

Response to the comments

We would like to thank all the reviews for their contributions, as we feel the quality of the manuscript has substantially improved due to their comments. We have addressed all the comments, and we have added some extra changes, that include:

- Pressure altitude has been changed to GPS altitude, which is more precise.
- Colours in the SEM graphs changed to be more intuitive and distinguishable. Bins of SEM analysis simplified, numbers of surface area have been updated (the changes are very minor).
- The back trajectories have been run as ensembles and moved to the Supplementary Information.

RC3

General comments:

The introduction is too short and does not include enough relevant background. For instance, there is only a very brief introduction to Arctic mixed-phase clouds here. Some more effort should go into describing AMPCs, INP observations, and how limited they are relative to other latitudes. For the INP aspect, since this is an Arctic manuscript, the introduction material on INPs should focus more on what has been observed in that region. There are some missing key references and studies of Arctic INPs. See my specific comments below on these and what should be included. In general, *some* of the relevant literature is included, but should be more complete and involve discerning where and when the observations were made, since those can vary quite a bit depending on location, sea ice extent, etc. This is hinted at on lines 150-151, but should be elaborated on in the introduction to develop the broader picture for the reported observations.

We have expanded the discussion of previous INP research in the Arctic and also included a specific mention of Arctic mixed phase clouds in the introduction. We do not wish to go into great depth on the arctic mixed phase clouds because we did not actually study them, but rather focused on the aerosol. We disagree that measurements are particularly 'limited' in the Arctic relative to elsewhere. There is a dearth of INP measurements around the globe. Regarding the comment on being 'more complete and involve discerning where and when the observations were made: We did exactly this and state that we compare our data to ' literature data collected in a similar location and time of the year in Fig. 3'. We then go on to set the data in the context of a compilation of Arctic data (Fig. 3b); this is now much clearer in the revised figure 3.

Even though some of the methodologies are described elsewhere, pertinent information should at least be distilled here so the readers do not have to go on a wild goose chase to obtain such information. Much of it is buried in the appendix, but why not include it in the main text? Here are some questions I had after reading through the methods, information that could be briefly added for clarity:

Overall, we have expanded the explanation of all the used methods through the text.

- When I first read it, it was not clear if only 1 filter was collected per flight, for how long, which altitudes, etc. Table A1 has this information, which I think would be important to include in the main paper.

The methodology has been expanded in the main text, the table A1 has been moved to the main paper.

- There is not enough information on filter collection, preservation, and analysis in the main text. Again, this information is in Table A1 and its caption. Should be moved to the main text. For the altitude, was the aircraft spiraling at that one altitude per filter? If so, what was the spatial coverage of

these spirals? Or were these profiles and the one altitude value provided is an average? Need more details here.

The flight tracks have been added to Fig. 1. The table is now in the main paper.

- I originally assumed the processing was not done on site, however processing was indeed done on site as stated in the caption for Table A1. This may seem like mundane information, but how the filters are preserved and handled can have a large impact on their results, so is important enough to include in the main text.

We have added a few lines on the advantages of doing this analysis on-site.

- What filter pore size was used for the filters? Different filters have different collection transmission efficiency curves.

Information about the filters and a reference to our discussion on filter collection efficiency has been added to the second paragraph of Sect 2.

- Why was Teflon used for INPs? Teflon tends to have high backgrounds compared to polycarbonate. And why polycarbonate for microscopy?

We have added a few lines on why we used it: A disadvantage of the droplet-on-filter technique is that each sample can only be analysed once, which makes it incompatible with standard heat tests, for example. However, the great advantage of the droplet-on-filter technique over techniques where particles are washed off a filter into a volume of water is that it is around 20 times more sensitive than a typical wash-off assay employing 1 μ l droplets (depending on the details of the freezing assays). This enhanced sensitivity is very important given that aerosol sampling durations are typically only a few tens of minutes long.

Polycarbonate filters were used for SEM because these give a flat surface on which we can identify particles.

- How were carbonaceous particles classified if the filter substrate material is carbon-rich?

