Author comments: Response to Referees and Changes to Manuscript

Anonymous Referee #1

5 This scientific paper gives an over view of the Number Size Distributions (NSD) measured at several monitoring sites located around the Arctic. Data is collected from 2010 to 2015 and using k-Means cluster analysis identifies 4 NSD - one nucleation and 3 accumulation. The paper then reviews the seasonality of these data and clusters - nucleation in the summer and accumulation (associated with Haze) in the darker spring-winter months. It also considers the sources giving rise to these NSD using the analysis of air mass back

10 trajectories pointing to EuroAsia for the accumulation modes. The paper deals mainly with the accumulation mode particles and also shows that rain fall is important for aerosol scavenging. The paper gives a very useful observation of Arctic aerosol and recommended for publication.

Major Points:

15   1.   None - well written paper.

Minor Points:

    1.   Not wanting to interrupt the science, it would be useful to be able to refer to details of methodology in the Supplementary information. What type of normalisation was used before clustering and if one was
20   not used how can a size dependence be avoided in the analysis?
        Authors' Response: In sections 2.4 and 3.3 we describe the clustering methodology and provide relevant references. No "normalization" is done for this purpose so distributions with similar shapes but different concentrations, would be assigned a different cluster. However, we do homogenize the dataset, i.e. interpolate the number size distributions to a common number and bin-sizes (equally
25   spaced on a logarithmic scale). The clustering algorithm itself doesn't "care" about the sizes of the particles but the same weight is given to every bin. If the bins would have been equally spaced on a linear scale, more (undesired) weight would have been given to the larger particles. We have clarified this in the revision.
    2.   The opening sentences of Section 3.3 are not completely clear. How was the decision to use 4 clusters
30   arrived at? What happens if more clusters are chosen? Is there a spacial dependence when using 4 or more number of clusters? The output used to justify 4 clusters would be useful in the SI. Were validation statistics used and if so please present. Furthermore, was the whole 30,000 hour data set clustered at once or was there a limit set by the memory of the computer?
        Authors' Response: We have revised the first paragraph in section 3.3 to justify the selection of four
35   clusters as most suitable for the purpose of this paper as well as added a few references. The main points are that even if there are some objective ways to determine the "right" number of clusters, they may suggest a different number, and there is some room for subjective input.
        There were no memory problems for running the cluster analysis on ~30,000 distributions.
    3.   In figure 5, it would be interesting to see the minimum and maximum values plotted for each size bin
40   to get a complete picture of the cluster output.
        Authors' Response: We chose to show the interquartile range, as showing a wider range (including minimum and maximum) would make the separation between the clusters very difficult and "overcrowd" the figure. The last paragraph in Sect. 3.3 discusses those "outliers".
    4.   Pg 10 lines 35 onward, the explanation is not totally clear as to why the retreating ice and the
45   trajectories in Figure 7 explain the skewed purple bars representing cluster 4 for Alert and Nord. Please clarify.
        Authors' Response: Page 11 lines 3-4 provides the explanation: high concentrations of DMS near the retreating ice edge (which continues to get closer to Alert and Nord even after the days get shorter and the sun is lower in the sky). We've added a reference to support that.

5. Pg 11 lin 18. For the mountain measurement site ZEP (looking beyond the widening of the Hoppel gap), how often does the boundary layer drop below the height of the station and what affect does this changing boundary layer have on the NSDs? And does a higher number of clusters reveal a different shape of NSD possibly associated with trajectories not associated with cluster 1 distributions?

Authors' Response: The determination of whether Zeppelin is within the mixed layer or not is not straight forward, due to the complex topography and often stratified atmosphere with varying aerosol concentrations in the vertical dimension (we've added this and a reference to the manuscript). We only suggest this as another possible explanation for the reduced frequency of "Cluster 1" distributions at Zeppelin, compared to the Alert and Nord (along with increased North Atlantic effect, which may be more important). The slightly different median shapes of the NSD at Zeppelin might as well be due to its elevation, as stated in page 8 lines 28-36. However, increasing the number of clusters, as stated in the revised opening paragraph of section 3.3, mainly introduced additional clusters with differences in the concentrations of the smallest particles (i.e. splitting cluster 4 into sub-clusters). The effect of the added "exposure" to the free troposphere at Zeppelin is seemingly the increase in the frequency of higher cluster numbers (i.e. lower accumulation mode concentrations), rather than a significant change in the shape of the NSD.

6. Pg 13 line 40. How does the previous sentence referring to the height of the trajectories above 1km above the surface 'reinforce the claim that these areas are source region for accumulaton-mode particles in the Arctic? A detail or link is missing here - don't assume the reader will have prior knowledge.

Authors' Response: The idea is that the air mass can "pick up" particles (and/or precursors) from sources at the surface, only if the trajectory height is close to the surface (e.g. within the PBL). Air masses passing high above the ground sources would not be affected by them. Therefore, to identify as source region for accumulation mode particles by this methodology, at least two conditions need to be met: 1) high mean/median Nacc concentration; 2) a low trajectory height.
We rephrased and added some text in the manuscript for clarification.

7. Pg 14 lines 7-13. If they are of interest but you don't want to inflate the main text then put such figures into the SI.

Authors' Response: Keeping the full data rather than plotting just the mean or median could provide additional information. However, in the case of the effect of the extensive wild fires in Alaska in the beginning of July 2015, there is no need to provide a map of a higher percentile (or the maximum values), because they even show on the map of the median N_{acc} for Zeppelin (Fig. 8) as well as the mean N_{acc} map for all sites (Fig. 9b).
We've omitted the links in the footnote that refer to this event, and added a couple of recent references discussing the effects of this extreme event on aerosol observations in the European Arctic. We also rephrased the text to point to the effect of this event on the Zeppelin map.

8. Appendix A should go in the SI because the values are not discussed or add to the value of the main paper.

Authors' Response: We chose to include the log-normal fitting parameters for the monthly median number distributions as an appendix, because they describe quantitatively the curves in Fig. 4 and are referred to from the main text (end of section 3.2). In addition, the numbers are more accessible to the readers this way so they can get a general impression of the values, without the need to download another file to their computer and open it in a different program (which they will probably do in case they'd like to actually use these fitting parameters in some way). The fitting parameters for the median number distributions, as well as for other percentiles, will be available in a digital format as supplementary material. See also point 3 in the response to Referee #2.

Anonymous Referee #2

This is a scientifically sound, original and very well-written manuscript that should definitely be published in Atmospheric Chemistry and Physics. I have a few, mostly minor suggestions for clarifications when revising the paper.

1. Page 2, line 9: I would avoid using the term Anthropocene. Its starting point is ill-defined and it is very much debated how it should be defined.
   Authors' Response: OK. Changed that.

2. Page 8, lines 31-36: I do not fully agree on the argumentation here. In a remote environment, why would an elevated site have more cloud-processed aerosols as a site close to the surface? For example, a marine boundary layer is usually well-mixed, so all the air in the BL is expected to circulated at about equal efficiency through clouds. When exposed to free-tropospheric air masses, air measured in Zeppelin is expected to be even less cloud-processed as air originating from the marine boundary layer, as there tends to be less clouds in the free troposphere. I would rather think that are, in general, more clouds around Zeppelin than around other sites, causing more cloud-processing, or that clouds produce secondary material more efficiently around Zeppelin.
   Authors' Response: This is an important point. But as the last sentence in this paragraph states it is not trivial to identify whether the observed aerosol sample is from an air mass that is coupled of de-coupled from the surface. The frequent stratification of the arctic lower troposphere, together with effects of the local topography and the surface properties, make the picture much more complex than e.g. subtropical marine boundary layer. In addition, Zeppelin is often in a cloud, but that doesn't necessarily mean that it is coupled with the surface layer. Anyway, the main probable cause for the difference between the sites, as you suggest, is the increased cloudiness around Zeppelin on a long term and regional scale. We have revised this paragraph in that light.

3. Page 9, end of section 3.2: I would like to see a few sentences here discussing the interpretation of the observed modal structure of the aerosol. When looking at the modal parameter in the tables in Appendix A, one can easily see that the recorded modes have some features that differ from the traditional classification of the submicron aerosol into the nucleation, Aitken and accumulation mode. First, there are always three modes, even though the nucleation mode is in most cases absent. Second, there are usually two modes in the size range usually called the accumulation mode. What are these modes? Condensation and droplet mode, as defined in some earlier studies, or something else. Third, the two largest mode are sometime very close to each other. Are they really separable when fitting the modes?
   Authors' Response: This is more of a technical thing. The fitting procedure takes user –defined inputs, such as the number of desired modes as well as ranges of acceptable values. These should be based on physical principles, otherwise the solution may converge to a "mathematical" solution, as there is more than one set of parameters that can fit the observations quite well (especially when it is a relatively narrow range that does not often have three distinct modes). The acceptable ranges for the fitting parameters that we originally used, were quite wide and single modes could overlap. The sum of the modes though still represented very well the observed number distributions, with the relative mean absolute error well below 0.2% with respect to the total number concentration for every fit. However, in order to place the single modes within the ranges of the nucleation, Aitken and accumulation-mode ranges, we have rerun this analysis and updated the tables (this did not affect the overall quality of the fit), as well as added a few clarifying sentences in this revision.

4. Page 16, lines 31-35: There is something wrong here. The text has a totally incomplete sentence that refers to a box model which is not explained anywhere and which does not seem to fit to the context of the surrounding text.
   Authors' response: We've added a sentence to put this conceptual model in context and link it with the previous paragraph, as well as slightly modified the text and removed the term "box model".

Anonymous Referee #3

The authors analyzed a multi-year observational dataset of aerosol number size distributions from five sites around the Arctic Ocean. They conducted cluster analysis and back-trajectory analysis to investigate the seasonality and transport patterns. This study could improve the understanding on the spatiotemporal variation and transport of air pollution over the Arctic region. Before this manuscript can be considered for publication, I have a few comments for the authors to address.

1. Page 2, Lines 30-31: For the authors' consideration, a very recent study (Qi et al., 2017) used an adjoint analysis to identify source regions of black carbon over the Arctic, which could be cited here.
   References: Qi, L., Li, Q., Henze, D. K., Tseng, H.-L., and He, C.: Sources of Spring- time Surface

Black Carbon in the Arctic: An Adjoint Analysis, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1112, in review, 2017.
Authors' Response: We added this citation.

2. Measurement Section: the authors have provided qualitative descriptions on measurement sites and instrumentations. But more quantitative descriptions on the uncertainty associated with measurement instruments and methods as well as data processing should be provided in order to assess the validity of the analysis results in the text. Thus, I suggest including some quantitative discussions on the measurement uncertainties.

Authors' Response: The main point of this work, is to combine the observations from the individual sites and discuss them in the large perspective. We therefore include a rather short and mostly quantitative description of each site with the information that we find relevant. We also provide the references to the papers from the groups that were involved in making the measurements at the specific sites, where more details about the instrumentation, setup, calibrations and uncertainties can be found (although we have added in this revision the DMPS uncertainties for Zeppelin).

What we have done was using the original datasets, and homogenzing them by interpolating the data to common bin sizes, range and temporal resolution (hourly). This part doesn't introduce any uncertainties, as we discuss features of a much longer temporal scales.

We've also filtered the data to remove observations that were potentially affected by local pollution and describe how this is done, and how that may affect our results (section 2.2). However, when a number is given, we provide different percentiles to demonstrate that range of variability, which is much larger that the uncertainty in the observations or analysis methods (e.g. Table 1, Figs. 3 and 4).

We do provide an uncertainty estimate for the trajectory analysis (section 2.3), and also added in this revision the information about the relative error of the lognormal fitting results (appendix A).

3. Page 4, Line 5: Please provide the full name of "DMA" here, since this is the first time when such abbreviation appears.
Authors' response: Done.

4. Page 6, Line13: "A 240-hour 3D back-trajectory ...". A more recent study (Qi et al., 2017, see comment #1 for reference) conducted adjoint model analyses to quantify source contributions of black carbon over the Arctic region and they found that the large contributions from Asian anthropogenic sources are mainly on 1-2 month timescales, which suggested that it is likely 5-day or 10-day trajectory analyses underestimate Asian contribution to surface BC in the Arctic. Could the authors add some discussions on this aspect, since a 10-day trajectory analysis is used in this study? Also for Line 17, Page 13, would the argument here still be valid if a longer time trajectory analysis is used?

Authors' response: The reason for choosing 10 days is explained in the second paragraph of section 2.3 (balance between uncertainty in the trajectories and the typical lifetime of the particles). More than 10 days would not be suitable for the kind of analysis and geographical resolution we use in this work.

In the discussion of the results, we do state that 10 days may not be enough to cover all possible sources when removal processes are not effective (last paragraph in section 3.5), and show that the sources are likely further away than the extent of the shaded area of the maps (page 13 last paragraph and Fig. 9). This is in line with the references provided, including with Qi et. al, 2017, which was added to the discussion.

The argument in Line 17 page 13 s only relevant, as stated, for a 10-day period. Extending this time frame would extend the areal coverage of the Nord, Alert and Barrow trajectories so they may cover the area between the Caspian and Aral seas, but the uncertainties of the trajectory locations around there would be greater/much greater than 1000 km, which is double the distance between the Aral and Caspian seas.

5. Page 7, Line 16: ". . . comprising 29 bins . . ." How much uncertainty could this re- scaling/re-distribution of size bins cause for the final analysis? Is there any specific reason for choosing 29 bins?
Authors' Response: This doesn't add uncertainty as 29 bins can represent the hourly distributions (let alone the monthly distributions that are presented and analysed in the manuscript) very well and in detail.

The range of the homogenized dataset is between 20 and 500 (actually 502) nm. Bin spacing of exactly 0.05 in the log space, yields 29 bins. Smaller spacing would not add any real information. We've added this information to the manuscript.

6. Section 3.3 (Cluster analysis): What is the accuracy of the k-means analysis to divide different clusters? Is there any way to quantify this? Would the different clusters also imply different aerosol components?
Authors' response: The clustering of the aerosol number size distributions, using the k-means algorithm, has been done before and is very robust. By itself it does not introduce any errors. There are

different ways to evaluate how well the clusters represent the data, but the more relevant question is how many clusters to choose. Requesting more output clusters would inevitably reduce the total sum of errors and hence improve the accuracy, but at some point, only by little. There is no absolute "right" number of clusters to choose (even some different objective methods may suggest a different number for the same dataset), but we've evaluated the output using the "elbow" and "silhouette" methods. We have revised the first paragraph in section 3.3 (and added some references), where we try to justify our choice of four clusters.

Once we have done that, rerunning the analysis resulted in the very same output centroids (to confirm that there was no problem of the result converging to a local minimum). Of course, if different input is used (i.e. including more years etc.), the cluster means/medians may vary a little bit, but their general properties/features that are relevant for the discussion in our study would remain the same.

Different clusters may be related to different aerosol components, as they are linked with different source regions and seasons (Figs. 6 and 7). However, we only analysed the size distribution of the aerosols, and their chemical composition is not within the scope of this manuscript.

7. Figure 2: How would the data availability affect the final analysis? For example, for those years with a small fraction of available data at Tiksi and Barrow sites, would this cause any bias in the analysis of seasonality and transport?

Authors' response: This depends on the specific analysis and how the data availability varies through the year. Generally, there is no problem if observations are missing for a part of a specific year – it is more important to have a decent representation for each of the months.

