ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MASTER THESIS

ENVIRONMENTAL DRIVERS OF SOUTHERN OCEAN AEROSOL SIZE DISTRIBUTION

Student:
Julien Clark[1]

Supervisors:
Prof. Julia SCHMALE[2]
Dr. Jakob PERNOV[2]

[1] EPFL, Environmental Sciences and Engineering Section (SSIE)

[2] EPFL, Extreme Environments Research Laboratory (EERL)

JUNE 2022

Table of Contents

Abstract

To improve our predictions of future climates we need to understand past and present processes that influence them. In particular, characterising pre-industrial aerosol properties and formation processes has proven to be particularly difficult due to long-range transport of anthropogenic emissions reaching all other the globe. The Southern Ocean (SO) is one of the few places on Earth where pristine aerosol conditions can still be found thereby making it of prime importance for studying natural aerosol processes. To characterise the aerosol population, we carried out k-means clustering on submicron Particle Number Size Distribution (PNSD) data collected during the Antarctic Circumnavigation Expedition (ACE), and found three clusters described best its dynamics. Each cluster had a main mode, namely nucleation, Aitken, and accumulation mode. Aitken and accumulation clusters south of 60 °S showed pronounced bimodality, while the Aitken mode north of 60 °S was more unimodal, hinting to more important cloud processing near the Antarctic continent. We found in-situ meteorological data, such as wind speed or solar radiation, did not seem to influence the cluster type occurrence, confirming the importance of transport processes for local particle size distributions. Processes happening exclusively in the Marine Boundary Layer (MBL) were at least and if not more important than processes in the Free Troposphere (FT) in driving the nucleation and Aitken cluster, while airmasses coming from marginal sea ice zones were found to predominantly drive the nucleation cluster. Higher particulate sulphate and gas-phase methanesulphonic acid (MSA) concentrations were associated with higher CCN concentrations, particularly south of 60 °S, while increased wind speeds were found not to increase Cloud Condensation Nuclei (CCN) concentrations in the MBL. We build on previous results found during ACE and show the usefulness of clustering to compare PNSDs over different geographical regions.

1. Introduction

In the sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), it was noted how clouds, aerosols, and their interactions, still contribute the most to uncertainties in estimates of the changing energy budget of Earth due to climate change (Masson-Delmotte et al., 2021). Aerosols can affect the energy budget in two ways: aerosol-radiation interactions by the direct effects of scattering and absorption of solar and terrestrial radiation, and aerosol-cloud interactions, called indirect effects, by serving as CCN. CCN are particles that can become activated to grow to cloud droplets in the presence of supersaturated water vapor. Aerosol-radiation interactions have a global mean effective radiative forcing of $-0.45\ (-0.95\ \text{to}\ +0.05)\ \mathrm{W\,m^{-2}}$, for a 5 to 95 % confidence interval, while the IPCC has not given aerosol-cloud interactions effective radiative forcing due to the difficulty in separating their effect on rapid cloud adjustment. However, they give a cautious combined effect of aerosol-radiation and aerosol-cloud interactions of $-0.9\ (-1.9\ \text{to}\ -0.1)\ \mathrm{W\,m^{-2}}$ (Boucher et al., 2013). These negative radiative forcing offset part of the greenhouse gas forcing. They also give, with low confidence, an aerosol-climate feedback of $\pm 0.2\ \mathrm{W\,m^{-2}\,°C}$ due to a weak dimethylsulphide-CCN-cloud albedo feedback. One way, but not the only way, CCN can affect indirect radiative forcing is by changing cloud albedo: for fixed water content an increase in CCN concentration will result in more numerous smaller cloud droplets, which scatter more radiation compared to larger droplets thereby increasing cloud albedo (Twomey, 1977). The difficulty in estimating this aerosol indirect forcing is because changing aerosol concentration can have cascading effects on clouds. For example, smaller cloud droplets caused by increased CCN concentration could increase cloud lifetime by reducing rain formation (Albrecht, 1989). Increased cloud lifetime in turn causes other effects, such as increased night-time temperatures due to increased downward longwave radiation forcing (Huang et al., 2006), increasing uncertainty in indirect aerosol forcing. This is especially important in the polar regions which experiences 24-hour darkness in the wintertime.

