer source of oxygenated volatile organic compounds in the summertime marine Arctic boundary layer, possibly attributed to the surface waters carrying terrigenous dissolved organic matter (tDOM) to Nares Strait in the Canadian Archipelago. Abiotic processing of the sea-surface microlayer—including photochemical or heterogeneous oxidation reactions^{11,12}—is therefore a potential source of vapours which drive NPF events. This source may be very different from those in pelagic and sympagic marine or terrestrial settings.”

We now also make the following statement in the discussion (new text in bold):

“A comprehensive understanding of Arctic climate change necessitates an appreciation of all NPF precursor sources. In sunlit ocean waters, the strong absorption of high-energy UV radiation by chromophoric dissolved organic matter leads to the photochemical oxidation and degradation of DOM³⁴, generating secondary organic precursors **and small oxygenated organics**^{35,36}. Experiments simulating this, **alongside direct ozone exposure**, produce large fluxes of organic compounds, especially nonanal^{11,12} which is known to undergo autoxidation and produce highly oxygenated molecules³⁷. **Crucially, studies using Arctic seawater show that the total flux of these VOCs correlates strongly with total organic carbon**¹². **This laboratory finding is consistent with field data from Mungall et al., who also observed a strong correlation between a marine source of oxygenated-VOCs and the concentration of dissolved organic carbon (DOC)**¹⁰.”

Line 295-296: Leaitch et al., (2016) also showed that particles smaller than 50 nm can act as CCN in summertime Arctic marine boundary layer.

Great study, thank you. We include it as follows (new text in bold):

“Aerosol-radiative forcing in certain Arctic regions is constrained by CCN scarcity¹⁸. Long-term measurements show that cloud residual concentrations at Ny-Ålesund range from 1 – 1000 cm⁻³⁵⁰. Measurements in the high Arctic indicate that cloud residuals can be even smaller than 50 nm if particle concentrations are sufficiently low⁶, **and an airborne study over Nunavut, Canada, observed that in about 40% of cases, particles ≤ 50 nm activated, with some as small as 20 nm forming cloud droplets in higher-altitude clouds**⁵¹. Aerosol radiative forcing is most notably sensitive to production of CCN when concentrations are below 10 cm⁻³. At Tiksi during summertime marine air mass conditions N_{>50} concentration concentrations fall below 10 cm⁻³ for 10% of the time indicating that this region is very sensitive to NPF (Fig. S8).”

Line 303-304: I don't think the term “Arctic emission inventory” is an appropriate term to use in this context. You could simply say that continental runoff may be an important source for marine organic aerosols in the Arctic which have not been accounted for (or represented) in current atmospheric models.

Great point. We have amended as follows

“We propose that continental runoff is a major missing source of **marine organic aerosols in tDOM influenced regions of the Arctic currently absent in current atmospheric models.**”

Line 313-315: this maybe somewhat an overstatement. Not sure if “aerosol vertical flux” is an appropriate term to use to describe NPF process.

We agree and have amended the words “vertical flux”. The passage now reads

“...highlighting an overlooked **oceanic aerosol source** in this understudied region”

References

- Beddows, D. C. S., Dall'Osto, M., and Harrison, R. M.: Cluster Analysis of Rural, Urban, and Curbside Atmospheric Particle Size Data, *Environmental Science & Technology*, 43, 4694-4700, 10.1021/es803121t, 2009.
- Brean, J., Beddows, D. C. S., Asmi, E., Virkkula, A., Quéléver, L. L. J., Sipilä, M., Van Den Heuvel, F., Lachlan-Cope, T., Jones, A., Frey, M., Lupi, A., Park, J., Yoon, Y. J., Weller, R., Marinovich, G. L., Mulena, G. C., Harrison, R. M., and Dall'Osto, M.: Multiple eco-regions contribute to the seasonal cycle of Antarctic aerosol size distributions, *Atmos. Chem. Phys.*, 25, 1145-1162, 10.5194/acp-25-1145-2025, 2025.
- Carpenter, L. J., and Nightingale, P. D.: Chemistry and Release of Gases from the Surface Ocean, *Chemical Reviews*, 115, 4015-4034, 10.1021/cr5007123, 2015.
- Dall'Osto, M., Beddows, D. C. S., Tunved, P., Krejci, R., Ström, J., Hansson, H. C., Yoon, Y. J., Park, K.-T., Becagli, S., Udisti, R., Onasch, T., O'Dowd, C. D., Simó, R., and Harrison, R. M.: Arctic sea ice melt leads to atmospheric new particle formation, *Scientific Reports*, 7, 3318, 10.1038/s41598-017-03328-1, 2017.
- Dall'Osto, M., Ovadnevaite, J., Paglione, M., Beddows, D. C. S., Ceburnis, D., Cree, C., Cortés, P., Zamanillo, M., Nunes, S. O., Pérez, G. L., Ortega-Retuerta, E., Emelianov, M., Vaqué, D., Marrasé, C., Estrada, M., Sala, M. M., Vidal, M., Fitzsimons, M. F., Beale, R., Airs, R., Rinaldi, M., Decesari, S., Cristina Facchini, M., Harrison, R. M., O'Dowd, C., and Simó, R.: Antarctic sea ice region as a source of biogenic organic nitrogen in aerosols, *Scientific Reports*, 7, 6047, 10.1038/s41598-017-06188-x, 2017.
- Kirkby, J., Amorim, A., Baltensperger, U., Carslaw, K. S., Christoudias, T., Curtius, J., Donahue, N. M., Haddad, I. E., Flagan, R. C., Gordon, H., Hansel, A., Harder, H., Junninen, H., Kulmala, M., Kürten, A., Laaksonen, A., Lehtipalo, K., Lelieveld, J., Möhler, O., Riipinen, I., Stratmann, F., Tomé, A., Virtanen, A., Volkamer, R., Winkler, P. M., and Worsnop, D. R.: Atmospheric new particle formation from the CERN CLOUD experiment, *Nature Geoscience*, 16, 948-957, 10.1038/s41561-023-01305-0, 2023.