For example, for the analysis of the relative monthly frequencies of the clusters (Fig. 6), the input data needs to be representative. Our requirement was to have at least two months (same month different years) with more than 50% data coverage in each of them. Barrow and Tiksi does not fulfil this requirement and therefore were excluded from this specific analysis.

A potential issue with Tiksi and Barrow is the exclusion of the data from a specific (local) wind sector (the local wind direction may have some correlation with the large-scale wind flow). This may lead to a bias in the mean/median concentrations and reduce the representativeness of the remaining data (as stated in page 5 line 29). However, as discussed in the last paragraph in page 8, it doesn't seem to change the fact the both Tiksi and Barrow tend to have greater concentrations of accumulation-mode particles than at the other sites.

For the seasonality of the transport (Fig. 12) the right thing to do is to include all trajectories – even for the data that was suspected to be affected by local pollution. We have rerun this analysis for Barrow and Tiksi, updated Fig. 12 and added this information in section 3.5 (although the changes in the maps were very minor and did not affect the discussion).

8. The HYSPLIT model analysis basically represents the transport of air pollution following the wind/flow (i.e., physical process). However, there are a lot of chemical productions of secondary aerosols during transport. How to deal with and interpret the source regions of these secondary aerosols formed in the middle of transport pathways? Would such secondary productions of aerosols affect the trajectory analysis?

Authors' response: It is true that for obtaining the results shown in Fig. 8 (median N_{acc} concentrations) the assumption is that no aerosol dynamics take place during the 10-day transport to the observation site. This requires an aerosol lifetime on that order or longer. This is why this analysis is not directly applicable of mapping the source regions of nucleation- or Aitken-mode particles (that lose their "identity" on shorter time scales). Also, part of the mass of the accumulation-mode particles observed at the Arctic sites, could originate from gaseous constituents (as stated in page 9 line 16), and be affected by cloud processing, deposition etc., which certainly take place during transport. This is why we don't claim that the median concentrations shown in Fig. 8, are directly comparable with real world values that would have been measured anywhere on the map, but rather refer to those concentrations "by association" and only use relative terms. It is definitely not an emission inventory map and where the particles have actually been emitted/formed could also be upwind or downwind of what seems to be a source region using this methodology. Without simulating the aerosol-related processes along the trajectories (and accounting for the mixing), it is very difficult to say how much the accumulation-mode particles changed in number/mass during the transport. This is highly variable, required different modelling tools and not within the scope of this paper. Even state-of-the-art models do not perform very well in simulating the aerosol concentrations and size distributions within the Arctic. This is why we decided not to assume anything about the aerosol dynamics of the accumulation mode particles (Figs. 8 and 9), but we do discuss in section 3.5 whether different processes would increase or decrease their concentrations during transport (and leave the more detailed analysis to a potential follow-up study (page 17 line 16)).

We've added a few sentences in section 3.4 to clarify this point.

Pan-Arctic Aerosol Number Size Distributions: Seasonality and Transport Patterns

Eyal Freud[1], Radovan Krejci[1], Peter Tunved[1], Richard Leaitch[2], Quynh T. Nguyen[3], Andreas Massling[4], Henrik Skov[4], Leonard Barrie[5]

[1]Department of Environmental Science and Analytical Chemistry & Bolin Centre of Climate Research, Stockholm University, Stockholm 10691, Sweden
[2]Climate Research Division, Environment and Climate Change Canada
[3]Department of Engineering, Aarhus University, Aarhus 8200, Denmark
[4]Arctic Research Center, Department of Environmental Science, Aarhus University, Roskilde 4000, Denmark
[5]Department of Geological Sciences & Bolin Centre of Climate Research, Stockholm University, Stockholm 10691, Sweden

Correspondence to: Eyal Freud (eyal.freud@aces.su.se)

Abstract. The Arctic environment has an amplified response to global climatic change. It is sensitive to human activities that mostly take place elsewhere. For this study, a multi-year set of observed aerosol number size distributions in the diameter range of 10 to 500 nm from five sites around the Arctic Ocean (Alert, Villum Research Station - Station Nord, Zeppelin, Tiksi and Barrow) was assembled and analysed.

A cluster analysis of the aerosol number size distributions, revealed four distinct distributions. Together with Lagrangian air parcel back-trajectories, they were used to link the observed aerosol number size distributions with a variety of transport regimes. This analysis yields insight into aerosol dynamics, transport and removal processes, on both an intra- and inter-monthly scales. For instance, the relative occurrence of aerosol number size distributions that indicate new particle formation (NPF) event is near zero during the dark months, and increases gradually to ~40% from spring to summer, and then collapses in autumn. Also, the likelihood of Arctic Haze aerosols is minimal in summer and peaks in April at all sites.

The residence time of accumulation-mode particles in the Arctic troposphere is typically long enough to allow tracking them back to their source regions. Air flow that passes at low altitude over central Siberia and Western Russia is associated with relatively high concentrations of accumulation-mode particles (N_{acc}) at all five sites – often above 150 cm^{-3}. There are also indications of air descending into the Arctic boundary layer after transport from lower latitudes.

The analysis of the back-trajectories together with the meteorological fields along them indicates that the main driver of the Arctic annual cycle of N_{acc}, on the larger scale, is when atmospheric transport covers the source regions for these particles in the 10-day period preceding the observations in the Arctic. The scavenging of these particles by precipitation is shown to be important on a regional scale and it is most active in summer. Cloud processing is an additional factor that enhances the N_{acc} annual cycle.

There are some consistent differences between the sites that are beyond the year-to-year variability. They are the result of differences in the proximity to the aerosol source regions and to the Arctic Ocean sea-ice edge, as well as in the exposure to free tropospheric air and in precipitation patterns – to mention a few. Hence, for most purposes, aerosol observations from a single Arctic site cannot represent the entire Arctic region. Therefore, the results presented here are a powerful observational benchmark for evaluation of detailed climate and air chemistry modelling studies of aerosols throughout the vast Arctic region.

1   Introduction

Aerosols affect climate and weather in various ways. For example, they scatter and or absorb solar radiation, reducing surface insolation and altering the atmospheric radiation budget, which is referred to as "the aerosol direct effect" (Yu et al., 2006). They also affect Earth's radiation budget by altering the cloud and precipitation properties, which is known as "the aerosol indirect effect" (Lohmann and Feichter, 2005). Meteorology is the main driver of the aerosol life cycle, so changes in weather patterns change the aerosol characteristics, which can in turn feedback and affect the meteorology. Improving the understanding of the aerosol-cloud-climate interactions is therefore crucial for reducing the uncertainties in future climatic projections.

The total radiative forcing by anthropogenic aerosol particles in the last couple of centuries ~~during the Anthropocene~~ is highly uncertain, but according to the Intergovernmental Panel on Climate Change assessment (IPCC), it is most likely negative (Boucher et al., 2013). This means that it has been masking some of the observed global warming due to increased greenhouse gases. The Arctic region is especially sensitive to perturbations of the radiative budget. It has been shown that the temperature increase rate in the Arctic region has been more than twice that of the global average since the 1980s (Cohen et al., 2014), thus highlighting the sensitive nature of this region. A recent modelling study by Acosta Navarro et al. (2016) states that one of the main causes for the so-called Arctic Amplification is the reduction in aerosol emissions in recent decades in the developed countries surrounding the Arctic, especially in Europe, consistent with a study showing the overall past influence of the Arctic aerosol as one of cooling (Najafi et al., 2015). However, warming is not homogeneous across the Arctic region and throughout the year.

There were reports of dirty snow on a large scale in the high Arctic (>70°N) already in the 19th century (Garrett and Verzella, 2008). More reports on hazy skies, especially in the springtime, were published later. The signature of anthropogenic activity on this Arctic Haze was revealed only in the late 1970s (Flyger et al., 1980; Heidam and Z., 1984; Quinn et al., 2007; Shaw, 1981), which led to the understanding that long-range transport brings the pollutants to the Arctic from distant sources.

Arctic Haze is characterized by increased atmospheric turbidity as the result of higher-than-average concentrations of accumulation-mode aerosols (Radke and Lyons, 1984; Rahn et al., 1977). It is often seen as distinct dark bands when flying above the Arctic during daylight. It is accompanied by gaseous constituents, and it exhibits a strong annual cycle with a maximum in spring and a minimum in summer (Barrie, 1986). Inefficient removal processes and a rather stable lower troposphere in winter allow the particles to stay airborne for a long time, and to travel great distances.

There is an increasing number of studies using different approaches to identify the source regions of the major Arctic short-lived pollutants and their seasonality (Croft et al., 2016b; Heidam et al., 2004; Hirdman et al., 2010b; Huang et al., 2015a; Liu et al., 2015; Massling et al., 2015; Nguyen et al., 2013; Polissar et al., 2001; Qi et al., 2017; Tunved et al., 2013). All of them indicate that periods with high levels of anthropogenic pollutants, are mostly associated with transport from northern Eurasia to the Arctic sites. These support several previous long-range transport modelling studies of the origin of Arctic haze (Barrie et al., 1989; Christensen, 1997; Heidam et al., 2004) that showed the dominance of Eurasian sources and the strong transport in the lower troposphere. In addition, (Hirdman et al., 2010a) showed that the long-term decreasing trends in black carbon and sulfate aerosol concentrations in the Arctic are dominated by changes in emissions rather than long-term trends in atmospheric transport patterns.

However, aerosols are not only transported to the Arctic, they are also formed in situ via gas to particle conversion processes. Episodes with high concentrations of nucleation-mode aerosols (dry diameter < ~20 nm), following new particle formation (NPF) events, have been documented in various climatic zones both in the boundary layer and in the free atmosphere (Kulmala et al., 2004). In order to produce a large number of new stable molecular

5   clusters in the atmosphere, some preconditions are required. These include supersaturation of the condensing vapours, such as the oxidation products of dimethyl-sulphide (DMS) and ammonia, which increases the nucleation rate (Kirkby et al., 2011), together with a low condensation sink, i.e. less particle surface for the molecules to condense upon. The most favourable conditions for NPF formation in the Arctic are in the summer months (e.g. Asmi et al., 2016; Croft et al., 2016a; Leaitch et al., 2013; Nguyen et al., 2016; Tunved et al., 2013), where

10   sulfuric acid plays a key role, but also in the spring new particle formation occurs but now initiated by HIO_3 (Sipilä et al., 2016).

Large-scale atmospheric and oceanic phenomena as well as persistent weather patterns might affect the intra-Arctic as well as year-to-year variability of Arctic Haze and NPF. On daily or weekly timescales, however, the aerosol properties are governed by the synoptic and meteorological conditions, which may induce considerable

15   variations from the mean annual pattern. This highlights the importance of conducting continuous, long-term and high-resolution aerosol measurements at multiple locations in order to characterize the aerosols across the Arctic throughout the year. Until date, the authors are not aware of any previous study that has compared observations of aerosol number size distributions from multiple Arctic sites and linked them to the atmospheric transport patterns and the general meteorology.

20   The aim of this paper is to present and discuss the differences and the similarities of the aerosol general characteristics – as inferred from their number size distributions, between different sites across the Arctic, as well as exploring their common transport pathways and the main source regions of the precursor gases and accumulation-mode particles. This allows the assessment of the spatial representativeness of the aerosol measurements at each one of the sites on varying timescales and could provide a benchmark for atmospheric

25   models with resolved aerosol number size distributions.

2   Methodology

2.1   Measurement sites and Instrumentation

The foundations of this study are observations of aerosol number size distributions from five Arctic locations (Figure 1). The Zeppelin research station in western Svalbard (78.9° N, 11.9° E) is right below the top of mount

30   Zeppelin at an elevation of 474 m (all absolute heights in this paper are with respect to the mean sea level). Villum Research Station - Station Nord (VRS) in northeastern Greenland (81.6° N, 16.7° W, 24 m) is 600 km to the west-northwest of Zeppelin. The measurement site at Alert is 700 km to the west-northwest of Station Nord. Alert, Nunavut is at the northernmost tip of the Canadian Arctic (82.5° N, 62.3° W, 210 m) only 800 km from the North Pole. Point Barrow in northern Alaska, US (71.32° N, 156.6° W, 5 m), is the southernmost site and is 3300 km

35   from Zeppelin on the opposite side of the Arctic Ocean. The observation site at Tiksi (71.6° N, 128.89° W, 35 m) completes the list and represents the Russian sector of the Arctic.

The Zeppelin research station is located ~2 km south of the small community of Ny-Ålesund, but the elevation difference of 474m as well as the prevailing wind patterns inhibit pollution from nearby sources reaching the

measurement site (Beine et al., 2001). The Norwegian Polar Institute (NP) is the station owner, and the scientific coordination is done by the Norwegian Institute of Air Research (NILU). The Department of Environmental Sciences and Analytical Chemistry (ACES) at Stockholm University has been measuring the aerosol number size distribution with a closed-loop Differential Mobility Particle Sizer (DMPS) continuously since 2000. The DMPS-system comprises a custom-built twin DMA (differential mobility analyzer) setup including one Vienna-type medium DMA coupled to a TSI CPC 3010 covering sizes between 25-800 nm and a Vienna-type short DMA coupled with at TSI CPC 3772 effectively covering sizes between 5 and 60 nm. The number size distributions from the two systems are transferred to a common size grid and then merged. Both systems use a closed-loop setup. The instrument has been inter-calibrated during an ACTRIS (www.actris.eu) workshop. Sizing and number concentrations are within 1% and 5% from the standard DMPS, respectively.

The whole air inlet conforms to WMO/GAW standards (Baltensperger et al., 2003) and EUSAAR recommendations. The current setup has the inlet drawing a flow of 100 litres per minute (lpm). The inlet is further heated and kept above 0°C to allow gradual evaporation of any droplets or ice crystals as well as to prevent freezing and build-up of ice. Inside the sampling station, the temperature is typically around 20°C.

VRS is located outside 2 km outside Station Nord, a small military airfield on a ~100 km^2 fairly flat and ice-free peninsula (Goodsite et al., 2014). The dominating south-westerly winds are caused by the katabatic flow from the ice cap. In order to minimize the effect of local pollution, the sampling site is placed south-east of the main complex (Heidam et al., 1999, 2004; Nguyen et al., 2013). VRS is located west of the ice stream that floats out from the Arctic Ocean and thus there is both seasonal and multiyear ice right at the doorstep of the station. The aerosol number size distribution in the diameter range of 10 to 900 nm is recorded in 66 bins every 5 minutes with a Scanning Mobility Particle Sizer (SMPS) (Wiedensohler et al., 2012) that is maintained and calibrated by the Department of Environmental Sciences at University of Aarhus, Denmark. The DMA part of the SMPS is a medium Vienna-type and it is followed by a butanol-based TSI CPC (model 3772).

The Dr Neil Trivett Global Atmosphere Watch Observatory at Alert Alert is the northernmost continuous atmospheric environmental monitoring site in the world. It lies 8 km from the north-eastern shore of Ellesmere Island, which is mostly snow-covered 10 months of the year. The nearby Lincoln Sea typically remains frozen year round. The aerosol number size distribution in the range of 10 to 500 nm is measured with a TSI 3034 SMPS that is calibrated on site (Leaitch et al., 2013; Steffen et al., 2014).