To understand the evolution of the radiative forcing of aerosols due to anthropogenic influences (due to both increased sources of aerosols of anthropogenic origin and changing climate), models compare a pre-industrial reference to present-day conditions. Uncertainty in this pre-industrial reference, as information about aerosol levels before human interference is difficult to obtain, is one of the driver of high uncertainty in the models (Carslaw et al., 2013). This uncertainty can be reduced by understanding aerosol processes in current pristine conditions, which resemble pre-industrial conditions, but these regions are rare due to anthropogenic influences, particularly in the Northern Hemisphere. The SO, especially in summertime, is of particular interest because it is one of the few regions in the world that has close to pristine condition (Hamilton et al., 2014).

The SO is defined as all waters south of 60 °S but sometimes more loosely as all waters south of the Antarctic Convergence marine belt. This marine belt varies seasonally and is set at the convergence of cold surface Antarctic waters and warmer sub-Antarctic waters. These cold surface Antarctic waters are driven by a current called the Antarctic Circumpolar current which flows from East to West, driven in part by the Southern Hemisphere westerly wind belt

(SWW). Both wind and current flowing nearly unhindered due to the absence of land masses at those latitudes (Barker et al., 2007). Together, the SWW and the Antarctic circumpolar current control the upwelling of carbon rich waters coming from higher latitudes at the Antarctic Convergence marine belt due to the sinking of the colder Antarctic surface waters (Lamy et al., 2019). The upwelling of carbon rich waters drives large phytoplankton blooms in summer, significantly influencing the aerosol budget due to increased biogenic sources of precursor gases (Tortell and Long, 2009). The SWW is believed to be one of the main reasons behind the pristine conditions found in the SO, effectively serving as a barrier for anthropogenic aerosols (Uetake et al., 2020).

In pristine seasons and locations of the SO, there are believed to be two main sources of aerosols: sea spray, where particles are directly emitted from the ocean and are thus called primary aerosols, and biogenic marine emissions. Condensation of these biogenic vapours on existing particles are called secondary aerosols, while formation of new particles from gas-to-particle processes is a third process. When measuring the number of aerosols at different sizes, particle size distributions often exhibit modes at certain particle size ranges, with each mode representing different atmospheric processes. For submicron aerosols, particles can belong to three main modes: the nucleation mode, the Aitken mode, and the accumulation mode. The exact boundaries of each mode can vary in the literature depending on the context, but nucleation modes peak at particle sizes between 10 and 30 nm, Aitken modes at particle sizes between 20 and 100 nm and accumulation mode above 100 nm (Dinoi et al., 2020; Hussein et al., 2004). For supermicron particles, another mode called the coarse mode is formed by sea salt aerosol and dust. However, due to their size and therefore fast sedimentation rates, coarse mode particles have a low residence time in the atmosphere and represent only a very small fraction of CCN.

Nucleation mode particles include new particles formed by nucleation from the gas phase. This requires gaseous vapours to be supersaturated and to overcome the nucleation barrier. Sulphuric acid for example has a low vapour pressure and if there are sufficiently few pre-existing particles on which it can condense, then new particles are formed (Curtius, 2009). Furthermore, because water vapor is abundant in the atmosphere, water-vapor will co-condense with sulphuric acid in a process called binary homogeneous nucleation. However, the rate at which this co-condensation occurs does not explain new particle formation rates observed in the atmosphere, but studies have shown nucleation can also be enhanced by the presence of other molecules or ions, such as ammonia or organic compounds (Kirkby et al., 2011). Globally, nucleated particles that grow to CCN sizes account for approximately 50 % of the total CCN budget, while the other 50 % are directly emitted into the atmosphere (Merikanto et al., 2009). To reach CCN sizes, nucleating particles must grow very quickly to avoid scavenging by larger particles (Riipinen et al., 2011), with lifetimes of particles of 1 nm being in the order of 10 minutes due to scavenging.

