Multiple eco-regions contribute to the seasonal cycle of Antarctic aerosol size distributions, response to reviewers.

Note: Review comments are displayed in black, our responses to those comments are coloured blue and sections that from the manuscript are coloured green. We thank the reviewers for their insightful comments and provide responses below.

Reviewer: 1

This paper presents the results of k-means cluster analysis on the particle number size distributions (PNSDs) measured at 4 different Antarctic research stations. It is an extension of the work in Lachlan-Cope et al., ACP, 2020 which applied this analysis to the same year of PNSD data from one of the stations (Halley). Applying cluster analysis to the combined data is an interesting idea but it is not clear what new insights are gained. In addition, a lot of the material in the discussion and conclusion sections seems more like introductory material. I would suggest restructuring the paper to move all introductory material to the introduction and to make clearer in the discussion what new information is gained from the combined cluster analysis. Specific comments are:

Major comments:

1. Please be specific about new insights gained from this analysis. Is it confirmation of the cluster types observed in Lachlan-Cope et al. (2020). That paper found 8 clusters and this one finds 6, so a comparison of the two sets of clusters would be helpful. Is there new information about the source of a particular cluster type? If so, please add this to the discussion.

Thank you for the comment. The new insights from this paper are 1) the direct comparison of simultaneous PNSD data, where we show that coastal sites are dominated by the nucleation mode relative to the inland site, and that particle counts are greatest in the peninsula; 2) the insights from cluster analyses, which let us see the yearly cycle of NPF-dominated and primary-dominated clusters at each site; 3) the relation of these clusters to air mass back trajectories, where we show that NPF is more related to sea ice than other clusters are, and highlights the Bellingshausen sea as an NPF hotspot. This paper builds on Lachlan-Cope et al. (2020) by extending the analyses to four sites. It is not a confirmation.

Cluster analyses produce a large number of clusters which we then recombine into a smaller number. Lachlan-Cope et al. (2020) decide to not combine the three "pristine" clusters, but in our analyses we decided to do so as they follow the same trend. We state the following in the methods:

"These were assigned into 6 categories typical of Antarctic PNSDs (Lachlan-Cope et al., 2020). Compared to previous work, we combined the three "pristine" clusters identified by Lachlan-Cope et al. (2020) into one, as they follow the same seasonal trend, producing 6 clusters instead of 8."

Further, to highlight the novel findings of the study, we have changed the abstract to read as follows

"In order to reduce the uncertainty of aerosol radiative forcing in global climate models, we need to better understand natural aerosol sources which are important to constrain the current

and pre-industrial climate. Here, we analyze Particle Number Size Distributions (PNSD) collected during a year (2015) across four coastal and inland Antarctic research bases (Halley, Marambio, Dome C and King Sejong). We utilise k-means cluster analysis to separate the PNSD data into six main categories. Nucleation and Bursting PNSDs occur 28-48% of the time between sites, most commonly at coastal sites Marambio and King Sejong where air masses mostly come from the west and travel over extensive regions of sea ice, marginal ice, and open ocean, and likely arise from new particle formation. Aitken high, Aitken low, and bimodal PNSDs occur 37-68% of the time, most commonly at Dome C on the Antarctic Plateau, and likely arise from atmospheric transport and aging from aerosol originating likely in both coastal boundary layer and free troposphere. Pristine PNSDs with low aerosol concentrations occur 12-45% of the time, most common at Halley located at low altitudes and far from the coastal melting ice and influenced by air masses from the west. We infer that both primary and secondary components from pelagic and sympagic regions strongly contribute to the annual seasonal cycle of Antarctic aerosols. Our simultaneous aerosols measurements stress the importance of the variation in atmospheric biogeochemistry across the Antarctic region.”

2. The paragraph from line 601 to line 625, the dangling sentence on lines 626 to 628, and the paragraphs from line 662 to line 691 do not seem like material that is about interpreting the results in this paper. Rather this seems like broad overview material that should be combined with the introduction.

Great suggestion, thanks, we have moved these paragraphs.