The measurement site Barrow is located 3 km from the Arctic Ocean, and ~5 km northeast of the centre of the town of Barrow. It is surrounded by rather flat tundra and shallow water bodies. The dominant winds at Barrow are from the eastern sector, which most often bring Arctic marine air mass to the site. The global monitoring division (GMD) at the US National Oceanic & Atmospheric Administration (NOAA) has been measuring aerosols at Barrow for a few decades, however, only recently a custom-built SMPS, measuring aerosol dry diameters in the range of 10 to 990 nm, has been installed there by the Leibniz Institute for Tropospheric Research (IfT) in Leipzig, Germany.

Tiksi is a town of ~5000 inhabitants in northern Siberia on the shore of Laptev Sea and south of the delta of the Lena river. The aerosol number size distributions are measured at the clean air facility, which is located approximately 5 km south of town, about 500 m from the shore. There is a 200 m tall hill between the site and town. The dominant local winds are from the western sector, but between April and August light winds from the sea are more common. This was accounted for when choosing the location of the measurement site in order to

minimize the contamination of the observations by local pollution. The site, as the other observations sites, is part of the International Arctic Systems for Observing the Atmosphere (IASOA), and it is run by a number of institutions (for more details see: http://www.iasoa.org). The Finnish Meteorological Institute (FMI) is responsible for the twin-DMPS system currently covering particle diameters between 3 and 800 nm. The raw data is available

5 for download from the NOAA FTP site (ftp://ftp.etl.noaa.gov/psd3/arctic/tiksi/aerosol). Asmi et al. (2016) provide a complete description of the system setup, inlet and the various routines for assuring the high quality of the data.

2.2 Data and quality control

The raw SMPS/DMPS measurements from all sites were corrected for diffusional losses, multiple charging and particle counting efficiency. In order to facilitate direct comparisons between the measurements from the different

10 stations, all integral parameters (e.g. total number, aerosol volume and effective diameter), were calculated over a size range covered by all sites, i.e. 10 to 500 nm. This range includes most of the particles that serve as cloud condensation nuclei (CCN) and, in this environment, most of the particles that scatter and absorb light. Hence, this size range has strong relevance for climate. All data were recalculated as hourly averages to match the resolution of the back-trajectory analysis that is discussed in Sect. 2.3.

15 Local pollution may affect the aerosol observations to various degrees at the different sites, although at Zeppelin this is less of an issue (e.g. Hansen et al., 2014). The Alert station is about 7 km NNE of the observatory. Data are filtered for wind directions at the observatory between 0° and 45° (true north) as well as for local events. A comparison of data at Alert indicates that particles between 20 nm and 100 nm unfiltered by wind direction are 5% higher than the filtered data, and there is no significant difference between filtered and unfiltered data for

20 particles larger than 100 nm (Leaitch: personal communication). At VRS NO_x measurements detect local pollution from the military base or the cars servicing the station, and the corresponding SMPS measurements during local pollution events, which are typically very short, were removed from the dataset (Nguyen et al., 2016). At Barrow and Tiksi, the risk of contamination of the measurements by pollution from the nearby towns is higher and observations need to be screened carefully. For Barrow, all recorded aerosol data is normally dismissed when

25 wind directions are between 130° and 360° or when the winds are weaker than 0.5 m/s (Polissar et al., 2001). However, we have noticed that despite this, there were still indications of local pollution both in the SMPS and Nephelometer data just after a local wind shift from the potentially "polluted" to the "clean" sector. An extra prerequisite was thus added, requiring that the winds from the clean sector need to be persistent for at least 24 hours. In addition, data were omitted during the first two hours after the wind shift. This procedure does not

30 necessarily filter observations that are potentially affected by the large oil and gas extraction fields at Prudhoe Bay ~300 km to the east-southeast (Kolesar et al., 2017), but this should not be regarded as local pollution anyway. The same filters were applied for the Tiksi data, but the polluted sector was confined to the azimuths between 330° and 20°.

While these conditions are useful for limiting the analysis to the background aerosol and facilitating the

35 comparison with the other sites, they inevitably somewhat reduce the temporal and spatial representativeness of the observations – for some stations more than others. Nevertheless, these datasets provide valuable information about the aerosol characteristics around the Arctic Ocean.

For a more robust statistical analysis, we included all available Arctic DMPS/SMPS observations from recent years. This is because there is no period with overlapping observations from all sites that is long enough. Covering

2.5 to 5 years of observations from each of the sites allows learning about the year-to-year variability without biasing the results. An example of environmental changes that may correlate with aerosol properties within the observation period, are the minimum and maximum extents of the Arctic Ocean sea-ice (Figure 1). In September 2012, the Arctic sea-ice coverage reached a record low in modern history (Parkinson and Comiso, 2013). On the

5   other hand , the year of 2013, had the highest mean September ice extent (together with 2009) since 2006 (Serreze and Stroeve, 2015). Together, these two years capture the last decade's sea-ice extent variability, which is driven by the synoptic conditions and the meteorology (Tilling et al., 2015).

The monthly data availability for all sites is presented in Figure 2. Each bar shows the fraction of the hourly aerosol measurements that passed the quality control and filtration procedures from the total number of hours in

10   each month. The observation sites at Zeppelin, Station Nord and Alert have monthly data coverage greater than 50% in two years or more. The site near Tiksi, however, has an annual coverage of ~40%, and the one outside Barrow is left with the poorest data availability (<25% annually) after filtration, but still every month is somewhat represented by measurements from at least one of the years.

2.3   Analysis of air-mass back-trajectories

15   In this study, the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT_4) model (Draxler and Hess, 1998) was used. The meteorological fields were obtained from the Global Data Assimilation System (GDAS) of NOAA at 1º resolution (http://ready.arl.noaa.gov/archives.php).

A 240-hour 3D back-trajectory was calculated for every hourly aerosol size distribution measured at each of the sites. The receptor altitude was set as 474 m for Zeppelin and 100 m for the other sites. The length of the back-

20   trajectory calculation was chosen as a balance between the typical lifetime of the aerosols in the arctic troposphere, which is up to two weeks (shorter in summer and longer in winter/spring) for the accumulation-mode particles (Stohl, 2006; Williams et al., 2002), and the increasing uncertainty in the calculation the further back in time it goes. The meteorological parameters along the trajectories were also saved and used for the assessment of their interplay with the aerosol properties. Furthermore, to assess the characteristics of NPF events and to allow a

25   deeper analysis in a following study, the satellite-derived sea-ice concentrations (http://nsidc.org/data/nsidc-0051) as well as the ocean depth data (http://www.ngdc.noaa.gov/mgg/bathymetry/relief.html) were added as trajectory-related parameters, but only when the trajectories were within the atmospheric mixed layer.

There are various potential sources of error in trajectory calculations. A position error of ~20% of the travelled distance is considered typical (Stohl and Seibert, 1998). One way to estimate the trajectory uncertainty is to use

30   ensembles of trajectories calculated for the same time and location. Another way is to use multi-particle dispersion models, such as FLEXTRA (Stohl et al., 1998). In this study, however, the large number of single-particle trajectories was sufficient for providing a statistically robust dataset for identifying large aerosol source regions, among other things.

The trajectory analysis used here for each of the sites (and also for all of them at once), was an adaptation to the

35   one used in Tunved et al. (2013). This was done in the following way: a concentric coordinate system with the measurement site at its pole was defined around each station. The distance between the centres of neighbouring grid cells was set as 0.5 latitude degrees and 4 longitude degrees. The increased grid cell area further away from the site due to the concentric form, offsets the fact that all trajectories converge to the aerosol site.

The hourly coordinates of the trajectory points were then projected onto the coordinate system, and all the grid cells, which had one or more trajectory points in them, were considered "hits". This was repeated for all trajectories and the integrated number of hits in each grid cell was divided by the total number of trajectories to provide an estimation of the trajectory probability (i.e. likelihood of the back-trajectories crossing a certain grid

5   cell). On a larger scale, this highlights the main transport pathways of the observed aerosols at a site, for those particles which are not locally originated.

A number of parameters that were derived from the measured aerosol number distribution at each site (e.g. aerosol total volume, number concentration of accumulation-mode particles) or simulated/integrated along the trajectory (e.g. mixing height, distance travelled, integrated precipitation), were compared, and where relevant associated

10   with their corresponding trajectory grid cells. No explicit assumption was made about aerosol dynamic processes during the transport.

2.4   Clustering the aerosol number size distributions

Aerosol number size distribution exhibits a large degree of spatial and temporal variability, reflecting the variety

15   of processes that has taken place in the air mass before the aerosols were measured. Cluster analysis serves as an excellent method for data-mining. The method relies on the grouping of data to minimize the differences within the data groups, or clusters, while simultaneously maximizing the differences between various clusters. Beddows et al., (2009) demonstrated that the k-means method (Lloyd, 1982) is the most favourable clustering method for aerosol number size distribution data. For this purpose, MATLAB programing tools were used to run the k-

20   means++ algorithm (Arthur and Vassilvitskii, 2007) and to calculate the centroids of the given number of clusters of the aerosol size-segregated number distributions. To allow this, as well as to facilitate the comparison of other aerosol integral properties between the sites, the original observations with the different size ranges and numbers of bins were homogenized and transformed to a common size grid comprising 29 bins equally distributed on a logarithmic scale ($\Delta log(Dp)$=0.05) over the 20 to 502 nm diameter range. This also ensures that the clustering

25   algorithm does not have a size dependence, as equal weight is given to each bin. Furthermore, applying finer spacing by increasing the number of bins in the homogenized dataset would not add any useful information, but just inflate the dataset.

The monthly relative frequency of the different clusters is discussed and linked with the trajectory analysis in order to evaluate the spatial association of the trajectories and the aerosol clusters. However, due to the screening

30   of the Barrow and Tiski dataset based on the local wind directions and their resulting poor data availability, the analysis for these two sites may be biased. The Barrow and Tiksi datasets are therefore excluded from the cluster analysis.

3   Results and Discussion

3.1   Annual cycle of the aerosol number, surface and volume concentrations

35   Figure 3 presents the monthly median and 10th to 90th percentile range of the total aerosol (>10 nm) and accumulation-mode (100-500 nm) number concentrations (henceforth N_{10} and N_{acc}, respectively), as well as the

same percentiles of the aerosol surface area (S_{10}) and volume (V_{10}) concentrations. All five Arctic sites exhibit an annual cycle with common features, although there are some differences between the sites.

N_{acc} (the solid grey curve in Figure 3) peaks around April at all sites, with median values between 100 and 200 cm^{-3}. The minimum median concentrations, between 20 and 50 cm^{-3}, are observed in September or October at all sites except for Tiksi, where the median N_{acc} during these months is around 70 to 80 cm^{-3} and its minimum is below 50 cm^{-3} in July.

The bottom whiskers in Figure 3 indicate that during the summer and autumn months the air occasionally becomes highly pristine; when N_{acc} drops below 10 cm^{-3} at Zeppelin, Nord and Alert, and below 20 cm^{-3} at Barrow and Tiksi. This coincides with the season of maximum precipitation in the Arctic basin (Serreze and Hurst, 2000), and implies that enhanced wet deposition is a factor in the removal of accumulation-mode particles from the lower Arctic troposphere. Further discussion on the factors that drive the aerosol annual cycle is provided in Sect. 3.5. The increased precipitation, the destabilization of the lower troposphere as well as the more heterogeneous surface properties (in the melting season) and more variable shortwave radiation in summer can modify the aerosol bulk properties and potentially contribute to an increased spatial and temporal summertime heterogeneity. The large variability in summer as well as the typical particle concentrations at the five sites, are generally in line with the observations made over the Arctic sea further from land, on board the Swedish icebreaker Oden in August and September 1991 (Covert et al., 1996).

The median S_{10} and V_{10}, denoted by the orange and blue curves in Figure 3, respectively, follow the annual cycle of the median N_{acc}. They also have the smallest variability and they peak in late spring, as well as increased variability with lower values in summer and autumn. Correlations among these three parameters are driven by the fact that most aerosol number distributions are dominated by accumulation mode particles (see Sect. Annual variation of the aerosol clusters3.3.1) and because each of these larger particles contributes more to the aerosol total surface area and volume compared to the Aitken- and nucleation-mode particles. It is important to note though that the V_{10}, more than the other bulk parameters, is underestimated due to the exclusion of the particles >500 nm from the integration. At the smaller end of the size distribution, the contribution of particles smaller than 10 nm to the total aerosol surface area and volume, is negligible.

The grey dashed curve of N_{10} in Figure 3, does not follow the other curves, which follow each other closely. Its most evident feature is the gap that opens between it and the N_{acc} curve in the summer months, and the resulting second peak in N_{10} around July-August. This double-peak in Arctic particle concentration has been observed previously (Croft et al., 2016b; Polissar et al., 2001; Tunved et al., 2013).

The two peaks in N_{10} are of different nature. While the spring maximum is governed by the number of accumulation-mode particles, the summertime peak is due to the increased concentrations of smaller nucleation- and Aitken-mode particles (more in Sect. 3.3.2). It is during the summer months when N_{10} is likely to exceed 500 or 1000 cm^{-3}, 10% of the time (depending on the site), as indicated by the top whiskers of the N_{10} curve in Figure 3.

3.2 Annual variations of the aerosol number distributions

The total number concentration, surface area and volume of the particles, do not provide explicit information regarding the shape of the aerosol size distribution. Figure 4 displays the monthly median aerosol number distributions and the interquartile ranges for all five Arctic sites. The general features of the distributions at all

locations, month by month, seem quite comparable. However, it could be noticed that in all months except between May and October, the Zeppelin curves (in grey) tend to be below the other curves, i.e. exhibit median lower number concentrations, especially for particles with diameters between 50 and 200 nm.

All monthly Zeppelin distributions exhibit a "Hoppel gap" (Frick and Hoppel, 1993; Hoppel et al., 1986) around
5   60-80 nm that is more pronounced than in the distributions of the other sites. This implies that the Zeppelin aerosol is more influenced by aerosol-cloud interactions. This is probably due to its proximity to the cloudier north Atlantic air mass, but may also be because Zeppelin is higher ~~due to its considerably higher elevation~~ than the other sites and is frequently within a cloud layer that may or may not be coupled with the surface layer ~~and proximity to the cloudier north Atlantic air mass~~. The determination whether there is such coupling at any given
10   time is far from trivial due to the frequently stratified lower atmosphere, the local topography and the heterogeneous properties of the surface. But this is ~~The altitude of Zeppelin may contribute to an increased exposure to free tropospheric air at Zeppelin, with respect to the other low-elevation sites, but the identification of these instances is not trivial and is~~ beyond the scope of this paper.

The distributions from Barrow (in blue) and Tiksi (in pink), on the other hand, have greater particle concentrations
15   than at the other sites in most months, especially in the accumulation-mode range. This is apparently not the result of omitting a subset of the data when the winds arrived from a specific sector, because the filtration of the data mainly reduced the concentrations of the Aitken-mode particles emitted by local sources. However, the filtration has in fact caused a slight increase in the median concentrations of the accumulation-mode particles at Barrow, but this still not does not change the finding that Barrow and Tiksi have the highest concentration of N_{acc} with or
20   without filtration. It is difficult to separate the contribution of the local pollution from the background aerosol based on the aerosol size distribution alone. The screening, however, seems effective in reducing the measured concentrations of the Aitken-mode particles to their actual background levels.