Aitken mode particles grow from nucleation mode particles through coagulation or condensational growth as nucleation particles are transported in the atmosphere. Nucleation particles are believed to be the largest source of particles in this range (Seinfeld and Pandis, 2016). Accumulation mode particles originate from primary sources or grow from Aitken mode

particles through coagulation of smaller particles, condensational growth, or cloud processing. Cloud processing is a combination of processes such as collision/coalescence and aqueous phase oxidation of gases that increase the size and change hygroscopicity of particles that activate during cloud formation. For aqueous phase oxidation, soluble gases dissolve into the droplets and react, forming products that stay in the particle phase (for example the transformation of sulphur dioxide to sulphate). When the clouds evaporate, volatile species evaporate, leaving behind the species in the particle phase (Noble and Hudson, 2019). After cloud processing there is a gap in PNSDs called the Hoppel minimum, between particles that activated and have grown to accumulation mode sizes, and between particles that did not activate or grow enough and stayed in the Aitken mode (Hoppel and Frick, 1990).

The Antarctic Circumnavigation Expedition took place during the SO summer of 2016-2017, one of its objectives being to advance our understanding of climate-relevant aerosol processes by "capturing their summertime variability and characteristics, providing an in situ reference of CCN number concentration for remotely sensed cloud droplet number concentration and to improve representations of pre-industrial-like aerosol properties for global climate models" (Schmale et al., 2019). Using the data collected during the expedition, the goal of this master thesis was to perform cluster analyses on submicron aerosol size distribution data and interpret the resulting clusters based on ancillary data, including several environmental datasets and back-trajectory analysis, to determine the drivers of the size distribution shape and regional differences. In section Data and Methods, the data used in this thesis as well as the clustering algorithm used is presented. In section 3, Results and Discussion, is presented the results of the clustering and how the optimal number of clusters was chosen. I then present how an exploratory approach was taken to investigate differences in distributions of the meteorological variables depending on the cluster type. South of 60°S, drivers of cluster type and variables found in conjunction with increased CCN concentration are presented. The conclusion summarizes the new insights found during this work and relates them to further possible investigation of the ACE dataset and in a broader way to investigation of aerosol processes in the Southern Ocean.

2. Data and Methods

2.1 The Antarctic Circumnavigation Expedition

All data in this master thesis come from the ACE, which took place from December 2016 to March 2017. The expedition took place on board the Russian icebreaker Akademik Tryoshnikov with an international team of scientists studying a broad number of fields, from climatology to marine biology.

The journey was divided into three legs: Leg 1 from Cape Town (South Africa) to Hobart (Tasmania), Leg 2 from Hobart to Punta Arenas (Chile) and Leg 3 from Punta Arenas to Cape Town (see Figure 1). Each leg took approximately one month, with several stops along the way. Latitudes varied between 34 °S in Cape Town and 75 °S during leg 2. A more detailed presentation of the expedition can be found in Schmale et al. (2019) and Landwehr et al. (2021).

Figure 1: Travel plan of the Antarctic Circumnavigation Expedition. Source: https://swisspolar.ch/expeditions/ace/

The main interest of this master thesis, namely particle number size distributions, were collected with a custom-built scanning mobility particle sizer (SMPS), as described in Schmale et al. (2017), with a range of 11 to 400 nm. The SMPS returns information about the number of particles in a hundred discrete bins logarithmically spaced. The SMPS is based on the principle that the ability of a charged particle to move across an electric field (electrical mobility) is related to its size (Intra and Tippayawong, 2007). It is composed of two main elements: a Differential Mobility Analyser (DMA) and a Condensation Particle Counter (CPC).

Before entering the DMA, the air sample is dried then passes through a bipolar diffusion charger that brings the particles to an equilibrium charge distribution (Wiedensohler, 1988). The bipolar diffusion charger uses an x-ray source which generates high concentrations of ion pairs to bring concentrations of positive and negative ions in the air sample to almost equal concentrations. The DMA consists of a cylinder with a high negative voltage rod in its centre, which attracts the positively charged particles. Particles entering the cylinder will move towards the central rod at the rate given by their electrical mobility. Particle with a certain electrical mobility will exit the cylinder through a slit while other particles exit with the exhaust flow. By varying the voltage of the central rod, different particle sizes can be chosen to exit the DMA at the slit, where they then pass through the CPC to count the number of particles at the given particle size range. In this case, the SMPS will cycle through 100 different voltages which are calculated to select particle sizes corresponding to the discrete logarithmically spaced bins. Figure 2a shows a schematic diagram of a DMA, where the particles in blue then pass on to the CPC. The CPC uses an Optical Particle Counter, which employs a laser to measure the amount of light scattered by individual particles (Glantschnig and Chen, 1981). The scattering being dependent on the particle diameter, detection of very small particles is difficult and as such the lower limit of detection is for diameters around 0.1 μm. To be able to count smaller particles, the CPC draws the particles through warm butanol vapour which then passes in a condenser. The alcohol vapour condenses onto the particles which grow to sizes visible by the OPC.