3. I don’t understand what Figure 7 adds to the discussion. Everything is the same between summer and winter except with slightly more snow and sea ice in the winter. Perhaps the goal was to show that boundary layer NPF does not contribute as much in winter, but the symbol is really hard to see. Why are the panels slanted? Is the sun shining in winter or is that supposed to be the moon?

Figure 7 illustrates the geographic and biological variation both spatially across Antarctica as well as seasonally. Regarding the findings in our study it shows 1) the increase of NPF and dominance of secondary aerosols in the summertime; 2) the dominance of primary aerosols in-land versus at the coast. More generally it shows 1) the sea ice retreat in the summer, increasing marine emissions; 2) Reduced terrestrial biological activity in the summertime; 3) darkening in the wintertime. We argue figure 7 ties our arguments regarding Antarctic aerosols together. In the text we now state

“Figure 7 shows a schematic illustration of the sea ice, microbiota, sea-to-air emissions, and primary and secondary aerosols in Antarctica. Figure 7 highlights the dominance of NPF in summertime PNSDs, and a dominance of primary aerosols during the wintertime, with these primary aerosols being more prevalent inland than at the coast, a key finding of this study. It also highlights the retreat of sea ice in the summer, leading to increased marine emissions, alongside a reduction in terrestrial biological activity and sunlight intensity during winter months.”

4. I find Figure 5 really hard to interpret with backtrajectories from all of the stations overlapping. I would move Figure S9 to the main body of the paper and skip Figure 5.

Thanks for the suggestion, we have done this and updated the text accordingly. We agree the following sections make much more sense following this format.

"Our CWT analysis grids back trajectories to 1x1 degree squares and weighs each segment of the back trajectory with the corresponding N_{tot} observed upon the air mass's arrival, performed individually for each PNSD cluster. These are plotted in Figure 5. A map highlighting source regions for N_{tot} unseparated by cluster per site is shown in Figure S8. The CWTs aggregated together for each site are shown in Figure S9. Mean heights of these trajectories are shown in Figure S10."

Minor comments:

1. The paragraph in lines 196 to 205 is confusing. The first part talks about wintertime PNSDs at Halley and the second part talks about NMR analysis from a summertime cruise. How are they related?

The intent was to discuss primary aerosol sources. We have reworded it as follows (major changes in bold)

"Previous analyses of PNSDs at Halley using k-means cluster analysis has shown that wintertime PNSDs were characterised by extremely low particle concentrations, with a bimodal PNSD appearing with a blowing snow or sea spray origin (Lachlan-Cope et al., 2020). The Antarctic wintertime at this site is therefore mostly devoid of secondary aerosol sources **and is instead dominated by primary sources**. **Some of these primary aerosols will be organic, and** NMR analyses of ambient aerosol samples show that aerosols arising from the ice-free Southern Ocean are rich in lipids and sugars, and aerosols arising from coastal areas are rich in sugars associated with plant vegetation (Decesari et al., 2020**). These sources are likely primary**, and the primary aerosol sources are therefore many and varied across Antarctica."

2. In the paragraph about the Halley data (lines 238 to 244), please state that this same cluster analysis has already been applied to the Halley data and published in Lachlan-Cope et al. (2020).

Thank you for the suggestion. We include it in fact at the start of our discussion of k-means cluster analysis

"k-means Cluster Analysis was applied to the PNSD data to apportion the PNSDs according to their shape (Beddows et al., 2009), and is routinely applied in pristine environments to PNSD data (Dall'Osto et al., 2017b, 2019b; Lachlan-Cope et al., 2020), **including the Halley dataset included in this paper (Lachlan-Cope et al., 2020**)."

3. Please define the parameters in Equation 1.

This section now reads as follows

"The condensation sink (CS, s^{-1}) represents the rate at which a vapour phase molecule will collide with pre-existing particle surface, and was calculated from the PNSD data as follows (Kulmala et al., 2012):

$$CS = 2\pi D \sum_{d_p} \beta_{m,d_p} d_p N_{d_p} \quad (1)$$

Here, D represents the diffusion coefficient of the condensing vapor, which is assumed to be sulfuric acid. The transitional regime correction factor is denoted by $\beta_{m,dp}$, d_p is the diameter of a measurement bin, and N_{dp} is the number of particles in size bin dp."