The monthly medians of Station Nord (orange) and Alert (green) in Figure 4 show a remarkable resemblance, which makes it hard to separate between the two in some months. These sites are closer to the pole than the other
25   sites, relatively close to each other and are furthest away from the sources of anthropogenic and marine aerosols. The concentrations there are therefore typically lower than at the other coastal and lower latitude sites (Tiksi and Barrow).

Moreover, the shape of the Arctic background aerosol number size distribution at all sites, is dominated by the accumulation mode particles during most of the year, except in summer between June and August. Such
30   distributions normally indicate an aerosol population that is rather "aged" (e.g. Tunved et al., 2013), i.e. measured far from its sources, and possibly, but not necessarily, has been part of one or more cloud cycles before being sampled. It can also mean that there is considerable precursor material associated with their origins.

From Figure 3Figure 4 it is evident that although the general *shape* of the monthly median number size distribution at all sites does not change much during the period the between October and April, the accumulation mode number
35   concentrations have an increasing trend during this period. Figure 4 shows that both inter- and intra-site variability is lowest in April. This suggests that the aerosol properties within the lower Arctic troposphere are rather homogeneous due to the weak aerosol production within the Arctic, inefficient removal processes as well as strong north south transport and vertical stratification in the preceding months. These factors allow the particles to travel further from their sources, mix horizontally and result in increased concentrations and likelihood for the

appearance of Arctic Haze around April. In other months, the temporal and spatial variability of the aerosol size distributions is considerably larger.

Later in spring, as conditions become increasingly favourable for NPF events, there is a growing tail of Aitken-mode particles. By July it clearly dominates over the concentrations of the accumulation-mode particles, but that only lasts until September-October, depending on the site. When the Arctic gets dark and the ice sheet is starting to grow again, the NPF signature on the monthly number distributions is mostly lost.

In order to facilitate a quantitative comparison of the monthly aerosol number distributions between the sites, and between other observations, studies, periods, as well as modelling results, the distributions shown in Figure 4 were described as the sum of three log-normal distributions (Jaenicke and Davies, 1976). Each log-normal distributions is characterized by three parameters; the modal number concentration (N_i), the geometrical mean diameter ($\overline{D_{p,i}}$ ~~Dg,i~~) and the modal geometrical standard deviation ($\sigma_{g,i}$). The acceptable ranges of the geometrical mean diameter that were input to the fitting algorithm were 20-30, 40-80 and 100-250 nm for the nucleation-, Aitken- and accumulation-modes, respectively. This is in order to allow the separation between the modes and for obtaining more "physical" results – even when the monthly distributions did not show three distinct modes. In total, there are nine independent fitting parameters that describe each monthly percentile of the aerosol size distribution. Appendix A provides the nine fitting parameters for the median distribution. It also shows that the relative mean absolute error is only in the order of 0.1%, indicating a very good fit between the sum of the three log-normal modes and the observations. ~~in each of the months and for each site.~~ The fitting parameters for the 10th, 25th, 50th, 75th and 90th percentiles can be found in a digital format in the online supporting material.

3.3 Clusters of aerosol number distributions

More than 30,000 hourly aerosol number distributions from Mt. Zeppelin, Station Nord (VRS) and Alert, covering two full years, were used as the input for the k-means analysis, with four output centroids. Each one of the remaining ~30,000 size distributions was ascribed to the cluster whose sum of distances from it was minimal. Different numbers of output clusters were tested and the total sum of the distances from the centroids were evaluated with the "elbow" and silhouette methods (Ketchen and Shook, 1996; Rousseeuw, 1987; Thorndike, 1953). Four clusters were found optimal for this dataset and for providing a simplified picture of the most common number size distributions. Choosing a higher number resulted in clusters with too few members and mainly with differences in the concentrations of the smallest particles, as these can be highly variable, but they are not in the focus of this work. Choosing fewer clusters would have over-simplified the picture and put in the same group distributions with rather different characteristics or physical background. ~~The number of clusters was optimized as choosing more than four output clusters resulted in one or more clusters with very few members. In addition, the total sum of the distances from the centroids started levelling out for a greater number of output clusters than the four.~~ The resulting median centroids and the interquartile ranges are shown in Figure 5 as solid curves and shaded areas, respectively.

Table 1 provides some quantitative information regarding the characteristics of the centroids. The clusters were numbered by descending aerosol effective diameter, which is useful for comparing between observations and an indication of their age (Croft et al., 2016b).

Three of the four cluster medians in Figure 5 exhibit an aerosol number distribution that is dominated by accumulation mode particles, with concentrations dropping from 150 to 19 cm[-3] between clusters 1 and 3,

16

respectively. Thus, the centroid of cluster 1 has the highest total aerosol surface area, volume and mass of 0.21 cm^2 m^{-3}, $8.9 \cdot 10^{-7}$ cm^3 m^{-3} and 1.34 µg m^{-3}, respectively – assuming spherical particles with a density of 1.5 g cm^{-3}. Clusters 2 and 3 have decreasing values in all aerosol parameters in Table 1, representing the cleaner conditions, but still with rather aged particles – as indicated by their relatively large effective diameter. The centroid of cluster 4, on the other hand, has a different shape with a mode in the Aitken range and a relatively low effective diameter of 189 nm. The median N_{10} of all number distributions associated with cluster 4 is 273 cm^{-3}, considerably higher with respect to the other clusters, but the median aerosol mass is only 0.31 µg m^{-3} – less than a quarter than the integrated mass of the centroid of cluster 1.

The shaded bands around the centroids in Figure 5 denotes the inter-quartile range within each cluster for evaluating the spread of the bulk of the distributions within the clusters. It shows little overlapping, which indicates that the clusters are quite distinct from each other. However, the values of the lower and the upper quartiles still lie outside of the shaded bands. Individual observations can deviate considerably from the centroid of their assigned cluster, but these "outliers" were part of the cluster analysis and affected its output. They are therefore accounted for despite their invisibility in Figure 5. However, the focus of this work is on the general features of the aerosol number distributions rather than the fine details of individual observations.

3.3.1   Annual variation of the aerosol clusters

The aerosol bulk properties (Figure 3) and median size distribution (Figure 4) have a pronounced annual cycle. Although some percentile information is included in those figures, it is not sufficient for understanding the monthly variability of the aerosol size distributions. The clustering of the aerosol size distributions assists in resolving this. Figure 6 shows how the probability of occurrence of an aerosol size distribution assigned to a specific cluster varies between the different months.

There is a common annual pattern in the relative occurrences of the aerosol clusters at the different sites. Most notable is the increasing occurrence of cluster 4 distributions (purple bars) from late spring to late summer – to ~40% at all sites, and the swift drop in September to below ~10%, as the daylight hours rapidly decrease. Alert and Station Nord show more skewed distributions compared to Zeppelin. At Alert and Station Nord, the maximum occurrence of cluster 4 distributions is reached in August rather than in July. This might be due to the continued retreat of the ice edge in Baffin Bay during the summer (Figure 1), where most of the cluster 4 trajectories arrive from to Alert and Station Nord (Figure 7) as well as the stronger exposure to light at more southerly Zeppelin. A closer ice edge allows enhanced concentrations (less diluted) of biogenic aerosol precursor gases such as dimethyl sulphide and an increased probability for NPF events (Levasseur, 2013). It should be noted that NPF events were not confined to only Baffin Bay air masses, as for example (Nguyen et al., 2016) reported a higher chance of observing a NPF event at Nord with southerly air masses arriving from over the Greenland sea. The trajectories that are associated with cluster 4 distributions at Zeppelin according to Figure 7, mostly arrive from the open North Atlantic. This may be the reason for the closer relation between the irradiance and relative occurrence of cluster 4 distributions (Figure 6) for Zeppelin, compared with Alert and Station Nord.

Another feature of Figure 6, common to all sites, is the increase in the monthly occurrence of cluster 1 distributions (indicative of the accumulation mode dominated Arctic Haze) from November to April. The absolute values in April are, however, quite different with ~40%, ~65% and ~90% at Zeppelin, Alert and Station Nord, respectively.

A possible explanation for the decreased occurrences of Arctic Haze at Alert compared to Station Nord, which does not seem to be due to year-to-year variability, are the frequent katabatic winds from the high mountains that pull mid tropospheric and less polluted air to the site at Alert (Morin, 2005). This is clearly seen in day to day fluctuations of ozone depleted boundary layer air and un-depleted ozone containing free tropospheric air at Alert after polar sunrise (Barrie et al., 1988, 1994).

For Zeppelin, on the other hand, there is a fair number of trajectories arriving from the North Atlantic in April that are not associated with cluster 1 distributions. In addition, according to the trajectory analysis, the air masses arrive, on average, from higher elevations compared to Station Nord, as Zeppelin is sometimes above the regional mixed layer. This results in a relative low frequency of Arctic Haze at Zeppelin in April, compared to Station Nord and Alert. However, the determination of whether Zeppelin is within the mixed layer for the individual observations is not trivial due to the complex topography and missing information about the vertical profiles of aerosols and the meteorological parameters (Beine et al., 2001).

The mean annual relative occurrence of the clusters, indicated at the top of panels in Figure 6, provide a general view on the year-to-year variability. Only the years with an annual coverage >55% are included to minimize biases. However, the Zeppelin dataset is missing April 2015 and July 2012 (Figure 2), which results in an underestimation of ~3-4% in the occurrence of clusters 1 and 4 in the respective years (~40% average monthly occurrence divided by 12 months). There are systematic differences between the sites that are not a result of the year-to-year variability. These include the greater frequency of cluster 3 and 4 distributions at Zeppelin, as well as cluster 1 at Station Nord.

Barrow and Tiksi are excluded from this specific analysis and discussion due to the wind-direction-based data filtration required due to the local particle sources (see 2). This preferentially dismisses parts of the dataset associated with transport form certain sectors with varying representation through the year. For example, the North Pacific flow to Barrow, which is more frequent in the spring and summer (see Figure 12). This filtration increases the uncertainty in the monthly relative occurrence and annual frequency and does not allow an unbiased comparison with the other sites.

3.3.2   The relationship between the atmospheric flow and the aerosol clusters

The trajectory analysis – as described in Sect. 2.3, was done to identify any relationship between the aerosol properties recorded at the various sites and the geographical positions of the sampled air masses in the preceding days. Each subplot in Figure 7 shows a trajectory probability map for a given site and aerosol cluster. It indicates that the trajectory-occurrence density field is typically far from isotropic, i.e. there are preferred pathways to each site. Trajectories associated with cluster 1 distributions (highest N_{acc}) at Zeppelin, for instance, are far more likely to arrive from the eastern sector than from the western sector. The trajectory frequency of cluster 2 (middle N_{acc}) distributions at Station Nord is another example; two main branches are separated by Ellesmere Island. The southern branch mostly follows the western coast of Greenland (an ice sheet that is 2 to 3.2 km in altitude), and this demonstrates the channelling effect of the topography on the flow in that region.

A closer examination of the green and yellow shades in Figure 7, as well as the mean trajectory positions (red curves) for each site separately, reveals a counter-clockwise rotation of the air mass origin when moving from the cluster 1 aerosol distributions with their high aerosol effective diameters, to the relatively "fresh" cluster 4 members. This applies for all three sites, indicating that an aerosol size distribution of cluster 1 is more likely to

arrive from the Russian side of the Arctic Ocean, while trajectories from North of the Canadian arctic are more likely to be associated with cluster 3 (lowest N_{acc}) distributions.

The trajectory densities that are associated with cluster 4 (recent NPF event) aerosols at the three main aerosol sites have major southerly branches (Figure 7) that come either from the North Atlantic or Baffin Bay – although there is still some contribution from intra-Arctic flow, mostly over shallow-waters close to the shoreline. Elevated DMS fluxes from the open water compared to the frozen ocean (Lana et al., 2011; Leck et al., 2002; Mungall et al., 2016) and ammonia fluxes from coastal bird colonies (Croft et al., 2016b; Wentworth et al., 2015) may be an important factor (Croft et al., 2016a). It is also evident that the trajectories of cluster 4 distributions have an extremely low probability of passing over the more polluted Asian or the European mainland during the preceding 10 days.

The areas where the NPF events take place are much closer to the sites with respect to the extent of the shaded areas for the cluster 4 distributions, at the bottom of Figure 7. This is because the coloured shades cover the area of the full 10-day trajectories, while the newly formed aerosols are likely to grow into cloud condensation nuclei (CCN)-sized particles (>~70 nm) in a day or two (Dal Maso et al., 2005; Kulmala et al., 2001). At high-latitude sites including the Arctic it may take up to three days, due to a mean growth rate of newly formed particles of as low as 1 nm hr^{-1} (Asmi et al., 2016; Kulmala et al., 2004; Ruuskanen et al., 2007; Ström et al., 2009; Tunved et al., 2003), unless close to the surface of the open sea in the Arctic where such growth may occur over a few hours (Willis et al., 2016). A large part of the area covered by cluster 4 trajectories is where the air mass may have been exposed to marine and coastal precursor gases and where the condensation sinks are small due to low pre-existing particle concentrations, both of which enhance the probability of NPF.

3.4   Source regions of accumulation-mode particles

Arctic Haze is characterized by elevated concentrations of light scattering and absorbing accumulation-mode particles. The residence time of a particle in the accumulation-mode diameter range, is rather long with respect to the smaller or larger size ranges – especially when there is low precipitation such as in the winter/spring Arctic air mass (Barrie, 1986). This means that accumulation-mode particles can travel great distances in the Arctic and be traced back to their source regions.

Figure 8 shows by site, the spatial distribution of median concentrations of accumulation mode particles in the trajectory grid cells. It was derived from the N_{acc} concentrations observed at the sites at the time of the air mass arrival combined with the origin of air mass based on back trajectory analysis. Each site's entire dataset is included (Figure 2). To reduce uncertainties, the shaded areas in each panel include only grid cells that were crossed by at least five trajectories. Displaying the median N_{acc} is useful for observing the common features of its spatial distributions, because it is not affected by extreme cases (like the mean value).

The median N_{acc} values at Tiksi and Barrow are higher with respect to the other sites, as indicated by the greater extent of green and yellow shades. Zeppelin is most affected by continental Europe, although these trajectories are not associated with the highest median N_{acc}. The regionally elevated Zeppelin N_{acc} values, associated with North Atlantic trajectories, are possibly due to the contribution of the sea salt particles (Glantz et al., 2014).

The similarities between the highlighted regions in all Figure 8 panels indicate that the Arctic sites share the Asian side of the Arctic as the main large-scale source region of accumulation mode aerosols. This is consistent with modelling studies of Arctic haze transport (e.g. Barrie et al., 1989; Christensen, 1997; and many others). Analyses

involving observational data, modelling and or emissions (Hirdman et al., 2010b; Qi et al., 2017; Sharma et al., 2006) highlight the same regions as potential sources of black carbon for the Arctic sites.

According to Figure 8, Zeppelin and Tiski have an additional source region for accumulation-mode particles from western Russia and western Kazakhstan below 50° N – the area to the north of the Black, Caspian and Aral Seas.

5   The eastern part of this area is one of the global hot-spots for desert dust (Engelstaedter and Washington, 2007), which smaller particles are in the accumulation mode size range (e.g. Mahowald et al., 2014). Trajectories from this region apparently barely reach the other sites within the ten days' frame of the trajectory analysis, but would probably do with longer trajectories.