Polycarbonate were used for SEM-EDS analysis. Particles were detected based on their morphology, and they were classified carbonaceous if they do not contain further elements beyond C and O, the elements present on the background filter. We have expanded the section on the SEM, but refer to the original paper regarding the details of the classification scheme.

- How exactly were the sample suspensions prepared? Shaken in water? For how long? What type of water was used? And was a blank water spectrum tested as well to see how much of the background is from the water itself versus the filter handling?

INP concentrations were analysed using a droplet-on-filter technique, explained in the second paragraph of the Sect. 2 (i.e. an advantage of this technique is that we did not prepare a suspension). Blanks and handling blanks were performed, as explain on this paragraph and shown in Fig. 2a.

- How many drops and how many separate tests for the drop freezing experiments?

We have added this information to the section 2: "On average, we pipetted 54 (with and standard deviation of 5) droplets per filter."

- What type of cold stage is this? Since it is not a commercial instrument, there needs to be some description of the method, even if just a few sentences.

Additional details have been added to the third paragraph of Sec. 2. We note that we have published papers with this cold stage for over a decade and citations to details are given.

- What is the temperature range of the cold stage?

The cold stage can work at temperatures within the range of 30 to -100°C, but the experiments were performed at a range of ~-10 to -35 °C.

- Was freezing detection manual (by human eye) or automatic?

The videos of the freezing were manually analysed. This has been added to the text.

- Why was a 68% confidence interval chosen? That seems like an oddly low one...

The important thing here is to report what confidence interval we report the data to. We chose to report it to 1 standard deviation (i.e. 68%) to be consistent with our previous work.

- I realize the inlet is described elsewhere, but clear referencing is needed, given aircraft inlets can be riddled with artefact issues.

Added to second paragraph of Sec. 2.

- For HYSPLIT, why were the heights and 5-days chosen? A topic of discussion in the community is how long back is long enough. I don't have a good answer to that, but the authors should at least justify why they chose the parameters that were used. Additionally, typically a good approach is to run trajectory ensembles (i.e., every so many hours during filter collection and at multiple heights around the sampling point). One trajectory per sample is not enough, given the uncertainties with HYSPLIT, especially near the poles.

We have moved the back trajectory analysis to the supplementary material, and run trajectory ensembles for each sample. In terms of why 5 days we are aware that confidence decreases the further back in time we go.

- There is hardly any information on the microscopy work. Even though it is in a previous paper, a few more sentences on the methodology should be included here. Where was the SEM-EDX analysis done? What about the details of how the samples were processed on the same filters? Were the samples frozen or stored at room temp and why? There should also be a brief synopsis of how the classifications were defined.

A more detailed description of the SEM-EDS technique has been added in response to referee 1. Some more information about the S rich and Carbonaceous categories has been added to the third paragraph of Sect. 4, alongside the existing information on the Na rich and dust categories.

These are just a handful of questions I had, but effectively, the authors should take care to include the relevant methodology information in a succinct manner in the main text instead of in the appendix. That would alleviate most of my concerns here. The details on the calculations and blank corrections could remain in the appendix, if the authors desire, but the rest would be more helpful in the main text.

We have substantially improved the amount of information about the methodology throughout the text. We hope this makes the reading more clear.

I am concerned about the background INP spectra. These are very high, and I suspect part of it is due to the use of Teflon and the fact that only a few hundred liters of air were collected per filter.

We are aware the Teflon sometimes have a higher background, however, as stated above the droplet-on-filter technique has a greater sensitivity. We cannot use the polycarbonate filters for the droplet-on-filter technique because they are not sufficiently hydrophobic. In addition, we find that the background is defined by the environment that we work in as much as the filter type.

Sure, these volumes would be sufficient in the midlatitudes, but we are dealing with the Arctic here!

We disagree with implications of this statement. The volume of air sampled is limited by the length of time we can sample for. This is dictated by the demands on time within the aircraft sorties. The total

duration is four hrs including transit, hence a few filter runs of 20 minutes each is what we have to work with. The results we produced are still meaningful, despite many being upper limits. Aircraft sampling is very challenging for many reasons and a great deal of hard and valuable work should not be dismissed with statements like the referee's.