4. Please list the cluster types in the same order in the text and the figures. For example, they are listed in one order in lines 313-314, in a different order in the paragraphs in lines 320 to 348, and in yet another order in Figure 4.

Thanks for pointing this out – we have amended this throughout (and slightly tidied up Figure 3 also).

5. I find the name Bursting confusing because I think of bubble bursting but you are using it to mean new particles that fail to grow. In fact, you use bursting in both senses in the paragraph on lines 485 to 497. Is it possible to change the name of this cluster?

We opt to keep the name of this cluster the same for the sake of consistency with literature, but we do now note this in the text

"Note, the name "bursting" refers to "bursts" in particle number concentrations due to secondary formation, rather than bubble bursting."

6. I would refer to Figures S4 and S5 at the end of line 357.

We have included this.

7. Lines 378-379: Is the increased contribution of Nucleation and Bursting in September due to the sun returning?

NPF is indeed likely a function of photochemistry, as nighttime NPF is essentially unheard of, with only a few instances in urban areas (where primary precursor emissions or NO_3^- chemistry are likely the culprit). We don't explicitly make this observation as we don't want to make any unfounded mechanistic interpretations.

8. In lines 107 to114, you talk about the debate between upper troposphere NPF and boundary layer NPF. On lines 429-430, you note that trajectories corresponding to Nucleation have a lower than average height. Does this provide an answer to the debate? You don't come back to this observation in the discussion.

No, unfortunately we don't think so. Upper-tropospheric NPF followed by transportation down to the measurement site could result in many different shapes of PNSD, in fact, this is one of the key uncertainties in the *Aitken* and *bursting* PNSDs, which we highlight as follows

"*Aitken high, Aitken low, and bimodal* PNSDs occur 37-68% of the time, most commonly at Dome C on the Antarctic Plateau, and likely arise from atmospheric transport and aging from aerosol originating likely in both coastal boundary layer and free troposphere"

Further, the number of PNSDs falling into the *nucleation* cluster at Dome C is actually quite small, which we now highlight as follows

"All trajectories corresponding to Nucleation PNSDs have a lower-than-average trajectory height, although the total number of trajectories this corresponds to is relatively small (Figure S10)."

9. Line 489: Can you explain briefly what the Hoppel minimum is?

Good suggestion we have done this as follows

"The two initial clusters differ slightly because of the Hoppel minimum at 55 nm and 75 nm, respectively; the Hoppel minimum refers to a specific dip in the number concentration of aerosol particles at these sizes, suggesting variations in particle stability, growth, or origin."

10. Figure 1. Why not show the ice extent in 2015 rather than the 30-year average?

Thanks, we have done this and include the figure below here:

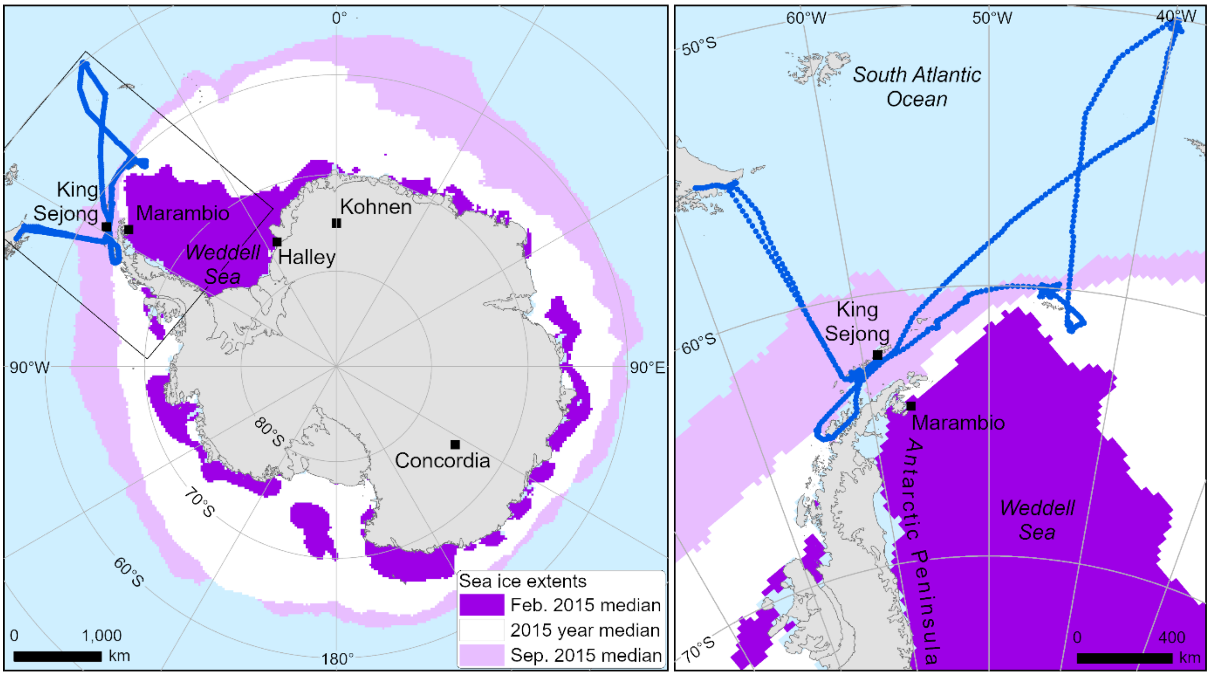
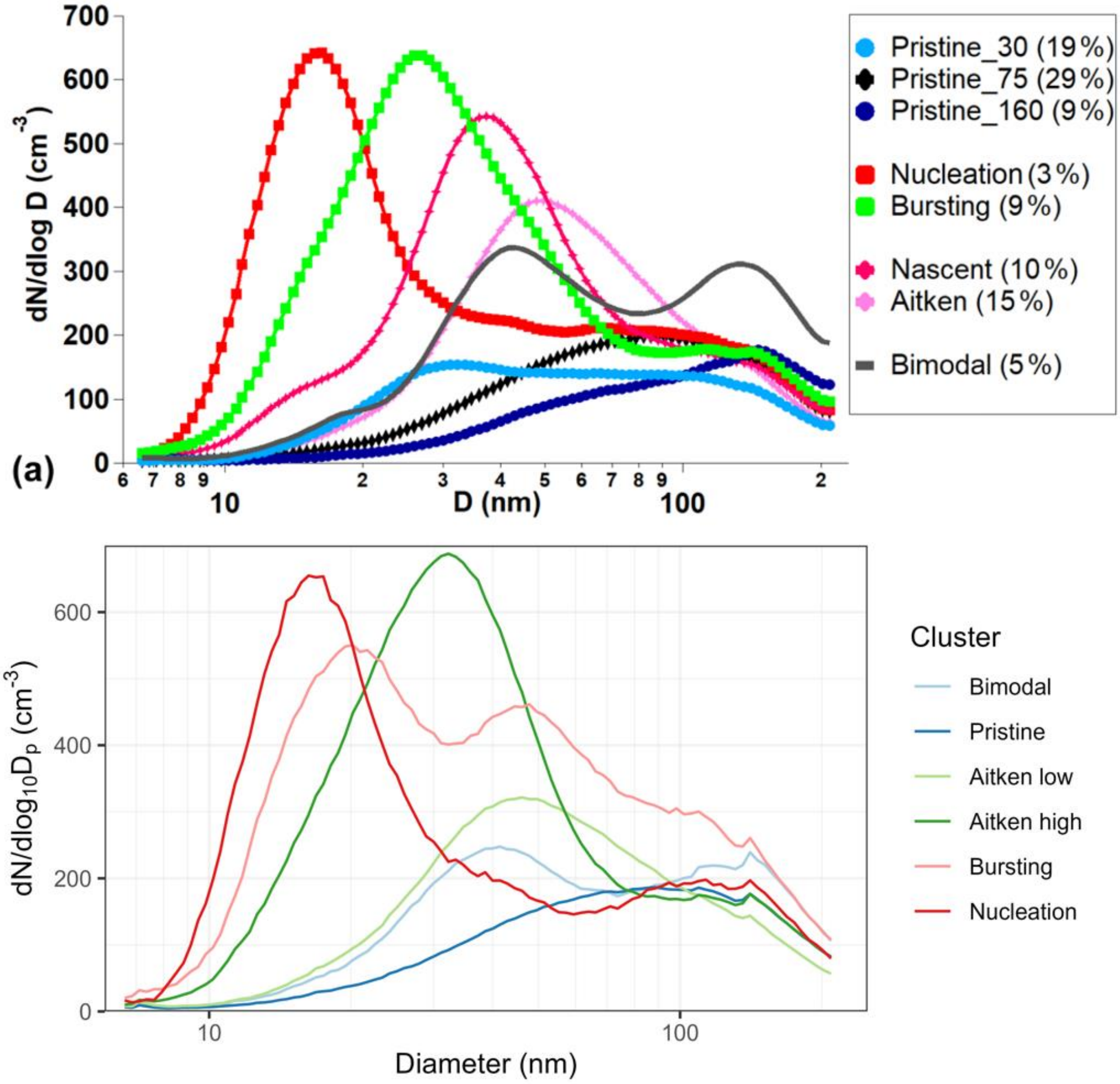