The median accumulation-mode particle concentration over the Arctic Ocean is lower than that over land, because

10   it is frequented more by trajectories associated with lower concentrations (higher cluster numbers) that also pass over the ocean (Figure 7).

It is also important to note that no assumption was made about aerosol dynamics during the transport and thus the N_{acc} values were kept constant. This is to avoid introducing any additional complexities and uncertainties, and the justification is the longer lifetime of the accumulation-mode particles with respect to the time span of the

15   trajectories. The N_{acc} values in Figure 8 are therefore not a direct estimate of the actual median N_{acc} concentrations that would have been measured at an observation site anywhere on the map.

Furthermore, when high N_{acc} occur near the edge of the analysed domain in Figure 8, it is possible that the source regions are farther away. It may also be possible that some of these regions are "in the shade" or "behind" the actual source regions, so the trajectories first pass over those regions before reaching the source regions. However,

20   without trajectories that cover the "shaded" regions and avoid the actual source regions, it is not possible to separate those regions using this methodology. As the durations of high quality observations increase, these analyses will reveal even more clear results with greater accuracy.

To avoid further filtering and to improve the statistics, the results of the trajectory analysis shown in Figure 8 do not account for the altitude of the trajectory above the surface. This means that the median N_{acc} was derived from

25   all trajectory heights for each of the grid cells. To ensure that this does not affect the results of the analysis, similar maps were derived for two subsets of trajectory points: (i) those that were within twice mixing level height according to the meteorological dataset; and (ii) those above that – representing the planetary boundary layer (PBL) and free troposphere (FT), respectively. The distributions did not show any considerable difference (not shown here). This may be due to the increasing uncertainties in the trajectory height with time, which makes in

30   many cases the determination of whether a single-particle trajectory, which is a few days old, is within the PBL or in the FT very unreliable. For that, running trajectory ensembles with different methods for treating vertical motions may be beneficial, but it is not within the scope of this analysis.

Instead, panel (a) in Figure 9 shows the map of the mean trajectory altitude for the combined dataset – containing all trajectories from all five sites. Panel (b) displays the median N_{acc} concentrations derived from this dataset.

35   They show that over central Siberia and western Russia, where the median N_{acc} is rather high, the mean trajectory altitude is ~~around~~ often less than 1000 m above the surface, i.e. it is likely that most trajectories in this region are influenced by the surface. This reinforces the claim that these areas are source regions for accumulation-mode particles in the Arctic. The area to the south and east of lake Baikal is also highlighted with high N_{acc} concentrations, but the mean trajectory altitude there is more than 2000 m, which may be an indication that the

40   aerosols are originated from sources further away. For example, Huang et al. (2015b) showed that Asian dust

occasionally reaches the Arctic region, ~~and~~ (Liu et al., (2015) identified the main source of the~~ir~~ observed black carbon in the lower troposphere of the European Arctic as the region between 50 and 60 degrees north in Asia, ~~and~~ (Qi et al., (2017) showed that springtime BC sources for Barrow, Alert and Zeppelin is dominated of Asian anthropogenic origin, while later in the year it is dominated by Siberian open biomass burning.

5   Additional information about the variability can be obtained by plotting and comparing higher and lower percentile values, rather than only the median or mean, as they do not provide the full information and may be affected by ~~.~~ ~~However, sometimes the~~ extreme ~~cases~~ events. ~~are those of interest and they show up in the map displaying the maximum values (not shown here).~~ One such case occurred around July 11th, 2015, which brought heavy pollution from Central Alaska to Zeppelin, after the occurrence of an exceptionally high number of ~~forest~~

10   wild fires in the preceding weeks[1] (Markowicz et al., 2016; Moroni et al., 2017). The associated extreme N_{acc} values ~~considerably~~ even affect the grid cell ~~mean~~median value~~s~~ for the region around Alaska (Figure 8 top right) because they were not counteracted by trajectories associated with low N_{acc} values at Zeppelin during the study period.~~, and thus demonstrate why keeping the full information is useful for avoiding misinterpretation of the results.~~ This effect of this biomass burning event also shows up in the mean N_{acc} values when all sites are mapped

15   together (Figure 9b).

3.5   Main drivers of the annual cycle of N_{acc}

During the course of a year, the Arctic environment undergoes various changes with the potential to affect local and regional aerosol properties. In a modelling study, (Croft et al., 2016b), identified wet removal by snow or rain as the main sink for accumulation-mode particles. Condensation (including cloud processing) and transport were

20   found to be as the main sources of these particles. As modelled, all these processes were most active in the summer months. Their results are consistent with observed Arctic precipitation distributions although there is considerable variability and uncertainty in the precipitation amounts and distributions in the Arctic (Serreze and Hurst, 2000). There is general agreement that most of the Arctic receives more precipitation in summer than in winter, except for the North Atlantic sector (including the Spitzbergen region). This is due to increased moisture and heat fluxes

25   from the sea when sea ice retreats, which favours cloud formation and precipitation and the lack of moisture in very cold Arctic air masses.

For wet removal of aerosols, precipitation along the trajectory (P_{traj}) is more relevant than the mean monthly precipitation around the measurement site itself. This information can be derived from the meteorological fields used for the trajectory analysis. Although there is a large uncertainty in P_{traj}, for individual days, combining many

30   days allows for a more accurate estimate of the monthly median. Panel (a) in Figure 10 shows that the peak median accumulated precipitation along the 240-hr back-trajectories is around August for all sites except for Zeppelin – where this peak is in September (the dark season values at Zeppelin are considerably higher too). This is a couple of months later than the maximum monthly local precipitation (Serreze and Hurst, 2000) – possibly and partly due to the change in the transport patterns (panel (c) in Figure 10) and the continued melting of the sea ice as well

---

[1] ~~The URLs for the fire map of the period 30 June to 10 July, 2015, and a post by the Alaska Division of Forestry from 1 July 2015:~~
~~https://lance.modaps.eosdis.nasa.gov/imagery/firemaps/firemap.2015181-2015190.2048x1024.jpg~~
~~https://akfireinfo.com/2015/07/01/could-the-2015-alaska-fire-season-top-the-record-year-of-2004/~~

as changes in available moisture as air temperatures and absolute humidity rise. On average, the air masses spend more time in August than in June at lower/wetter latitudes over open waters, potentially increasing P_{traj}.

The trajectory-precipitation maximum is nearly coincident with the lowest monthly N_{acc} concentrations (Figure 3), and they are anti-correlated. This suggests that wet removal is the main driver of the N_{acc} annual cycle, or at least an important contributor to atmospheric lifetime of accumulation-mode particles in the autumn. However, a more complex picture emerges when analysing the fitting parameters, a (slope) and b (intercept), in the following regression - Eq. (1):

$$\log(N_{acc}) = a \cdot X_{traj} + b, \tag{1}$$

where X_{traj} denotes the median value of a trajectory-derived parameter, such as P_{traj}, or any of the others that are shown in Figure 10.

The slope a in Eq. (1) is the *relative* change in N_{acc} for every additional unit of X_{traj}, i.e. *dlog(N_{acc})/dX$_{traj}$*. Panels (a) through (d) in Figure 11 displays the values of a (adjusted to make the units more intuitive) for the corresponding parameters shown in Figure 10. Only statistically significant values (P<0.05) are shown.

Negative values in Figure 11a indicate that N_{acc} is reduced with increasing precipitation – as expected if wet removal is the main process controlling the concentration of the accumulation-mode particles. While this claim appears to be valid for Zeppelin for most of the year (8 out of 12 months), it seems to be the opposite for Tiksi, i.e. N_{acc} rises with increased trajectory precipitation. It is certainly not a causal relation – it just means that samples with greater N_{acc} are associated with higher P_{traj}. In a storm, for instance, transport and precipitation may be coincidental. Even if the scavenging is reasonably efficient, the residual transported aerosol may be sufficient to give a positive association of N_{acc} and P_{traj}. Alternatively, the addition of the particles to the sampled air mass may sometimes take place *after* most precipitation fell.

The other sites, however, show a weak negative tendency, but the relative effect of wet deposition is lowest between June and October – when it precipitates the most. This may be surprising, but it could be partly due to the high absolute uncertainty in the individual P_{traj} values– especially in the summer months when there are more convective clouds, which are not resolved by the coarse meteorological fields. However, higher median summertime P_{traj} suggests a greater potential for aerosol wet scavenging on a regional scale.

Panels (b), (c) and (d) in Figure 10 and Figure 11 explore other trajectory-derived potential candidates for playing a key role in the N_{acc} annual cycle and or its monthly median concentrations. X_{traj} in Eq. 1 is replaced by the time the trajectory spent in cloud (T_{cloud}), time in the planetary boundary layer (T_{PBL}) and the trajectory length (*Dist*) in panels (b), (c) and (d), respectively.

Figure 10b indicates that T_{cloud} is minimal for all sites around June. This may not seem in line with the observations of increased summertime cloudiness in the Arctic. But the vertical dimension should be accounted for in such a comparison as well as the fact that parts of the trajectories may be outside the Arctic region. There are additional causes for the apparent discrepancy between satellite- and reanalysis-derived annual cycle of Arctic cloudiness which are explained in Chernokulsky et al. (2012). Nevertheless, the results of the trajectory analysis shown in Figure 10b, suggest that in the dark season, there is more available time for cloud processing, which acts to increase the aerosol sizes while not considerably affecting their number concentrations in non-precipitating clouds.

The opposite phases of P_{traj} and T_{cloud} (cf. panels (a) and (b) in Figure 10) indicate that in the summer and spring months, more precipitation falls for every hour spent in a cloud, compared to the dark season. According to Figure 11b, the little time spent in precipitating clouds between June and October, still tends to lower N_{acc}, but in the dark season the extended cloud processing and low precipitation rates are associated with slightly increased N_{acc}

5   concentrations.

Figure 10c shows the annual cycle of the median T_{PBL}. The rationale is that most of the sources of the aerosols and their precursors are within the PBL, but higher T_{PBL} are not more common in the polluted season hence cannot explain the annual cycle of N_{acc}. Actually in Tiksi, and to a lesser extent Alert, there is a greater PBL influence in the summer/spring months – rather than the opposite. However, the relative change in N_{acc} for every additional

10   hour spent in the PBL (Figure 11c) is negative for most of the sites and especially between May and October. This is an indication that the PBL is a net sink for N_{acc} during these months – probably mainly due to wet removal. Another trajectory-derived parameter is *Dist*, which is also a measure of mean wind speed along the 10-day trajectory. Figure 10d Figure 10indicates that this distance is minimal in late spring/early summer between May and July for all sites and maximal in early winter December-January (Tiksi's cycle is less pronounced). This

15   means that in early winter there is potential for transporting aerosols from greater distances compared to late spring/early summer. Figure 11d does not provide a consistent picture of whether greater distances are linked with lower or higher concentrations of accumulation-mode particles on a monthly scale. This is because it does not contain information from where exactly the air is arriving during the different seasons and whether it travels over the source regions of the accumulation-mode particles. This information is shown in Figure 12.

20   The top maps in Figure 12 focus on the differences between the main summer and winter months. It shows that in June and July (the pink shade) the sampled air at all sites is rather restricted to the Arctic Ocean, with flow over the Canadian Arctic and Greenland that reaches ~~only~~ mostly Alert, Station Nord and Zeppelin. There is an occasional intrusion of North Pacific maritime air through the Bering Strait in the summer months. In December and January, however, the trajectory areal coverage is greater for Alert, Station Nord and Zeppelin, and all sites

25   indicate an expansion into the Asian side of the Arctic, with Zeppelin having some north European influence. This means that in wintertime, as the polar vortex expands southward, the 10-day trajectories are much more likely to travel over regions with anthropogenic influence compared to the summer months, when the Arctic is more isolated. It is worth noting this analysis includes also the trajectories for the data that was suspected to be affected by local pollution. This is to minimize any potential bias caused by the relationship between the large scale flow

30   and the local wind direction, which was used to filter the data.

The bottom maps in Figure 12 display the trajectory coverage in spring versus autumn, but the differences are less pronounced compared ~~with respect~~ to the top maps. This is because the annual cycle of the median trajectory length has its extremes in winter and summer (Figure 10d). However, with respect to the geographical coverage, the spring trajectories are more likely to arrive from lower latitudes than in autumn, and hence still have the

35   potential to carry more accumulation mode particles. In effect, the north does not experience a spring but rather winter, short summer and a short autumn.

By only interpreting Figure 12 and comparing the blue shaded areas in the top and bottom panels, one might expect that winter concentrations of N_{acc} in the Arctic would be higher in winter than is spring, as the aerial coverage of the winter trajectories contains more potential sources. However, Figure 3 shows that t~~T~~he median

40   N_{acc} continues to rise from winter to spring at all sites (Figure 3) despite the slight reduction in the Asian influence,

causing a "phase shift" between "maximum" transport (December-January) and highest annual median N_{acc} (April). This suggests that transport ~~alone~~ cannot ~~explain~~ control the N_{acc} cycle in the Arctic alone. To explain this conceptually, ~~However, a simple conceptual "box" model that considers~~ the Arctic dome, bounded by the Polar vortex, can be regarded as an aerosol reservoir (or a mixing chamber). The same mass of air comes in and out of the dome due to continuity, but if more particles enters the dome (both from the outside as well as those produced/emitted inside) compared with the number of particles that exit the dome and removed within it, the particle concentration inside the dome would increase. This seems to be the case between January and April; although the Arctic dome is shrinking and thus excluding some of the potential anthropogenic sources from the dome, the *net N_{acc}* flux into the Arctic is still positive due to the low removal rate. Around May, when most of the winter and early spring N_{acc} sources are outside of the shrinking Arctic dome, and when at the same time wet scavenging increases (Figure 10), the regional aerosol concentrations within the dome start to decrease.

Another point to consider is the mixing between the transported air mass and its surroundings during the transport. This is required when running single-particle trajectories for obtaining a better representation of reality. In that case, the mixing is the reason why two identical trajectories that do not pass near any N_{acc} sources, and even without any clouds or precipitation along them, are expected to have higher N_{acc} in the Arctic in April versus October or January. Without profound aerosol dynamics, as the trajectories extend in time, the initial particle concentration is less important, and the mixing with the surrounding air plays an increasingly important role. This is most strongly indicated at the high Arctic sites of Alert and Nord, by the relatively small inter-quartile variability and high 25th percentile of N_{acc} in Figure 3.

The analysis discussed in this section explores the effects of various trajectory-derived global parameters on the observed N_{acc} at the end-point of the trajectory one by one. It may be possible to perform a more detailed study that accounts for the various processes simultaneously and iteratively with the same dataset, and it would be beneficial to use one or more observed aerosol distributions along the trajectory. For instance, an analysis of trajectories passing Alert and then Nord and or Zeppelin would potentially reveal or support some of the discussed results. This might be done in a follow-up study.

The results reported above suggest that on a larger scale the patterns of the airflow (transport and mixing), seasonal precipitation and their link to aerosol processes and source regions are the main drivers of the pronounced annual cycle of the accumulation-mode aerosols observed in the Arctic. Wet removal plays an increasingly important role in reducing N_{acc}, regionally and locally, from spring to autumn. However, the small amount of precipitation in winter and early spring, the long time that the air spends in clouds and the extension of the polar vortex, allow average regional concentrations of the accumulation-mode particles to increase during this period.

