Hartmann et al. present findings from shipborne INP measurements during the late spring/early summer of 2017 in the European Arctic. The study is very comprehensive, and I appreciate the targeted scientific questions presented and addressed by the research on aerosol, SML, and BSW INP analyses in tandem with supporting aerosol, meteorological, oceanographic, and air mass modelling data. The authors provide a detailed account of observations and closure resutls from the entire campaign while ultimately focusing on a case study linking SML and airborne INPs. On the flip side, there are some substantial weaknesses that should be remedied. This paper is certainly suitable for publication after the comments below are addressed.

Major comments:

I appreciate how the authors introduced scientific questions that cover crucial gaps in Arctic INP research. However, they are indeed quite broad and cannot solely be answered by one study in one region during one time period. They are indeed a fantastic objective, but the authors should be clear that their results provide insight into these questions, i.e., specifically for the European Arctic in the summer. These findings are likely not consistent with other locations and certainly, not other seasons (e.g., the spring Arctic haze season when long-range transport is prominent or winter polar night where open water and biological productivity are at their minima).

The combination of LINA and INDA is very useful, however, it is not clear why: (1) both techniques are not presented per sample even though the methods imply sample prep for both was executed and (2) one technique was presented over the other in the figures instead of both. For example, are Fig 2 and 8 INDA or LINA data? Why were INDA data not shown for Fig 3 and LINA data not shown for Fig 5? Perhaps this issue would be resolved if the authors articulated which offline technique was conducted for which samples and why.

Were the blank spectra subtracted from the filter sample data? The only place I see a blank spectrum is in Fig 11a, but is this one filter or an average of all the blanks collected as indicated in the methods? And is this blank shown from the filter in ultrapure water or ultrapure water alone? Because the background for INPs can widely vary for the type of ultrapure water used, a bit more detail on the source of the ultrapure water used should be provided (e.g., Milli-Q, DI, etc.). Blanks for both plain ultrapure water and the blank filters in ultrapure water should be shown to demonstrate the reliability of the ultrapure water and filters used. Additionally, DI blanks can be all over the board depending on the container used for sampling things like seawater. Was a blank conducted for putting ultrapure water in similar sampling containers as the SML, BSW, and fog water were collected in? If so, these blanks should probably be shown in the SML, BSW, and fog water spectra as well (at least, the first occurrence of each).

Frankly, it is a bit difficult to discern the generalized difference in the offline INP spectra from ice-free ocean, ice pack, and the MIZ for the aerosol (e.g., in Fig 3, perhaps this is somewhat clear > -15C, but not the bulk of the spectra) and those + melt ponds for the SML and BSW (e.g., Fig 5 has highest spectra from the ice pack and MIZ for SML; and ice-free ocean, melt pond, and ice pack for BSW, depending on the temperature). To support the authors' claims regarding which regions have the highest INPs, perhaps a figure summarizing the data would be useful. For example, a figure showing: (1) aerosol, (2) SML, and (3) BSW INP concentrations at select temperatures (e.g., =10, - 15, -20, etc.). Then, this would clearly demonstrate which indeed had higher concentrations and at which freezing temperatures.

Where were the SML and BSW samples collected from in the ice pack, if not from melt ponds? Were these samples collected in leads I assume?

The authors state several times that the heat tests suggest the presence of biogenic INPs, but this is generalized without any discussion on how this statement came to be. In looking at the SI figure, indeed some spectra show a decrease in INPs after treatment, but it is difficult to tell if this is the case for all samples at all temperatures. In the results and discussion when first mentioned, the authors should elaborate briefly, meaning providing a more quantitative assessment of the decrease (how much of a decrease exactly?), if this was consistent for all samples and temperatures (was it?), and why or why not.

I assume this is because it would overcrowd the figures, but why are uncertainty bars not shown on any of the spectra, aside from the fog water derived INPs? Is there a way to possibly show uncertainty to corroborate the statements regarding differences between the different sample types?