Figure 1: Map of the sampling stations (Halley, Marambio, Concordia/Dome C, King Sejong) for the year dataset collected in 2015. Additional, data for shorter period are intercompared at Kohnen (Weller et al., 2018) and during the PEGASO cruise (Dall´Osto et al., 2017, blue line the PEGASO cruise track). The February sea ice extent signifies the annual minimum, while the September median signifies the annual maximum (data are from the National Snow and Ice Data Center – NSIDC – at https://nsidc.org/data/, last access: 30 July 2024, Fetterer et al., 2017).

11. Figure 3a: I would like to see a comparison of the Halley cluster distribution from this work with the one derived in Lachlan-Cope et al. (2020), maybe in a figure in the SI. I'm curious whether the distribution of clusters changes when you include data from the other sites.

We provide a comparison of the two datasets below. The top panel is lifted from Lachlan-Cope et al. (2020), while the below is constructed from our analyses. The main difference is the *bursting* PNSDs, which are bimodal in our data, but monomodal in Lachlan-Cope et al. (2020).



12. Figure 4. Is panel a) useful? Since Dome C is entirely over Land, it seems like averaging all 4 sites together just dampens any pattern among the different clusters. It might make more sense to show the average across clusters for each site since you discuss that in the text, i.e., add a 7th column for the average for each of the 4 sites. Can you use the same names in the legend as in the text, e.g., “Sea ice” rather than “Consolidated pack ice”? Can you shift the names of the clusters slightly left so they are under the tick marks?

Good suggestions. All this has been done, although we retain panel (a) as we still argue the average is useful. Please see the below figure.