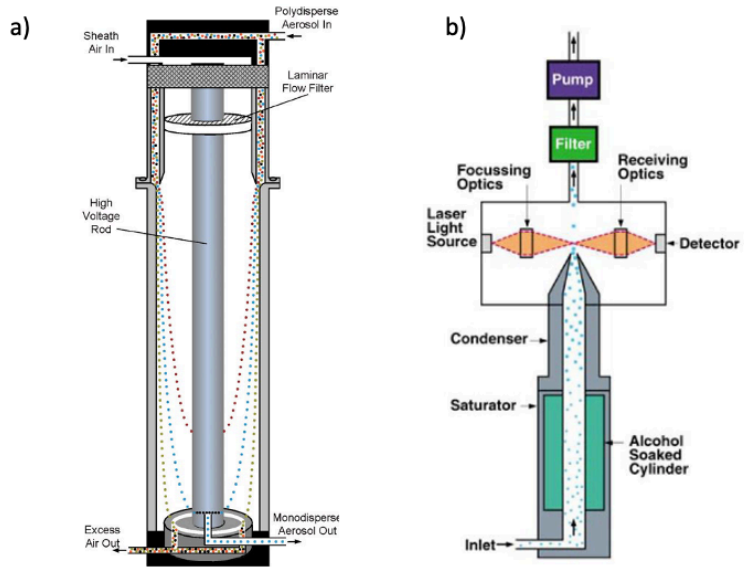

Figure 2: a) Flow schematic of a DMA. b) Flow schematic of a CPC. Source: https://tsi.com/

Among all measurements made aboard Akademik Tryoshnikov, we focused on 12 ancillary variables grouped into 3 categories according to Landwehr et al., Appendix B (2021) namely: "Atmospheric dynamics and thermodynamics", "Atmospheric side of the hydrological cycle" and "Atmospheric chemistry", which can be found in Table 1. The time resolutions for the

measurements of the variables are not the same, ranging from 1 minute to 1 h. All observations were resampled to a 5-minute frequency, corresponding to the original frequency of the SMPS data. Observations with a higher frequency were averaged to non-overlapping 5-minute windows, regardless of the number of observations in the window, while lower frequency observations were assigned to the nearest 5-minute window, as in Landwehr et al. (2021). Non-available data was mostly due to data polluted by the ship exhaust, with some instrument malfunctions (the instrument for MSA measurement was not on-board during leg 1). Cleaning of the data was performed by the scientists of the project, with exhaust periods identified using equivalent black carbon and CO and CO_2 measurements as well as 10-s variability of particle number concentration, and account for approximately 50 % of the total data (Schmale et al., 2019). Additional quality checks were performed on the SMPS data, after discovering particle number concentration above natural concentrations (see Appendix C).

Table 1: Variables used in this master thesis their respective units, time resolutions and the percentage of data available for the whole cruise

Variable	Unit	Time resolution	Data availability
Sub-micron aerosol size distribution	cm^{-3}	5 min	34.1 %
Wind speed, derived from the flow-distortion-corrected in situ measurements	M s^{-1}	5 min	95.9 %
Surface cyclone mask	-	1 h	100 %
Cold and warm temperature advection mask	-	1 h	84.4 %
Solar radiation	W m^{-2}	1 min	97.3 %
Relative humidity	%	1 min	86.8 %
$\delta^{18}O$ of atmospheric water vapour	‰	5 min	92.4 %
Mass concentration of sulphate in non-refractory particulate matter	g m^{-3}	10 min	60.4 %
Mass concentration of chloride in nonrefractory particulate matter	g m^{-3}	10 min	60.2 %
Concentration of gaseous sulphuric acid	molec cm^{-3}	5 min	61 %
Concentration of gaseous methanesulphonic acid	molec cm^{-3}	5 min	61 %
Particle number concentration acting as CCN at 0.2 % supersaturation	cm^{-3}	1 h	100 %

Meteorological data were collected from a Vaisala weather station. The MAWS 420 system onboard includes pressure sensors, ultrasonic wind sensors, humidity and temperature sensors, and a ceilometer (Walton and Thomas, 2018).