It is also worth noting that the typical lifetime of the accumulation-mode aerosols is on average shorter in summer and autumn with respect to winter and spring due to the increased accumulated precipitation along a ten-day back-trajectory (Figure 10a). This means that in summer/autumn a shorter period than 240-hr could be used for the back-trajectory calculation to determine the areal extent with a potential to considerably affect the aerosol size distribution. In winter/spring more than 240 hours may be needed. This would make the differences between the shaded areas in Figure 12 even more pronounced. This further emphasises the important role of the large-scale flow in determining the properties of the aerosols throughout the year.

4   Summary and Conclusions

This paper discusses the annual cycle of several physical aerosol properties at five sites around the Arctic Ocean, with the focus on the accumulation-mode particles. In order to perform such task and to allow comparison among the sites, multi-year observations of the aerosol number size distributions were collected and compiled and the dataset was quality-controlled and homogenized with hourly aerosol concentrations that covers the diameter range of 20 to 500 nm in 29 bins.

The cluster analysis revealed four distinct aerosol number size distributions that were reconcilable with a mixture of aerosol dynamical processes and long-range transport and removal processes. One of these was a number size distribution characteristic of a recent new particle formation event in relatively clean air. This prevails mostly in June to August and may occur at higher elevation before descending into the surface-based observation site. These four modes of aerosol constitute an important diagnostic that climate models need to reproduce if they are accurately representing climate active aerosols.

On the large scale, all sites showed pronounced annual cycles with common features. The total aerosol surface area, volume and accumulation-mode concentration increased through the winter, peaked in April and decreased to a minimum around autumn. This is due to the interplay of a number of factors, with the large-scale flow, wet removal by rain and snow and cloud processing being the most important. The total number of aerosols experienced a second peak in late summer in addition to the spring peak. This was due to rather common regional events of new particle formation, which produced small particles in relatively high numbers. These events are commonly related to incoming shortwave radiation, low condensation sink and the accumulation of precursor gases prior to the nucleation event. This mostly took place over the Arctic Ocean and North Atlantic Ocean with some indication of association with coastlines and the ice edge.

The prevailing weather and the associated air-mass trajectories on a temporal scale of about a week determine the aerosol monthly characteristics in a specific year. This is because it affects the environmental conditions such as the ice extent, the regional heat and moisture fluxes and the biological activity, and hence the aerosols and the clouds, but there are consistent differences among the sites that are beyond the year-to-year variability. They are related to the location of the measurement site. For example, Alert is affected by frequent katabatic winds from nearby hills, which during winter/spring frequently draw free tropospheric air to the site. In a stable and stratified Arctic troposphere, this could mean the observation of very different aerosols compared to the regional surface based-inversion layer of the Arctic air mass.

Another example is the fact that the Zeppelin site is more than 400 m high above the nearby sea, which is open year round – unlike the sea near the other sites (Figure 1). This results in a difference in the cloud and precipitation patterns and hence the lowest background concentrations of accumulation-mode particles in April, compared to the other sites included in this study. On the other hand, its proximity to Europe (and rather frequent ship traffic) makes it more likely for some pollution to make it to the site during the cleaner months of summer and autumn. There is no single site that can be considered as fully representative for the entire Arctic region with respect to aerosol number concentrations and distributions. It is therefore important to understand which processes cause the differences between the sites and to test how well state-of-the-art aerosol models can capture these effects. The multi-site aerosol dataset could also be used for looking into how the aerosol number size distribution is altered when the air flows from one site to another.

It is expected that with a continued trend of reduction in Arctic sea ice, the emissions of biogenic sulphur gases that are aerosol precursors and hence affect aerosol growth and formation would increase in summer. This would alter the CCN properties and thus the clouds in the region. It is not clear whether this would result in a positive (Levasseur, 2013) or a negative (Gabric et al., 2005) feedback to the ongoing Arctic warming. This is because of the complex interactions and feedbacks between the aerosols, the clouds, the longwave and shortwave radiation, the ocean dynamics, the biota and the environment (Browse et al., 2014). Also, the potential for increased shipping emissions and other Arctic industrialization will make this highly uncertain. Eventually, an improved understanding of these interactions would reduce the uncertainties in future projections of Arctic climate changes and its implications for the rest of the world.

Appendix A

An observed aerosol number size distribution can often be well-approximated as the sum of three lognormal distributions. Each lognormal distribution is described with three parameters: N, σ and $\overline{D_p}$ [nm], which denote the particle concentration [cm^{-3}], the geometric standard deviation [dimensionless] and the mean geometric diameter [nm], respectively.

The sum of the three log-normal distributions is calculated in the following way-:

$$n_N^\circ(\log D) = \sum_{i=1}^{3} \frac{N_i}{(2\pi)^{1/2} \log \sigma_i} exp\left(-\frac{(\log D_p - \log \overline{D}_{pi})^2}{2 \log^2 \sigma_i}\right) \qquad (A1)$$

The tables below provide the fitting parameters of the three log-normal representations of the monthly median aerosol number distribution for each of the Arctic sites and each of the months, shown in Figure 4. It also provides the relative mean absolute error, which is the mean absolute deviation of the estimate from the observation for the 29 bins, normalized by the total number concentration. The values are in the order of 0.1%, which indicates that the sum of the three log-normal distribution fits the median monthly observations very well. The fitting parameters for the 10th, 25th, 75th and 90th percentiles are given in the online supporting material.

ALERT	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$\log \sigma$	$\overline{D_p}$	N	$\log \sigma$	$\overline{D_p}$	N	$\log \sigma$	$\overline{D_p}$	
Jan	2.84	0.127	30	22.83	0.220	80	62.09	0.151	184	0.05
Feb	2.50	0.139	30	19.40	0.202	69	71.18	0.161	172	0.06
Mar	2.96	0.116	30	39.05	0.243	77	79.14	0.167	179	0.04
Apr	2.50	0.115	30	44.10	0.239	78	112.44	0.163	178	0.06
May	2.50	0.121	30	7.49	0.183	50	72.15	0.189	145	0.08
Jun	12.03	0.145	30	2.82	0.114	50	60.58	0.198	102	0.04
Jul	2.50	0.243	30	23.67	0.167	43	38.06	0.185	100	0.04
Aug	19.79	0.165	30	21.26	0.241	73	27.39	0.124	126	0.06
Sep	2.50	0.088	26	10.91	0.243	40	13.48	0.172	121	0.10
Oct	2.80	0.149	30	4.26	0.158	58	24.74	0.162	141	0.07

	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$log\ \sigma$	$\overline{D_p}$	N	$log\ \sigma$	$\overline{D_p}$	N	$log\ \sigma$	$\overline{D_p}$	
Nov	2.50	0.134	30	20.55	0.226	70	57.03	0.176	173	0.05
Dec	2.50	0.131	30	17.70	0.226	67	48.22	0.176	161	0.08

BARROW	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$log\ \sigma$	$\overline{D_p}$	N	$log\ \sigma$	$\overline{D_p}$	N	$log\ \sigma$	$\overline{D_p}$	
Jan	8.51	0.165	30	9.73	0.109	49	169.03	0.231	152	0.03
Feb	13.03	0.181	30	6.22	0.119	49	215.85	0.248	143	0.02
Mar	19.42	0.177	30	7.48	0.110	49	214.69	0.230	149	0.03
Apr	7.94	0.142	29	6.47	0.090	47	221.09	0.212	153	0.03
May	5.00	0.119	30	5.61	0.075	43	266.19	0.203	151	0.10
Jun	17.42	0.154	26	16.05	0.106	41	91.39	0.207	114	0.08
Jul	46.40	0.142	29	53.78	0.136	53	107.68	0.232	100	0.07
Aug	59.71	0.154	29	24.57	0.073	40	87.55	0.245	100	0.12
Sep	12.27	0.150	30	5.82	0.126	80	64.17	0.197	113	0.17
Oct	6.91	0.138	28	5.00	0.165	80	57.97	0.233	133	0.15
Nov	13.78	0.183	30	5.00	0.243	80	122.54	0.221	156	0.09
Dec	7.65	0.139	28	5.00	0.096	45	198.04	0.243	157	0.02

STATION NORD	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$log\ \sigma$	$\overline{D_p}$	N	$log\ \sigma$	$\overline{D_p}$	N	$log\ \sigma$	$\overline{D_p}$	
Jan	16.03	0.229	30	15.78	0.172	60	69.03	0.206	168	0.08
Feb	17.16	0.195	30	12.42	0.152	55	77.98	0.206	158	0.04
Mar	20.32	0.208	30	29.63	0.215	66	102.10	0.196	167	0.05
Apr	5.00	0.243	24	47.60	0.226	45	176.62	0.203	162	0.04
May	5.90	0.157	30	35.43	0.205	41	99.10	0.220	145	0.07
Jun	39.59	0.200	28	5.00	0.105	48	49.82	0.237	101	0.07
Jul	34.46	0.243	20	53.31	0.177	41	43.16	0.231	100	0.07
Aug	89.76	0.243	23	20.79	0.113	41	31.29	0.213	100	0.04
Sep	42.77	0.243	25	16.17	0.243	80	18.66	0.169	127	0.08
Oct	13.25	0.179	27	5.88	0.122	46	38.39	0.192	124	0.11
Nov	12.97	0.232	30	23.39	0.243	61	65.96	0.182	171	0.03
Dec	15.56	0.238	30	19.29	0.243	80	50.81	0.185	175	0.07

TIKSI	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$log\ \sigma_g$	D_g	N	$log\ \sigma_g$	D_g	N	$log\ \sigma_g$	D_g	
Jan	11.86	0.243	26	77.38	0.243	76	73.83	0.185	206	0.04
Feb	15.07	0.243	20	65.60	0.208	57	130.62	0.219	162	0.03
Mar	37.80	0.243	30	92.32	0.243	78	123.74	0.187	180	0.06
Apr	21.84	0.172	30	82.14	0.243	65	151.61	0.179	170	0.09

	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$log\ \sigma_g$	D_g	N	$log\ \sigma_g$	D_g	N	$log\ \sigma_g$	D_g	
May	15.00	0.181	30	35.22	0.243	79	60.09	0.186	180	0.05
Jun	57.84	0.138	30	19.98	0.097	45	128.78	0.222	100	0.09
Jul	67.03	0.243	30	42.35	0.146	41	43.20	0.184	116	0.07
Aug	5.00	0.107	30	106.68	0.228	45	59.06	0.174	153	0.10
Sep	31.39	0.136	30	51.81	0.131	50	64.08	0.197	152	0.09
Oct	10.68	0.155	30	75.09	0.243	55	62.50	0.155	190	0.10
Nov	23.54	0.203	30	48.27	0.243	77	57.89	0.180	196	0.06
Dec	15.24	0.182	30	36.75	0.172	61	81.43	0.201	186	0.05

ZEPPELIN	Mode 1			Mode 2			Mode 3			Relative MAE [%]
	N	$log\ \sigma_g$	D_g	N	$log\ \sigma_g$	D_g	N	$log\ \sigma_g$	D_g	
Jan	6.47	0.138	28	13.40	0.243	80	36.36	0.152	213	0.21
Feb	10.14	0.153	29	19.13	0.243	80	51.38	0.143	197	0.16
Mar	9.56	0.154	28	27.62	0.243	80	71.10	0.149	196	0.13
Apr	20.07	0.179	30	38.81	0.243	74	82.21	0.171	185	0.07
May	58.18	0.243	20	46.58	0.172	41	106.84	0.210	155	0.09
Jun	80.38	0.225	30	13.66	0.103	40	82.45	0.215	125	0.04
Jul	79.09	0.184	30	24.25	0.109	50	84.46	0.228	111	0.06
Aug	74.89	0.232	30	23.17	0.139	40	45.48	0.201	114	0.07
Sep	15.90	0.168	22	19.55	0.161	40	31.34	0.215	110	0.05
Oct	8.35	0.168	29	5.00	0.147	54	24.87	0.195	148	0.07
Nov	9.17	0.179	29	8.96	0.189	80	29.11	0.193	188	0.10
Dec	7.53	0.164	27	12.37	0.243	80	27.13	0.148	211	0.15

We would also like to express our appreciation and gratitude for the work and effort of all the scientists and
5   engineers involved in setting up and maintaining the Arctic aerosol sites, as well as making the data available and providing useful documentation.

References

Acosta Navarro, J. C., Varma, V., Riipinen, I., Seland, Ø., Kirkevåg, A., Struthers, H., Iversen, T., Hansson, H.-C. and Ekman, A. M. L.: Amplification of Arctic warming by past air pollution reductions in Europe, Nat. Geosci., advance on, doi:10.1038/ngeo2673, 2016.

5 Arthur, D. and Vassilvitskii, S.: k-means++: the advantages of careful seeding, in Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027–1035, Society for Industrial and Applied Mathematics, New Orleans, Louisiana., 2007.

Asmi, E., Kondratyev, V., Brus, D., Laurila, T., Lihavainen, H., Backman, J., Vakkari, V., Aurela, M., Hatakka, J., Viisanen, Y., Uttal, T., Ivakhov, V. and Makshtas, A.: Aerosol size distribution seasonal characteristics

10 measured in Tiksi, Russian Arctic, Atmos. Chem. Phys., 16(3), 1271–1287, doi:10.5194/acp-16-1271-2016, 2016.

Baltensperger, U., Barrie, L., Fröhlich, C., Gras, J., Jäger, H., Jennings, S. G., Li, S.-M., Ogren, J., Wiedensohler, A., Wehrli, C. and Wilson, J.: Aerosol Measurement Procedures: Guidelines and Recommendations. Global Atmosphere Watch Report No. 153, Geneve, Switzerland., 2003.

Barrie, L.: Arctic air pollution: An overview of current knowledge, Atmos. Environ., 20(4), 643–663, 1986.

15 Barrie, L. A., Bottenheim, J. W., Schnell, R. C., Crutzen, P. J. and Rasmussen, R. A.: Ozone destruction and photochemical reactions at polar sunrise in the lower Arctic atmosphere, Nature, 334(6178), 138–141, doi:10.1038/334138a0, 1988.

Barrie, L. A., Olson, M. P. and Oikawa, K. K.: The flux of anthropogenic sulphur into the arctic from mid-latitudes in 1979/80, Atmos. Environ., 23(11), 2505–2512, doi:10.1016/0004-6981(89)90262-X, 1989.

20 Barrie, L. A., Staebler, R., Toom, D., Georgi, B., den Hartog, G., Landsberger, S. and Wu, D.: Arctic aerosol size-segregated chemical observations in relation to ozone depletion during Polar Sunrise Experiment 1992, J. Geophys. Res., 99(D12), 25439–25451, doi:10.1029/94JD01514, 1994.

Beddows, D. C. S., Dall'Osto, M. and Harrison, R. M.: Cluster Analysis of Rural, Urban, and Curbside Atmospheric Particle Size Data, Environ. Sci. Technol., 43(13), 4694–4700, doi:10.1021/es803121t, 2009.