I see in the appendix that the observations are often at or lower than the background filters. This makes interpretation of the results extremely difficult.

That is correct, we are only able to report upper limits for many of our measurements on this campaign because the INP concentrations in this location at this time were remarkably low. Nevertheless, upper limits are still very useful quantities and the results are treated and discussed accordingly.

I am also not clear on why differential spectra were used for the blank corrections,

The background subtraction must be done with differential spectrum (Vali, 2019).

or why there are still data points for the subtracted when the original datapoint was below the background.

We presume the referee is referring to the cumulative spectra. One has to remember that each point in the cumulative spectra is the integration of the differential quantity between 0°C and T. Hence, even if the differential quantity is zero at temperature T, the cumulative quantity will not be zero if the differential quantity is larger than zero in any of the points at greater T. For example, if we are confident that the INP concentration at -20 °C is 0.1 L⁻¹ (because the cumulative k is statistically above the limit of detection), then the INP concentration at -21 °C cannot be smaller than this value, regardless of what the k value at -21 °C is. This is because a cumulative spectrum cannot decrease with decreasing temperatures. This is already explained in the last paragraph of the Appendix 1 and is detailed extensively by Vali (2019).

It is evident the authors were branching into Arctic measurements (which is great!) and were perhaps not aware of the difficult sampling conditions.

We would like to point out that we have conducted a number of campaigns in the Arctic. We disagree with the idea that measurements in the Arctic are necessarily more challenging than elsewhere. This referee comment seems to be based on the notion that INP concentrations in the Arctic are always low, but this is far from the fact. For example, in our recent paper on measurements from the North Pole (Porter et al., 2022) we observed INP concentrations ranging from those representative of the remote Southern Ocean to as high as mid-latitude terrestrial locations, hence any sampling method and analysis needs to capture as much of this as possible, which is what we have aimed at.

Obviously, nothing can be done about this at this point, but the authors should spend the necessary time to describe these caveats in detail in the main text and CAREFULLY interpret the results given these caveats. The article is short as it stands, so certainly has the space to spend on appropriate descriptions of sampling issues that may lead to the results observed.

We have expanded on a number of points raised by this and other referees.

One of the novel aspects of this work is that the INP measurements were conducted ABOVE the ground. Arctic INP measurements are limited in general, but especially in the vertical. The authors should include more details on how these measurements were conducted exactly (e.g., which heights, etc.; see specific comments below) and include discussion and interpretation on these compared to previous ground based measurements, especially given the Arctic can be highly stratified year-round and certainly in the spring.

We have expanded the introduction to include the following 'Given there are strong aerosol sinks in the boundary layer, whereas the air above the boundary layer can be stratified with corresponding long aerosol lifetimes (Carslaw, 2022), vertical measurements are required. Hartmann et al. (2020)

report INP spectra for late March and early April north of 80° over the Fram Strait and Arctic Ocean and report that the highest INP concentrations ($2 \times 10^{-2} \text{ L}^{-1}$ at -15°C) correspond to the boundary layer, indicating a local marine source even though the region was mostly ice covered.'

Even if the filters suffer from high backgrounds and sampling statistical issues, this still affords key information on vertical distributions of INPs. The authors should also compare their results to previous airborne INP measurements such as those from MPACE and especially ISDAC (spring campaign in the same region).

We have included information on altitude and if the measurements was in the free troposphere or boundary layer in Table 1. The number of measurements clearly in the boundary is low, hence it is not possible to make statements about trends. The nature of the filter based technique means that it is not possible to build plots of vertical distributions of INP. This is where online instruments, such as CFDCs have an advantage.

In terms of the discussion of the relationship with other field data we have purposely focused on datasets in a similar region at the same time of year. A secondary focus is then showing how our results fit with the broader range of INP measurements that together indicate a massively variable INP concentration.