The surface cyclone mask and cold and warm advection mask were taken from Thurnherr and Wernli (2020); Thurnherr et al. (2020a). The surface cyclone mask is a binary indicator of the presence of a surface cyclone along the tracks of the research vessel, with 0 being the absence and 1 being the presence of a cyclone. The surface cyclones were calculated using a 2D cyclone algorithm (Sprenger et al., 2017) with data from the European Centre for Medium-Range Weather Forecasts (ECMWF). The cold and warm advection mask was calculated using the temperature difference between the sea surface and the air. The warm temperature advection is defined for temperature differences above 0°C and the cold temperature advection for temperature differences below 0 °C. To limit noise, the time gap between two advection events of the same type must be at least 6 hours, otherwise, all events between are considered of the same type.

Gas-phase sulphuric acid and methanesulphonic acid concentrations were measured with a nitrate chemical ionization Atmospheric Pressure Interface Time-of-flight (APi-TOF) mass spectrometer. Mass spectrometry measures the mass-to-charge ratio of molecules and their relative abundance to infer concentration in air samples. The time-of-flight mass spectrometry uses an electric field to accelerate ions. If the ions have all the same charge, their kinetic energy will be the same and therefore their velocities will only depend on their mass. By detecting how many ions arrive at the same time one can infer the ion concentration. To ionise sulphuric acid and methanesulphonic acid before they enter the TOF, a corona discharge needle is used to ionise nitric acid to nitrate. The acids are then ionised through proton transfer and directed towards the time-of-flight mass spectrometer. For details and a description of the APi-TOF instrument see Baccarini et al. (2021).

Particle number concentration acting as CCN at different levels of supersaturation were measured by a CCN counter (CCNc, type CCN-100 by DMT) (Tatzelt et al., 2020). The 0.2 % supersaturation CCN measurements were used in this thesis, as they are more representative of natural conditions than other supersaturations. A CCNc is based on the principle that the diffusion of water vapour is faster than the diffusion of heat (Roberts and Nenes, 2005) and is very similar to a CPC, as both use an OPC. The air sample goes through a cylinder in which the inner walls are porous and kept wet. A temperature gradient is established along the length of the cylinder and as thermal diffusion lags vapour diffusion, a constant supersaturation is established in the centre of the cylinder (Roberts and Nenes, 2005). This supersaturation can be controlled by modulating the airflow, temperature gradient or pressure inside the cylinder. Particles that activate at a supersaturation lower than the supersaturation of the instrument form droplets. The droplets that exit the cylinder are then counted using an OPC.

In addition to the ancillary variables measured on board the research vessel, hourly 10-day back-trajectories for the whole cruise are available in Thurnherr et al. (2020b). Calculations were performed with the "LAGRANgian analysis Tool" (LAGRANTO) (Sprenger and Wernli, 2015). Back-trajectories are the most probable path taken by air masses that arrive at a given time and place, in our case the research vessel. The LAGRANTO tool uses a Lagrangian approach to calculating back-trajectories, which involves using a moving frame of reference for resolving advection and diffusion equations with three-dimensional wind fields. The three-dimensional wind fields were taken from the 6-hourly ERA-interim global atmospheric dataset

from the ECMWF. Tatzelt et al. (2021) performed an air-mass origin analysis to identify the regions from which the air masses had come from in the last 10 days. The original dataset can be found in Tatzelt et al. (2020) . First, a precipitation threshold of 0.1 mm h^{-1} was applied to all trajectories and all rain events at the ship were removed too. Only the parts of the trajectories after the rain events were kept. Rain events are removed because aerosols are removed by scavenging and wet deposition therefore distorting measurements. Then, each time steps of back trajectories were classified as free tropospheric or from the planetary boundary layer (PBL) if the pressure level of the air parcel was below or above the pressure level of the PBL, respectively. If the time step is defined as free tropospheric, it is assumed no aerosols from the PBL were collected. If the air parcel was from the PBL, an additional mask was applied to classify the back-trajectories from coming from "land", "sea" or "coast". Finally, if the back-trajectories come from the sea, they are further classified as coming from the "open ocean" if the sea ice fraction is below 15%, from the "marginal sea ice zone" (MIZ) for sea ice fractions between 15 and 80% or from "sea ice" for sea ice fraction above 80 %. Additional masks were calculated, such as trajectories coming from Africa or South America, but were not used in this thesis due to the low prevalence of such cases (Radenz et al., 2021). The final dataset gives the fraction of the time steps in the 10-day back-trajectories for each classification.