The case study is indeed interesting but could be described in more detail beyond the one SML37 sample. For instance, what did the BSW spectra look like during this time period? If there was some level of blooming happening, that should be evident in the bulk seawater more so than the SML since phytoplankton reside in the upper trophic levels of the ocean. From the limited body of work on blooms and linkages to INPs, we would not expect chl-a to correlate directly with INPs (e.g., work by Creamean et al., Irish et al., McCluskey et al., and Zeppenfeld). It is more likely that INPs increase following a bloom due to enhancement in biogenic byproducts and bacterial growth following the peak of the bloom, versus INPs originating from the phytoplankton, especially considering phytoplankton have been shown to serve as only moderately effective INPs at colder temperatures. Looking at Fig 12, particularly at temps > -15C, it is interesting how LV195 had the highest warm-temperature “bump” of INPs followed by LV196 and LV197 towards the end of and following the spike in chl-a, while LV194 was also relatively high into the first half of the spike. It certainly would be interesting to highlight and discuss any SML samples following this period in addition to BSW samples before, during, and after, for comparison. Also, it looks like some of the relatively warmer temperature SPIN data were high at the beginning of this event (-26 to -32C), but it is difficult to tell from Fig 2. To provide a nice, holistic wrap up by presenting all the data during the case study, showing and discussing the SPIN data would be useful as well.

Along these lines, the comparison of the slopes is a bit tenuous, given the slope changes could be cause by numerous factors including but not limited to sample-to-sample variability, influences from other airborne sources as indicated, etc. etc. In my opinion, the slope analysis does not add value to the results and are not as convincing as, say, the relationship between bloom time and INPs described in my previous comment. I suggest eliminating the slope discussion and focusing more on comparing the increases and decreases in the INP concentrations between the water and air in the context of the supporting measurements (chl-a, NC, NCCN, etc.). These are what provide empirical evidence of INP sources instead of arbitrarily defining slopes for only select spectra.

The authors might consider reordering the sections in the results and discussion. As it stands, this section does not flow as well as it could, and it is not clear why the case study is section 4 when it is still a part of the results and discussion (section 3). This inherently causes some of the more important findings are buried. It makes more sense to order in the following as a combined section 3 of results and discussion (note that the titles do not need to be exactly these, but something along the lines). First, describe the sources: 3.1 Atmospheric INP sources over the entire campaign (this should include low temperature as a paragraph and not a subsection) and 3.2 INPs in the SML and BSW over the entire campaign. Next, describe linkages between the sources and airborne INPs: 3.3. connecting to sea spray. Then, possible effects on fog, which includes measured and derived: 3.4 Measured and derived fog INPs. Last, focusing on a unique case: 3.5 INPs from a phytoplankton bloom.

The conclusion that dust represents the cold temperature INPs solely based on the transect in Fig 4 is very tenuous. Dust is also present at temperatures warmer than -32C, so why is this generalization based on only one temperature? How can the authors be sure what they measured were not algal or diatom material INPs, which do glaciate at the same range as dust? What kind of drop in concentrations occurred in the heat treatments at these low temperatures? Can the authors link the air mass transport pathways to any major dust source like the Sahara? It does not look like it in Fig 13, especially considering how much time the air masses spent over water as compared to land. Certainly, aerosolized redistributed dust from the ocean surface could have been a possible source, but there is no evidence of this specific to this study region. From the evidence provided, the authors should not rule out other possible INP sources at these low temperatures, especially without proof of air masses interacting with dust sources along their transect, which would entail a more detailed analysis that might include remote sensing data.

Minor comments:

Lines 21-22: The statement on heat is redundant from earlier in the abstract; should remove.

Lines 111-112: How long were samples stored until analysis?

Lines 229-234: Since the online measurements are a primary focus of the manuscript (i.e., not supporting), they should maintain their own section in the methods, perhaps before or after the offline INP measurement section.

Fig 2: Why are SPIN data missing from the very beginning of the campaign? If these data were not obtained or missing, that should be noted in the caption or figure itself.

Lines 277-278: Which “others”? This statement is a bit vague.

Lines 281-283: This is not clear given the 8-h and 2-h samples are not visually distinguished in Fig 3. Perhaps use of different markers for each would help?

Fig 6: What does it mean by “stands out in terms of ice activity”? Was this qualitatively or quantitatively determined? Either way, there should be some description of how this was defined.

Fig 11: Clarify in the caption that the grey spectra are those from the entire study.