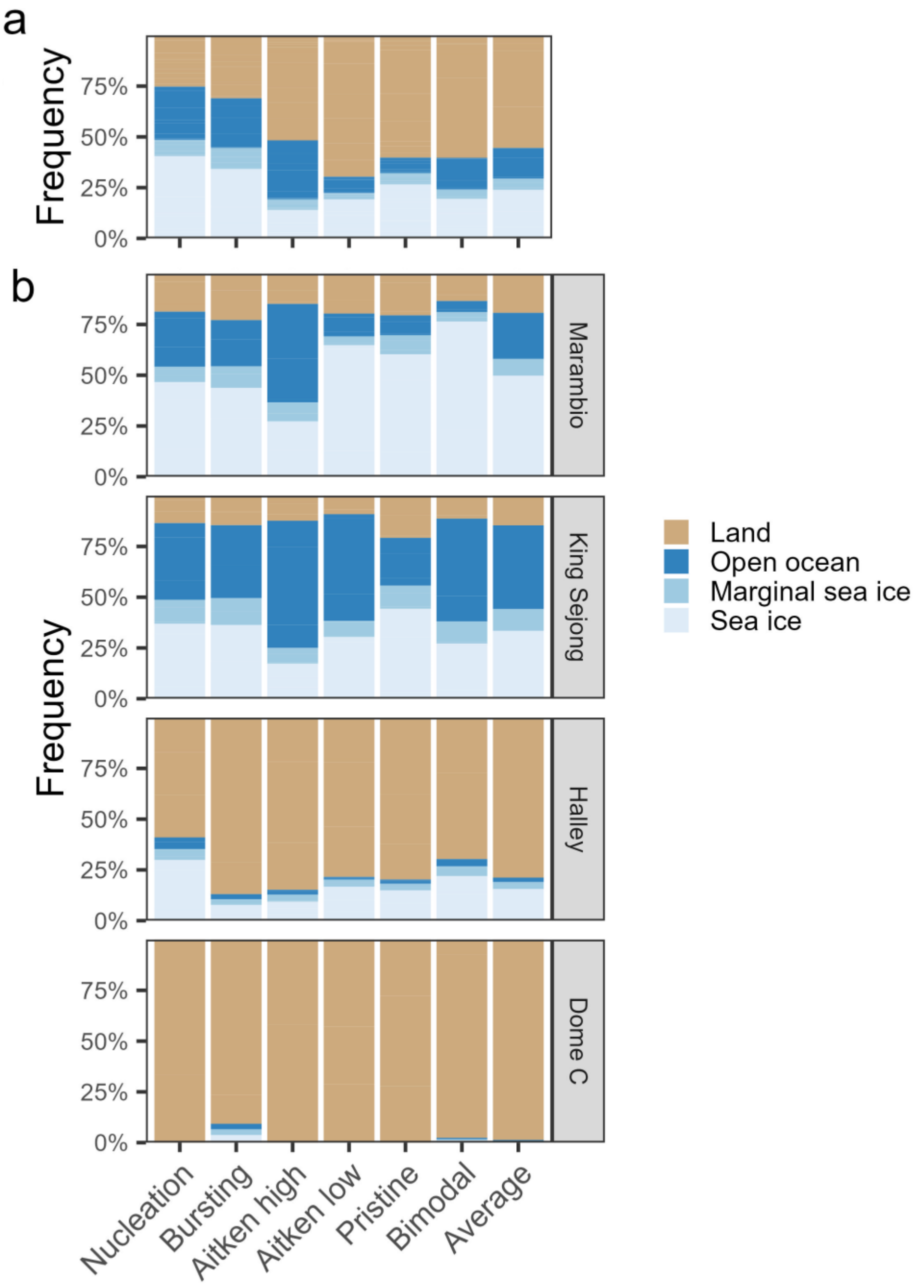


Figure 4: Land surface types associated with each cluster, showing (a) average association across all sites, and (b) association per site. ***"Average" is the mean of all clusters.***