We have added additional data to the comparison in Figure 3, including data from ISDAC and MPACE. Published data from ISDAC is unfortunately limited to just a few case studies (although there is more data on the ARM data repository, information on the context of this data is limited in the literature, hence we decided to limit what we present here to the published data).

Why is there no discussion on flying in clear air versus cloud? And if out-of-cloud, what percentage of time was spent below versus in? I would expect the results to vary quite a bit, depending on these conditions. The authors need to include details and discussion on cloud conditions, in addition to those conditions over which surface types (ice, snow, open water, land), and the interpretation of those conditions with respect to the observational results.

We do not sample aerosol in clouds or in regions of precipitation. The inlet is not designed to do this. This is now clearly stated in section 2.

Specific comments:

Abstract: One of the benefits of this work is INP measurements above ground, however, the abstract does not discern if there was any sort of vertical gradient in INPs. It would be important to note the altitude ranges somewhere in the abstract, and if the INPs were vertically resolved over the whole flight(s).

The altitude range has been included in the abstract.

Lines 54-55: This is only true for the Arctic haze season in the winter/spring.

The cited references clearly show transport of low latitude dust throughout the year. No change.

Lines 59-60: Some of these were samples collected then processed in the lab, so they *could* contribute to the INP population should they become airborne. It is not clear if they actually do in the real environment.

The cited paper, Sanchez-Marroquin (2020), is a study where dust was sampled from the aircraft where the dust from Iceland clearly had become airborne. No change made.

Lines 60-61: What about dust sources that contain biogenic material, like permafrost? Could just reference Creamean et al. (2020) here.

Yes, this is another possibility. We have added this reference to the list of citations.

Paragraph starting on line 67: Should reference the new study by Creamean et al. (2022) for Arctic interseasonal annual cycles. Specifically for the western Arctic, what about the other INP measurements in this region, including airborne studies (e.g., ISDAC, MPACE)? This is especially important to bolster since these measurements are above the ground. These references should be included here for the literature background.

This important new reference has been added in the revised text: 'Similarly, year-round measurements in the central Arctic indicate peak concentrations in the summer months of 2020 (Creamean, 2022). Creamean et al. (2022) suggested that local Arctic marine sources might contribute to the elevated INP populations in the summer.'

Lines 72-73: Seems like this is an incomplete sentence. I assume the authors mean that Rinaldi et al., in contrast, did not observe the seasonal cycle? They actually do report a small increase in INPs from spring to summer. Important to note here too that Svalbard can be partially ice-free in Apr versus western Arctic locations, and the presence of pack ice can modulate the local sources of INPs.

We have edited this sentence to read 'However, a recent study did not find strong seasonality of Arctic INPs at Ny-Ålesund, although these measurements were limited to being between April and August 2018'.

Line 90: Fig 1 shows air mass backward trajectories. Where are the flight paths? Certainly important to include those somewhere.

The flight tracks have been included.

Line 109-110: Need to describe what "handling experiment" is a bit more. I assume this means collecting a blank by handling it in the same manner as the sampled filters, but this needs to be explicitly stated.

This is defined in the methods section.

Lines 123-139: This information belongs in the methods section, not the results and discussion.

We have taken the referee's advice and moved this to the methods.

Lines 145-146: Was this expected, based on the conditions and previous literature? Why or why not? A discussion of how it fits with the literature comes in the next paragraph.

Lines 150 and on: The literature comparison is great, but the point to make here is that the authors measured ABOVE the ground. Given the Arctic can be highly stratified, it is challenging to compare what was measured at the ground to that aloft. The advantage of this work is that it is above ground, even if not vertically resolved. This concept needs to be highlighted, here and throughout.

We have modified the text to stress that our sampling was airborne in the text and also added this brief discussion: 'Given the variability in INP concentrations and the limited amount of data we cannot say anything about the vertical distribution of INPs in this location. Given the atmosphere is highly stratified, it would be interesting to perform simultaneous measurements at the surface and from an aircraft to explore how INP at the surface are related to those higher in the boundary layer and those in the free troposphere might be related.'