25 Beine, H. J., Argentini, S., Maurizi, A., Viola, A. and Mastrantonio, G.: The local wind field at Ny-Ålesund and the Zeppelin mountain at Svalbard, Meteorol. Atmos. Phys., 78(1–2), 107–113, doi:10.1007/s007030170009, 2001.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S. K., Sherwood, S., Stevens, B. and Zhang, X. Y.: Clouds and Aerosols,

30 in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Browse, J., Carslaw, K. S., Mann, G. W., Birch, C. E., Arnold, S. R. and Leck, C.: The complex response of Arctic

35 aerosol to sea-ice retreat, Atmos. Chem. Phys., 14(14), 7543–7557, doi:10.5194/acp-14-7543-2014, 2014.

Chernokulsky, A., Mokhov, I. I., Chernokulsky, A. and Mokhov, I. I.: Climatology of Total Cloudiness in the Arctic: An Intercomparison of Observations and Reanalyses, Adv. Meteorol., 2012, 1–15, doi:10.1155/2012/542093, 2012.

Christensen, J. H.: The Danish eulerian hemispheric model — a three-dimensional air pollution model used for

40 the arctic, Atmos. Environ., 31(24), 4169–4191, doi:10.1016/S1352-2310(97)00264-1, 1997.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J. and Jones, J.: Recent Arctic amplification and extreme mid-latitude weather, Nat. Geosci., 7(9), 627–637, doi:10.1038/ngeo2234, 2014.

Covert, D. S., Wiedensohler, A., Aalto, P., Heintzenberg, J., McMurry, P. H. and Leck, C.: Aerosol number size distributions from 3 to 500 nm diameter in the arctic marine boundary layer during summer and autumn, Tellus B, 48(2), 197–212, doi:10.1034/j.1600-0889.1996.t01-1-00005.x, 1996.

Croft, B., Wentworth, G. R., Martin, R. V., Leaitch, W. R., Murphy, J. G., Murphy, B. N., Kodros, J. K., Abbatt, J. P. D. and Pierce, J. R.: Contribution of Arctic seabird-colony ammonia to atmospheric particles and cloud-albedo radiative effect, Nat. Commun., 7, 13444, doi:10.1038/ncomms13444, 2016a.

Croft, B., Martin, R. V., Leaitch, W. R., Tunved, P., Breider, T. J., Andrea, S. D. and Pierce, J. R.: Processes controlling the annual cycle of Arctic aerosol number and size distributions, Atmos. Chem. Phys., 16(6), 3665–3682, doi:10.5194/acp-16-3665-2016, 2016b.

Dal Maso, M., Kulmala, M., Riipinen, I., Wagner, R., Hussein, T., Aalto, P. P. and Lehtinen, K. E. J.: Formation and growth of fresh atmospheric aerosols: eight years of aerosol size distribution data from SMEAR II, Hyytiala, Finland, Boreal Environ. Res., 10(5), 323, 2005.

Draxler, R. R. and Hess, G. D.: An overview of the HYSPLIT_4 modelling system for trajectories, Aust. Meteorol. Mag., 47(4), 295–308, 1998.

Engelstaedter, S. and Washington, R.: Temporal controls on global dust emissions: The role of surface gustiness, Geophys. Res. Lett., 34(15), L15805, doi:10.1029/2007GL029971, 2007.

Flyger, H., Heidam, N. Z., Hansen, K. A., Rasmussen, L. and Megaw, W. J.: The background levels of the summer tropospheric aerosol and trace gases in Greenland, J. Aerosol Sci., 11(1), 95–110, doi:10.1016/0021-8502(80)90149-4, 1980.

Frick, G. M. and Hoppel, W. A.: Airship Measurements of Aerosol Size Distributions, Cloud Droplet Spectra, and Trace Gas Concentrations in the Marine Boundary Layer, Bull. Am. Meteorol. Soc., 74(11), 2195–2202, doi:10.1175/1520-0477(1993)074<2195:AMOASD>2.0.CO;2, 1993.

Gabric, A. J., Qu, B., Matrai, P. and Hirst, A. C.: The simulated response of dimethylsulfide production in the Arctic Ocean to global warming, Tellus B, 57(5), 391–403, doi:10.1111/j.1600-0889.2005.00163.x, 2005.

Garrett, T. J. and Verzella, L. L.: An Evolving History of Arctic Aerosols, Bull. Am. Meteorol. Soc., 89(3), 299–302, doi:10.1175/BAMS-89-3-299, 2008.

Glantz, P., Bourassa, A., Herber, A., Iversen, T., Karlsson, J., Kirkevåg, A., Maturilli, M., Seland, Ø., Stebel, K., Struthers, H., Tesche, M. and Thomason, L.: Remote sensing of aerosols in the Arctic for an evaluation of global climate model simulations, J. Geophys. Res. Atmos., 119(13), 8169–8188, doi:10.1002/2013JD021279, 2014.

Goodsite, M., Skov, H., Asmund, G., Bennike, O., Feilberg, A., Glasius, M., Gross, A. and Hermanson, M. H.: Pilot Study of Contaminants near Station Nord, a Military Airbase and Research Station in NE Greenland, in NATO Science for Peace and Security Series C: Environmental Security, edited by I. Linkov, pp. 177–198, Springer, Dordrecht., 2014.

Hansen, A. M. K., Kristensen, K., Nguyen, Q. T., Zare, A., Cozzi, F., Nøjgaard, J. K., Skov, H., Brandt, J., Christensen, J. H., Ström, J., Tunved, P., Krejci, R. and Glasius, M.: Organosulfates and organic acids in Arctic aerosols: speciation, annual variation and concentration levels, Atmos. Chem. Phys., 14(15), 7807–7823, doi:10.5194/acp-14-7807-2014, 2014.

Heidam, N. Z. and Z., N.: The components of the Arctic aerosol, Atmos. Environ., 18(2), 329–343, doi:10.1016/0004-6981(84)90107-0, 1984.

Heidam, N. Z., Wåhlin, P. and Christensen, J. H.: Tropospheric Gases and Aerosols in Northeast Greenland, J. Atmos. Sci., 56(2), 261–278, doi:10.1175/1520-0469(1999)056<0261:TGAAIN>2.0.CO;2, 1999.

5 Heidam, N. Z., Christensen, J., Wåhlin, P. and Skov, H.: Arctic atmospheric contaminants in NE Greenland: levels, variations, origins, transport, transformations and trends 1990-2001., Sci. Total Environ., 331(1–3), 5–28, doi:10.1016/j.scitotenv.2004.03.033, 2004.

Hirdman, D., Burkhart, J. F., Sodemann, H., Eckhardt, S., Jefferson, a., Quinn, P. K., Sharma, S., Ström, J. and Stohl, a.: Long-term trends of black carbon and sulphate aerosol in the Arctic: changes in atmospheric transport

10 and source region emissions, Atmos. Chem. Phys., 10(19), 9351–9368, doi:10.5194/acp-10-9351-2010, 2010a.

Hirdman, D., Sodemann, H., Eckhardt, S., Burkhart, J. F., Jefferson, a., Mefford, T., Quinn, P. K., Sharma, S., Ström, J. and Stohl, a.: Source identification of short-lived air pollutants in the Arctic using statistical analysis of measurement data and particle dispersion model output, Atmos. Chem. Phys., 10(2), 669–693, doi:10.5194/acp-10-669-2010, 2010b.

15 Hoppel, W. A., Frick, G. M. and Larson, R. E.: Effect of nonprecipitating clouds on the aerosol size distribution in the marine boundary layer, Geophys. Res. Lett., 13(2), 125–128, doi:10.1029/GL013i002p00125, 1986.

Huang, K., Fu, J. S., Prikhodko, V. Y., Storey, J. M., Romanov, A., Hodson, E. L., Cresko, J., Morozova, I., Ignatieva, Y. and Cabaniss, J.: Russian anthropogenic black carbon: Emission reconstruction and Arctic black carbon simulation, J. Geophys. Res. Atmos., 120(21), 11,306-11,333, doi:10.1002/2015JD023358, 2015a.

20 Huang, Z., Huang, J., Hayasaka, T., Wang, S., Zhou, T. and Jin, H.: Short-cut transport path for Asian dust directly to the Arctic: a case study, Environ. Res. Lett., 10(11), 114018, doi:10.1088/1748-9326/10/11/114018, 2015b.

Jaenicke, R. and Davies, C. N.: The mathematical expression of the size distribution of atmospheric aerosols, J. Aerosol Sci., 7(3), 255–259, doi:10.1016/0021-8502(76)90040-9, 1976.

Ketchen, D. J. and Shook, C. L.: The Application of cluster analysis in strategic management research: an analysis

25 and critique, Strateg. Manag. J., 17(6), 441–458, doi:10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G, 1996.

Kirkby, J., Curtius, J., Almeida, J., Dunne, E., Duplissy, J., Ehrhart, S., Franchin, A., Gagné, S., Ickes, L., Kürten, A., Kupc, A., Metzger, A., Riccobono, F., Rondo, L., Schobesberger, S., Tsagkogeorgas, G., Wimmer, D., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Downard, A., Ehn, M., Flagan, R. C.,

30 Haider, S., Hansel, A., Hauser, D., Jud, W., Junninen, H., Kreissl, F., Kvashin, A., Laaksonen, A., Lehtipalo, K., Lima, J., Lovejoy, E. R., Makhmutov, V., Mathot, S., Mikkilä, J., Minginette, P., Mogo, S., Nieminen, T., Onnela, A., Pereira, P., Petäjä, T., Schnitzhofer, R., Seinfeld, J. H., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Vanhanen, J., Viisanen, Y., Vrtala, A., Wagner, P. E., Walther, H., Weingartner, E., Wex, H., Winkler, P. M., Carslaw, K. S., Worsnop, D. R., Baltensperger, U. and Kulmala, M.: Role of sulphuric acid, ammonia and galactic

35 cosmic rays in atmospheric aerosol nucleation., Nature, 476(7361), 429–33, doi:10.1038/nature10343, 2011.

Kolesar, K. R., Cellini, J., Peterson, P. K., Jefferson, A., Tuch, T., Birmili, W., Wiedensohler, A. and Pratt, K. A.: Effect of Prudhoe Bay emissions on atmospheric aerosol growth events observed in Utqiaġvik (Barrow), Alaska, Atmos. Environ., 152, 146–155, doi:10.1016/j.atmosenv.2016.12.019, 2017.

Kulmala, M., Dal Maso, M., Mäkelä, J. M., Pirjola, L., Väkevä, M., Aalto, P., Miikkulainen, P., Hämeri, K. and

40 O'Dowd, C. D.: On the formation, growth and composition of nucleation mode particles, Tellus B, 53(4), 479–

490, 2001.

Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V.-M., Birmili, W. and McMurry, P. H.: Formation and growth rates of ultrafine atmospheric particles: a review of observations, J. Aerosol Sci., 35(2), 143–176, doi:10.1016/j.jaerosci.2003.10.003, 2004.

5   Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., Dachs, J., Bopp, L., Saltzman, E. S., Stefels, J., Johnson, J. E. and Liss, P. S.: An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean, Global Biogeochem. Cycles, 25(1), GB1004, doi:10.1029/2010GB003850, 2011.

Leaitch, W. R., Sharma, S., Huang, L., Toom-Sauntry, D., Chivulescu, A., Macdonald, A. M., von Salzen, K.,
10   Pierce, J. R., Bertram, A. K., Schroder, J. C., Shantz, N. C., Chang, R. Y.-W. and Norman, A.-L.: Dimethyl sulfide control of the clean summertime Arctic aerosol and cloud, Elem. Sci. Anthr., 1(1), 17, doi:10.12952/journal.elementa.000017, 2013.

Leck, C., Norman, M., Bigg, E. K. and Hillamo, R.: Chemical composition and sources of the high Arctic aerosol relevant for cloud formation, J. Geophys. Res., 107(D12), 4135, doi:10.1029/2001JD001463, 2002.

15   Levasseur, M.: Impact of Arctic meltdown on the microbial cycling of sulphur, Nat. Geosci., 6(9), 691–700, doi:10.1038/ngeo1910, 2013.

Liu, D., Quennehen, B., Darbyshire, E., Allan, J. D., Williams, P. I., Taylor, J. W., J.-B. Bauguitte, S., Flynn, M. J., Lowe, D., Gallagher, M. W., Bower, K. N., Choularton, T. W. and Coe, H.: The importance of Asia as a source of black carbon to the European Arctic during springtime 2013, Atmos. Chem. Phys., 15(20), 11537–11555,
20   doi:10.5194/acp-15-11537-2015, 2015.

Lloyd, S.: Least squares quantization in PCM, IEEE Trans. Inf. Theory, 28(2), 129–137, doi:10.1109/TIT.1982.1056489, 1982.

Lohmann, U. and Feichter, J.: Global indirect aerosol effects: a review, Atmos. Chem. Phys., 5(3), 715–737, doi:10.5194/acp-5-715-2005, 2005.

25   Mahowald, N., Albani, S., Kok, J. F., Engelstaeder, S., Scanza, R., Ward, D. S. and Flanner, M. G.: The size distribution of desert dust aerosols and its impact on the Earth system, Aeolian Res., 15, 53–71, doi:10.1016/j.aeolia.2013.09.002, 2014.

Markowicz, K. M., Pakszys, P., Ritter, C., Zielinski, T., Udisti, R., Cappelletti, D., Mazzola, M., Shiobara, M., Xian, P., Zawadzka, O., Lisok, J., Petelski, T., Makuch, P. and Karasiński, G.: Impact of North American intense
30   fires on aerosol optical properties measured over the European Arctic in July 2015, J. Geophys. Res. Atmos., 121(24), 14,414-487,512, doi:10.1002/2016JD025310, 2016.

Massling, A., Nielsen, I. E., Kristensen, D., Christensen, J. H., Sørensen, L. L., Jensen, B., Nguyen, Q. T., Nøjgaard, J. K., Glasius, M. and Skov, H.: Atmospheric black carbon and sulfate concentrations in Northeast Greenland, Atmos. Chem. Phys., 15(16), 9681–9692, doi:10.5194/acp-15-9681-2015, 2015.

35   Morin, S.: A high time resolution study of boundary layer ozone chemistry and dynamics over the Arctic Ocean near Alert, Nunavut, Geophys. Res. Lett., 32(8), L08809, doi:10.1029/2004GL022098, 2005.

Moroni, B., Cappelletti, D., Crocchianti, S., Becagli, S., Caiazzo, L., Traversi, R., Udisti, R., Mazzola, M., Markowicz, K., Ritter, C. and Zielinski, T.: Morphochemical characteristics and mixing state of long range transported wildfire particles at Ny-Ålesund (Svalbard Islands), Atmos. Environ., 156, 135–145,
40   doi:10.1016/j.atmosenv.2017.02.037, 2017.

Mungall, E. L., Croft, B., Lizotte, M., Thomas, J. L., Murphy, J. G., Levasseur, M., Martin, R. V., Wentzell, J. J. B., Liggio, J. and Abbatt, J. P. D.: Dimethyl sulfide in the summertime Arctic atmosphere: measurements and source sensitivity simulations, Atmos. Chem. Phys., 16(11), 6665–6680, doi:10.5194/acp-16-6665-2016, 2016.

Najafi, M. R., Zwiers, F. W. and Gillett, N. P.: Attribution of Arctic temperature change to greenhouse-gas and aerosol influences, Nat. Clim. Chang., 5(3), 246–249, doi:10.1038/nclimate2524, 2015.