13. Figure 6. I would use more distinct colors for the different sites. Weddell is spelled with two l's.

We have fixed this figure, please see below

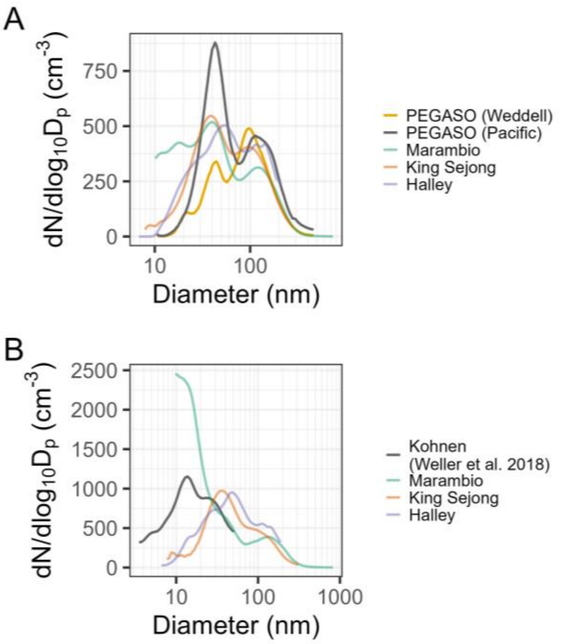


Figure 6: PNSD intercomparisons: (a) PNSD from the PEGASO cruise both when influenced from air masses from the Weddell Sea and Pacific Ocean and the stations used for this study where overlapping data are available (Marambio, Halley and King Sejong stations), and (b) PNSD from the Kohnen station (Weller et al., 2018) and the stations used for this study where overlapping data are available (Marambio, Halley and King Sejong stations)

14. Figure S6. A sentence is duplicated in the caption.

Thanks, this has been amended

15. References: Heinrichs is out of order and the two Humphries references are in the list twice.

This has been amended.

Reviewer: 2

In this study, the authors, have studied the particle number size distribution collected for one year in several stations in the Antarctic region plus some additional campaigns/measurements. They have found that the four different stations have very different size distribution possibly highlighting different sources. All the data are analysed with the k-means cluster analysis, a very common method often used by these authors. With this method they have identified six main categories (Nucleation, Bursting, Aitken high, Aitken low, Bimodal and Pristine).

The data analysis is well done and interesting however is not clear from the data analysis how they could get to their conclusions (Schematic Figure 7). Such as: “We provide evidence that both primary and secondary components from pelagic and sympagic regions strongly contribute to the annual seasonal cycle of Antarctic aerosols which add insight on the possible sources of aerosol production/activity in the whole Antarctic region.” I find it hard to connect the final schematic with the data analysis done in the study. I do understand that PSND can’t give you too much information on biological processes and chemistry. Finally, it is not really clear the advancement of this study compared to Lachlan-Cope et al., (2020).

My concern are clearly not major and probably with some change in the text and some “toning down” of the conclusions/results should be fine. I therefore think that it is a very good paper and deserve to be published in this journal.

Thanks for the comment. Firstly, we have rewritten out abstract to both be more concise, and to remove the overreaching conclusions. The new abstract reads as follows (important change to our wording in **bold**):

*“In order to reduce the uncertainty of aerosol radiative forcing in global climate models, we need to better understand natural aerosol sources which are important to constrain the current and pre-industrial climate. Here, we analyze Particle Number Size Distributions (PNSD) collected during a year (2015) across four coastal and inland Antarctic research bases (Halley, Marambio, Dome C and King Sejong). We utilise k-means cluster analysis to separate the PNSD data into six main categories. Nucleation and Bursting PNSDs occur 28-48% of the time between sites, most commonly at coastal sites Marambio and King Sejong where air masses mostly come from the west and travel over extensive regions of sea ice, marginal ice, and open ocean, and likely arise from new particle formation. Aitken high, Aitken low, and bimodal PNSDs occur 37-68% of the time, most commonly at Dome C on the Antarctic Plateau, and likely arise from atmospheric transport and aging from aerosol originating likely in both coastal boundary layer and free troposphere. Pristine PNSDs with low aerosol concentrations occur 12-45% of the time, most common at Halley located at low altitudes and far from the coastal melting ice and influenced by air masses from the west. We **infer** that both primary and secondary components from pelagic and sympagic regions strongly contribute to the annual seasonal cycle of Antarctic aerosols. Our simultaneous aerosols measurements stress the importance of the variation in atmospheric biogeochemistry across the Antarctic region.”*

We expand on this in the following sentences, *"Our study shows that the aerosol PNSDs across the Antarctic have striking differences,* ***likely*** *due to multiple eco-regions and subsequent atmospheric chemical and physical processes act as multiple aerosol sources around Antarctica."*