Line 173: No mention of OPCs in the methods. This information needs to be included in that section (i.e., which OPCs, the inlet they were sampling on, etc.). Were the optical data corrected or quality controlled at all? If comparing the SEM and OPC data, there should have been a conversion of the different measured diameters, but I do not see this anywhere.

We have added a section on this in the methodology

Section 4: The authors reference the unique properties of their samples to “other regions” they have conducted the sample analyses on, but which “other regions”? The Arctic is a very different place when it comes to aerosol sources and concentrations. It seems as if the authors have not conducted much work in the Arctic, which by no means is a problem, but doing the proper legwork is necessary here. What I mean here is, specifically comparing to other microscopy studies in the Arctic. Kerri Pratt’s group has done a number of studies using CCSEM-EDX in Northern Alaska. I strongly suggest looking up her group’s papers (<https://prattlab.chem.lsa.umich.edu/pubs.php>) and comparing/contrasting the reported observations to those. Essentially, more interpretation of the results is needed here with respect to previous work done in the same region (albeit, at the ground). We primarily use the size resolved composition to say something about the INP population, rather than a discussion of the composition more generally. Also, we are not aware of composition measurements at the same time of year as ours. We have added a reference to Gunsch from Pratt’s group: ‘This is consistent with other SEM-EDS studies of the aerosol samples collected on the Alaskan Arctic from the ground {Chen, 2022 #869;Creamean, 2018 #604;Kirpes, 2018 #871;Gunsch, 2017 #870} or during a ship campaign {Kirpes, 2020 #872}.’ We have been more specific about what we mean by ‘other locations’: ‘around Iceland, the eastern tropical Atlantic and the south east of the United Kingdom’.

Lines 180-181: I see the distributions for the SEM and OPC in Fig 4, but what are the percentages of particles detected from SEM via the total population in the overlapping size bins?

We refrain from doing this. The temptation would be to say one measurement is the standard and one is therefore biased. Both methods have several caveats and assumptions, hence we leave it as a comparison.

Lines 190-192: This could LARGELY affect interpretation of the results! But, what do the authors mean by “artefacts”? Do they mean cloud residuals, compared to interstitial aerosol? Then, were cloud particles collected via filters at times and water evaporated / ice sublimated during inlet residence time or once on the filters? See my general comment above about interstitial aerosol versus cloud residuals. Additionally, any information on the cloud phase? Temperature?

This discussion is about the OPCs, not the filters. This does not affect the interpretation of the results ‘largely’. Since submitting this paper we have been working on other datasets from the CDP and also looked back at some older datasets and it appears that there is something we don’t fully understand in these size distributions. This is ongoing work and we have adjusted the text to reflect the possibilities: ‘However, for samples C087_1 and C091_2, the optical counters detected a larger concentration of particles with sizes ~5 to 10 μm than the SEM analysis of the filters. Similar discrepancies have been observed previously with these instruments (Young et al. 2015) and were attributed to regions of high humidity even if the average humidity in a run should not have led to substantial hygroscopic growth. We also note that the instrument that reports size distributions in this range is designed for cloud droplets, and we are using it at the edge of its capability for larger aerosol particles, hence there may be some biases. Another possibility is that there are losses in the filters inlet system not accounted for by Sanchez-Marroquin et al. (2019).’

Lines 211-213: Need to cite this. But additionally, how likely is this given the air mass analysis in this region?

We would prefer to leave this. It is not inconceivable that combustion or volcanic ashes are present in this region.

Lines 219 and on: Sure, this is true for lower latitudes within the dust belt, but is this really relevant for here? Why not look at trajectories farther back in time if this is a possibility? With the evidence shown, this interpretation does not fit. There are no SEM images shown, so how sure are the authors

that what they are calling mineral dust is actually dust versus industrial particles from Prudhoe Bay? Did the authors evaluate the chemical spectra in the context of the morphology as well?