Nguyen, Q. T., Skov, H., Sørensen, L. L., Jensen, B. J., Grube, A. G., Massling, A., Glasius, M. and Nøjgaard, J. K.: Source apportionment of particles at Station Nord, North East Greenland during 2008–2010 using COPREM and PMF analysis, Atmos. Chem. Phys., 13(1), 35–49, doi:10.5194/acp-13-35-2013, 2013.

Nguyen, Q. T., Glasius, M., Sørensen, L. L., Jensen, B., Skov, H., Birmili, W., Wiedensohler, A., Kristensson, A., Nøjgaard, J. K. and Massling, A.: Seasonal variation of atmospheric particle number concentrations, new particle formation and atmospheric oxidation capacity at the high Arctic site Villum Research Station, Station Nord, Atmos. Chem. Phys., 16(17), 11319–11336, doi:10.5194/acp-16-11319-2016, 2016.

Parkinson, C. L. and Comiso, J. C.: On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm, Geophys. Res. Lett., 40(7), 1356–1361, doi:10.1002/grl.50349, 2013.

Polissar, A. V., Hopke, P. K. and Harris, J. M.: Source Regions for Atmospheric Aerosol Measured at Barrow, Alaska, Environ. Sci. Technol., 35(21), 4214–4226, doi:10.1021/es0107529, 2001.

Qi, L., Li, Q., Henze, D. K., Tseng, H.-L. and He, C.: Sources of Springtime Surface Black Carbon in the Arctic: An Adjoint Analysis, Atmos. Chem. Phys. Discuss., in review, doi:10.5194/acp-2016-1112, 2017.

Quinn, P. K., Shaw, G., Andrews, E., Dutton, E. G., Ruoho-Airola, T. and Gong, S. L.: Arctic haze: current trends and knowledge gaps, Tellus B, 59(1), 99–114, doi:10.1111/j.1600-0889.2006.00238.x, 2007.

Radke, L. and Lyons, J.: Airborne observations of Arctic aerosols. I: Characteristics of Arctic haze, Geophys. Res. Lett., 11(5), 393–396, 1984.

Rahn, K. A., Borys, R. D. and Shaw, G. E.: The Asian source of Arctic haze bands, Nature, 268(5622), 713–715, doi:10.1038/268713a0, 1977.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math., 20, 53–65, doi:10.1016/0377-0427(87)90125-7, 1987.

Ruuskanen, T. M., Kaasik, M., Aalto, P. P., Hõrrak, U., Vana, M., Mårtensson, M., Yoon, Y. J., Keronen, P., Mordas, G., Ceburnis, D., Nilsson, E. D., O'Dowd, C., Noppel, M., Alliksaar, T., Ivask, J., Sofiev, M., Prank, M. and Kulmala, M.: Concentrations and fluxes of aerosol particles during the LAPBIAT measurement campaign at Värriö field station, Atmos. Chem. Phys., 7(14), 3683–3700, doi:10.5194/acp-7-3683-2007, 2007.

Serreze, M. C. and Hurst, C. M.: Representation of Mean Arctic Precipitation from NCEP–NCAR and ERA Reanalyses, J. Clim., 13(1), 182–201, doi:10.1175/1520-0442(2000)013<0182:ROMAPF>2.0.CO;2, 2000.

Serreze, M. C. and Stroeve, J.: Arctic sea ice trends, variability and implications for seasonal ice forecasting., Philos. Trans. A. Math. Phys. Eng. Sci., 373(2045), doi:10.1098/rsta.2014.0159, 2015.

Sharma, S., Andrews, E., Barrie, L. a., Ogren, J. a. and Lavoué, D.: Variations and sources of the equivalent black carbon in the high Arctic revealed by long-term observations at Alert and Barrow: 1989-2003, J. Geophys. Res., 111, D14208, doi:10.1029/2005JD006581, 2006.

Shaw, G. E.: Eddy diffusion transport of Arctic pollution from the mid-latitudes: A preliminary model, Atmos. Environ., 15(8), 1483–1490, doi:10.1016/0004-6981(81)90356-5, 1981.

Sipilä, M., Sarnela, N., Jokinen, T., Henschel, H., Junninen, H., Kontkanen, J., Richters, S., Kangasluoma, J.,

Franchin, A., Peräkylä, O., Rissanen, M. P., Ehn, M., Vehkamäki, H., Kurten, T., Berndt, T., Petäjä, T., Worsnop, D., Ceburnis, D., Kerminen, V.-M., Kulmala, M. and O'Dowd, C.: Molecular-scale evidence of aerosol particle formation via sequential addition of HIO3, Nature, 537(7621), 532–534, doi:10.1038/nature19314, 2016.

Steffen, A., Bottenheim, J., Cole, A., Ebinghaus, R., Lawson, G. and Leaitch, W. R.: Atmospheric mercury speciation and mercury in snow over time at Alert, Canada, Atmos. Chem. Phys., 14(5), 2219–2231, doi:10.5194/acp-14-2219-2014, 2014.

Stohl, A.: Characteristics of atmospheric transport into the Arctic troposphere, J. Geophys. Res., 111, D11306, doi:10.1029/2005JD006888, 2006.

Stohl, A. and Seibert, P.: Accuracy of trajectories as determined from the conservation of meteorological tracers, Q. J. R. Meteorol. Soc., 124(549), 1465–1484, doi:10.1002/qj.49712454907, 1998.

Stohl, A., Hittenberger, M. and Wotawa, G.: Validation of the lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, Atmos. Environ., 32(24), 4245–4264, doi:10.1016/S1352-2310(98)00184-8, 1998.

Ström, J., Engvall, A.-C., Delbart, F., Krejci, R. and Treffeisen, R.: On small particles in the Arctic summer boundary layer: observations at two different heights near Ny-Ålesund, Svalbard, Tellus B, 61(2), 473–482, doi:10.1111/j.1600-0889.2008.00412.x, 2009.

Thorndike, R. L.: Who belongs in the family?, Psychometrika, 18(4), 267–276, doi:10.1007/BF02289263, 1953.

Tilling, R. L., Ridout, A., Shepherd, A. and Wingham, D. J.: Increased Arctic sea ice volume after anomalously low melting in 2013, Nat. Geosci., 8(8), 643–646, doi:10.1038/ngeo2489, 2015.

Tunved, P., Hansson, H.-C., Kulmala, M., Aalto, P., Viisanen, Y., Karlsson, H., Kristensson, A., Swietlicki, E., Dal Maso, M., Ström, J. and Komppula, M.: One year boundary layer aerosol size distribution data from five nordic background stations, Atmos. Chem. Phys., 3(6), 2183–2205, doi:10.5194/acp-3-2183-2003, 2003.

Tunved, P., Ström, J. and Krejci, R.: Arctic aerosol life cycle: linking aerosol size distributions observed between 2000 and 2010 with air mass transport and precipitation at Zeppelin station, Ny-Ålesund, Svalbard, Atmos. Chem. Phys., 13(7), 3643–3660, doi:10.5194/acp-13-3643-2013, 2013.

Wentworth, G. R., Murphy, J. G., Croft, B., Martin, R. V., Pierce, J. R., Côté, J.-S., Courchesne, I., Tremblay, J.-É., Gagnon, J., Thomas, J. L., Sharma, S., Toom-Sauntry, D., Chivulescu, A., Levasseur, M. and Abbatt, J. P. D.: Ammonia in the summertime Arctic marine boundary layer: sources, sinks and implications, Atmos. Chem. Phys. Discuss., 15(21), 29973–30016, doi:10.5194/acpd-15-29973-2015, 2015.

Wiedensohler, A., Birmili, W., Nowak, A., Sonntag, A., Weinhold, K., Merkel, M., Wehner, B., Tuch, T., Pfeifer, S., Fiebig, M., Fjäraa, A. M., Asmi, E., Sellegri, K., Depuy, R., Venzac, H., Villani, P., Laj, P., Aalto, P., Ogren, J. A., Swietlicki, E., Williams, P., Roldin, P., Quincey, P., Hüglin, C., Fierz-Schmidhauser, R., Gysel, M., Weingartner, E., Riccobono, F., Santos, S., Grüning, C., Faloon, K., Beddows, D., Harrison, R., Monahan, C., Jennings, S. G., O'Dowd, C. D., Marinoni, A., Horn, H.-G., Keck, L., Jiang, J., Scheckman, J., McMurry, P. H., Deng, Z., Zhao, C. S., Moerman, M., Henzing, B., de Leeuw, G., Löschau, G. and Bastian, S.: Mobility particle size spectrometers: harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions, Atmos. Meas. Tech., 5(3), 657–685, doi:10.5194/amt-5-657-2012, 2012.

Williams, J., de Reus, M., Krejci, R., Fischer, H. and Ström, J.: Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages, Atmos. Chem. Phys., 2(2), 133–145,

doi:10.5194/acp-2-133-2002, 2002.

Willis, M. D., Burkart, J., Thomas, J. L., Koellner, F., Schneider, J., Bozem, H., Hoor, P. M., Aliabadi, A. A., Schulz, H., Herber, A. B., Leaitch, R. and Abbatt, J. P. D.: Biogenic influence on the composition and growth of summertime Arctic aerosol, AGU Fall Meet., 2016.

5   Yu, H., Kaufman, Y. J., Chin, M., Feingold, G., Remer, L. A., Anderson, T. L., Balkanski, Y., Bellouin, N., Boucher, O., Christopher, S., DeCola, P., Kahn, R., Koch, D., Loeb, N., Reddy, M. S., Schulz, M., Takemura, T. and Zhou, M.: A review of measurement-based assessments of the aerosol direct radiative effect and forcing, Atmos. Chem. Phys., 6(3), 613–666, doi:10.5194/acp-6-613-2006, 2006.

10

List of Figures

Figure 1: A physical map of the Arctic region, with the five aerosol measurement sites marked with red squares around the Arctic ocean. The curved lines indicate the areal extent of the satellite-derived minimum and maximum ice edge in 2012 and 2013, which occurred in September and March, respectively. Here, the ice edge is defined as where the ice concentration is 50%.

Figure 2: Data availability. The bars specify the fraction of the time within each month with available aerosol data after filtering (see Sect. 2.2). The colours of the bars represent different years, as shown in the legend, and the percentages therein indicates the yearly total data coverage.

Figure 3: Annual cycle of aerosol bulk properties. The monthly median and interquartile ranges of the aerosol total and accumulation-mode number concentrations are indicated in grey dashed and solid curves and whiskers, respectively. The aerosol total surface area is in orange and the total volume in blue. The total values were calculated for the dry-diameter range of 10 to 500 nm, and the accumulation-mode size range is defined here between 100 and 500 nm. Particles were assumed spherical for the computation of the aerosol surface area and volume. The colour of the vertical axes and their labels, correspond to the colour of the curves, which are colour-blind friendly.

Figure 4: The monthly aerosol number size distributions. The solid curves indicate the median distributions of the 29-bin homogenized dataset, for each of the sites, with the colours matching the colours of the station name in the legend. The shaded semi-transparent areas denote the inter-quartile range and imply on the variability of the aerosol number size-distribution is each of the months. The horizontal and vertical black lines at 100 cm^{-3} and 100 nm, respectively, are guidelines for facilitating the comparison between the months. Error! Reference source not found.The fitting parameters for approximating these aerosol spectra as the sum of three lognormal distributions, are given in Appendix A.

Figure 5: The output of the K-means cluster analysis. The solid curves show the centroids (median) of the number size-distribution of the four clusters. The shaded area denotes the inter-quartile range of each of the size bins, for all distributions that were members of the given clusters.

Figure 6: The monthly probability of cluster occurrence. The bars indicate the mean relative frequency of the four aerosol clusters within each month (weighted by their number of occurrences in each of the included years). The colour coding is the same as in Figure 5. The percentages in the legend indicate the mean annual relative occurrence of each of the clusters, only for years that had a fair representation of all seasons (cf. Figure 2). Tiksi and Barrow were not part of the cluster analysis due to their poorer data coverage (Figure 2) and less representative wind-directional sampling (cf. Sect. 2.2).

Figure 7: The trajectory density associated with each aerosol cluster and site. The values represent the probability of the 240-hr long trajectories to cross the grid cells. Each cell is 0.5 and 4 degrees wide in latitude and longitude, respectively, in a concentric coordinate system whose pole is at the measurement site. The yellow shading indicates trajectory probability greater than 5%. The geographical mean location of the 240-hour long back-trajectories is indicated in red. They are much shorter than the individual trajectories because trajectories at one side of the site cancel out the trajectories on the opposite side in spatial averaging. Still, the mean location gives an indication of the direction from which most trajectories arrive to the site.

Figure 8: Maps of median concentrations of accumulation-mode particles. Every 240-hr long back-trajectory is associated with the concentration that was measured at the site at the air mass arrival time – according to the trajectory analysis. The colour of each grid cell denotes the median concentration related to all trajectories crossing that cell. Each subplot represents a different site, whose location is indicated in red. All available data was included for each of the sites, but only grid cells with at least 10 trajectory passes are presented for statistical robustness.

Figure 9: Like Figure 8, but the maps presents the combined the data for all sites together. Panel (a) displays the mean trajectory height above the surface, and panel (b) shows the mean concentration of accumulation mode particles. The locations of the sites are denoted by the small red circles.

Figure 10: Monthly statistics (median values) of trajectory-derived parameters for all sites. Panel (a): Accumulated precipitation along the 240-hr back-trajectory; Panel (b): Number of hours in a cloud. This was presumed when the relative humidity was greater than 95%; Panel (c): Fraction of trajectory time within the planetary boundary layer. This was assumed when its altitude above the surface was smaller than twice the model mixing layer height; Panel (d): Trajectory accumulated distance.

Figure 11: The derivative of aerosol accumulation-mode concentration with each of the trajectory-derived parameters shown in Figure 10 – for each month and site. Positive values indicate the percentage increase in accumulation-mode concentration with a unit increase in the corresponding parameter, while negative values show the opposite. Only statistically significant values (P<0.05) are shown.

Figure 12: The seasonal areal coverage of 240-hr long back-trajectories for each of the sites. The top maps compare winter and summer in blue and red, respectively, while the bottom maps compare the spring and autumn. Only grid-cells with a trajectory-crossing probability greater than 0.5% are shown.

List of Tables

Table 1: Quantitative information about the cluster centroids and members, which was the result of the k-means cluster analysis. The effective diameter, surface area, total volume and total mass are the centroid values, while the percentiles of the number concentrations are derived from all the cluster members.

	No. of Members	Effective Diameter [nm]	Surface Area [$cm^2 m^{-3}$]	Total Volume [$cm^3 m^{-3}$]	Total Mass [$\mu g\ m^{-3}$]	Number Conc. (10-500nm) [cm^{-3}]			Number Conc. (100-500nm) [cm^{-3}]		
						25th	median	75th	25th	median	75th
Cluster 1	5849	256	0.21	$8.9 \cdot 10^{-7}$	1.34	165	206	290	125	150	204
Cluster 2	12349	249	0.101	$4.18 \cdot 10^{-7}$	0.63	79	102	145	59	72	99
Cluster 3	12670	228	0.027	$1.03 \cdot 10^{-7}$	0.15	23	36	78	13	19	40
Cluster 4	3503	189	0.066	$2.09 \cdot 10^{-7}$	0.31	170	273	464	39	56	97