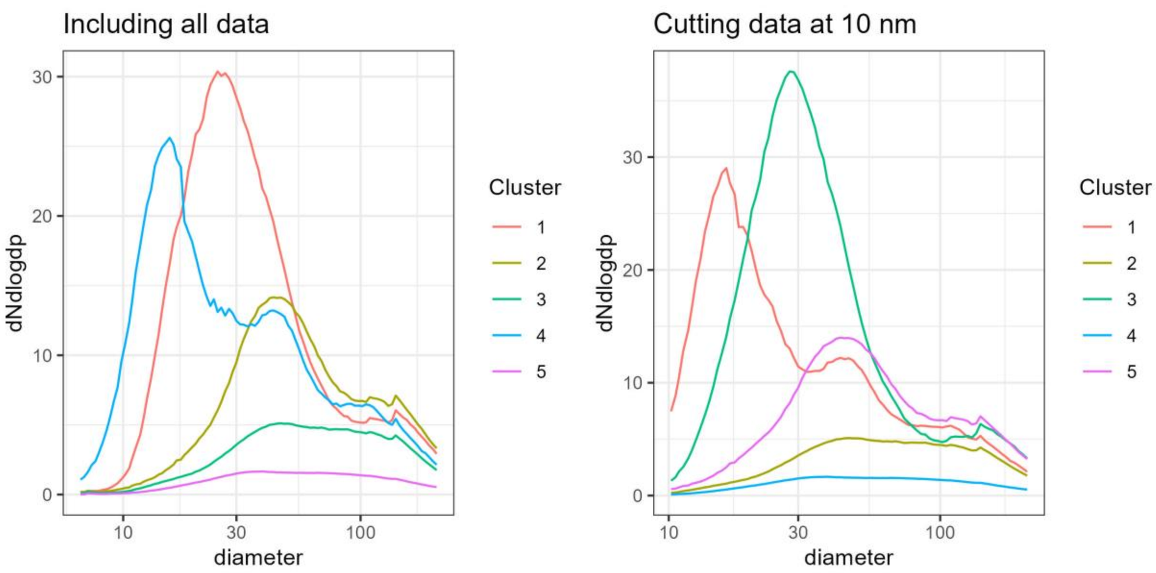
These statements are made at the end of our introduction,

"We show a prevalence of new particle formation at the coastal sites, and associate this new particle formation with air masses flowing over regions of sea ice and ocean. At the more southerly and inland sites, primary particles dominate the particle number concentrations, while air masses primarily travel over regions of land. Ambiguity remains in this analysis, as some PNSD clusters likely contains a substantial contribution from primary and secondary processes. Nonetheless, we provide further evidence for the roles of emissions from sympagic and pelagic ocean regions in new particle formation, and highlight the many and varied sources of particles across Antarctica."

Minor comments:

I have a question about the K-means clustering. How does it perform when the instruments have different cut-off (as the one shown in this study)? Especially in the nucleation mode where a small change in the cut-off can have a big impact on the particle number.

Great question. We decided not to give our instruments a common size range as the clustering works best when it has as much information as possible, however, some of the instruments have min cutoffs at 10 nm, others at lower diameters. We quickly re-analysed our Halley data (which has a min size of 6.6 nm), setting the number of clusters to just 5 (for ease of visualisation) and applied k-means cluster analysis twice, the second time we cut the smallest bins out, so our size distribution begins at 10 nm. The shapes of the resultant PNSD clusters are much the same, so we believe this makes not too much difference in our analyses.



Specify acronym such as VOC, DMS and so on.

Thank you, we have done this throughout

It feels that the discussion related to new particle formation around line 600-625 belongs more to the introduction then the results. I would consider of moving that part there.

Great, thanks for the suggestion, this has now been done.

I'm a bit puzzle by this statement in the code and data availability section: "The code and data used to produce all non-illustrative figures are available from the corresponding authors under reasonable request." What does it mean reasonable request? Shouldn't all the data and code be in a repository freely accessible. I think it's important to have these very important data (and possibly code) open access.

We aim to be as open as possible with our research materials. Researchers can request access by contacting the corresponding author, and we'll provide the data and code, following the guidelines set by those who collected each individual dataset.