Many studies show mineral dust from low latitude does make it into the Arctic (including those cited in the paper). Furthermore, the lifetime of aerosol in the Arctic is very long because the sinks are weak and the atmosphere is stratified.

We didn't look at trajectories further back in time simply because they become increasingly uncertain. Given the aerosol could have arrived in the Arctic many weeks or months before we sampled it, back trajectories are of limited use beyond establishing the likelihood of more local sources.

We have a paper on the characterisation of the SEM technique (Sanchez-Marroquin et al., 2020), so we think we do not need to take a lot of space in this paper proving this again.

We think our chemical evaluation of the particles is enough to rule them as "mineral dust" rather than some unspecified industrial aerosol since the industry in the area is mostly oilfields and these would produce other types of aerosol particles.

Lines 247-248: This cannot be confirmed without INP treatments. The authors should reword this to demonstrate this is a speculation, albeit a legitimate one.

We use the word 'suggest' which is already rather soft. We have added 'perhaps' as well: 'Hence, we suggest that the enhanced ice-nucleation ability of our samples is perhaps due to the presence of biological material'

Lines 254-257: Can the authors elaborate on this? What specifically would the biological material be from during this time of year? Should reference papers like Creamean et al. (2022) and Santl-Temkiv et al. (2019) here that do evaluate INPs/biological particles in the spring. Porter was in the late summer, so it is not exactly relevant for the Mar measurements here. Late summer sources can vary quite a bit from spring, given the contrasting transport conditions, surface open water, and marine versus sea ice biological productivity. Need to compare with previous ISDAC results (also a spring flight campaign).

It is very hard to say what it might be. Given our statement that there is some biological ice-nucleating material present is tentative, then we feel a lengthy discussion is not warranted. We have added the additional citations suggested by the referee.

We have added the limited amount of published ISDAC data to figure 3.

Fig 1: The flight IDs are difficult to discern without the table being in the main text.

The table has been moved to the main text.

Fig 2: Starts at -14C...is this a sample limit or instrumental bias? If the former, is this because of the blanks for sample volume?

The background has a tail to these temperatures. The data from this campaign unfortunately suffered from a higher background than other campaigns we have done. However, this is taken into consideration in our background subtraction and error analysis.

Fig 3: "All Arctic" is a misnomer. These are from Porter et al. from one study in Aug/Sep. Why are the authors showing data from other studies in the same timeframe, but from the late summer for Porter? The Porter data are not actually relevant for this figure. If the authors want to show the specific studies indicated in comparison to, truly all Arctic data, then other studies should be included from other times of the year as well.

We have replaced the All Arctic range by a new graph (Fig. 3b) which compares our measurements with all the existing literature of INP measurements in the Arctic.

5. Figure 4 and Table 1: I understand that Figure 4 left is the number size distribution. I assume Figure 4 right is the mass fraction of different components. My question is how to calculate the surface area of sea salt and dust in Table 1.

The left figures correspond to number size distribution while the right side fractions correspond to the number fraction of particles in each bin. This has been clarified. An explanation has also been added about the surface area: "The number size distribution is multiplied by the fraction of particles in each category and bin to calculate the number size distribution of each category. Then these number size distributions are turned into surface area size distributions and integrated to obtain the surface area of each category, as shown in Table 2."

6. Line 365 Appendix B: Did you use the mean background value of all handling blank samples and subtract this value? The frozen fraction of handling blank samples shows a larger variation (Figure 2a). How to explain the larger range of backgrounds?

In our experience handling blanks for INP analysis on some campaigns can be variable. It is not completely clear why, but very important to characterise with sufficient numbers of handling blanks from throughout the campaign.

7. Figure B3: Please use the correct legends (solid or hollow) for each sample. For example, legend markers of C091_4 should be hollow.

We could not do that as some samples have both upper limits (hollow markers) and full measurements (filled markers) at the same time. We have clarified it in the caption: "Note that full markers corresponds to measurements above the limit of detection, while hollow markers correspond to upper limits. This has not been specified in the legend as some samples have both upper limits and measurements at the same time"