2.2 The k-means clustering algorithm

Cluster analysis is an unsupervised machine learning approach dimensionality reduction and has been frequently used with SMPS datasets (Lachlan-Cope et al., 2020; Dall'Osto et al., 2019; Lange et al., 2018; Beddows et al., 2009). It is a powerful tool which separates the data into a predefined number of groups of similar size distributions. It is often used to compare the distributions over different periods at a single site as well as comparing distributions across several sites (Beddows et al., 2014). Here measurements are made on a moving ship, which does not allow us to study the evolution in time of the size distributions at a single site but instead to study the geographical differences that influence the size distributions.

Before clustering, the SMPS measurements are normalised. Normalisation ensures that we cluster on the shape of the distributions, irrespective of the magnitude of the total count of the observation. The SMPS returns values for each bin as dN/dlogDp, where Dp is the width of each bin in nanometres and N is the particle count. To normalise, we first multiply each value by the logarithm of the bin width, and we then divide by the sum of each observation. After normalisation, the values sum to 1.

The k-means algorithm aims to partition the observations into k clusters. The original and easier to understand algorithm is Lloyd's k-means clustering (Lloyd, 1982). In this algorithm, the observations are assigned to a cluster so that they are closer to their cluster centroid than to other cluster centroids, the cluster centroid being the arithmetic mean of all data points belonging to that cluster. By randomly selecting observations as starting centroids, the algorithm will then assign data points to the nearest cluster. Once every data point has been assigned, the centroids of each cluster are updated, data points are once again assigned to the nearest centroid and the algorithm iterates this datapoint until the algorithm converges. When

data points are no longer assigned to different clusters, the algorithm has converged and stops (Hartigan and Wong, 1979). Lloyd's k-means clustering does not guarantee convergence, as it can converge towards local optimums, and in some cases can be slow. In this master thesis, the Hartigan-Wong k-means clustering algorithm was used. The Hartigan-Wong algorithm tries to minimise an objective function, in this case the sum-of-squares. It will first iteratively assign the data point to a certain cluster and calculate the sum of the sum-of-squares of each cluster for each case. The data point is then assigned to the cluster which minimised this sum. This means a point can be assigned to a cluster even if its distance is higher, provided that the minimum sum of squares is achieved. This method improves convergence compared to Lloyd (Slonim et al., 2013).

In conjunction with the Hartigan-Wong clustering algorithm, the k-means++ algorithm was used ("kmeanspp" in the R library LICORS). It is used for choosing the initial data points for the centroids. The first centroid is chosen at random, and subsequent centroids are chosen from all other data points with a probability proportional to their squared distance to the other centroids. This helps in having well defined initial centroids and reduces the computation time, which is useful for large datasets (Arthur and Vassilvitskii, 2007). Moreover, the kmeanspp() function is run with 50 random sets, meaning it is run 50 times with different starting centroids, and the solution with the lowest objective function is kept.

The k-means clustering requires the user to choose the number of clusters. A variety of methods were created to aid in the decision, such as the Gap statistic, the Dunn Index or the Silhouette Width (SW) (Tibshirani et al., 2001; Bezdek and Pal, 1995; Rousseeuw, 1987). All these statistical tools try to find the number of clusters for which the clusters are most compact, with the smallest intra-variance possible and the highest differences in the mean between clusters. Relying only on these statistical tests can be useful when exploring underlying patterns without the user's preconception of the data. However, when the user wants to explore known trends in the data these statistical tools should only be used as a guide. Even in this case, this isn't a straightforward decision: the final goal of the clustering, the environmental conditions and other parameters can influence and ultimately decide how many clusters the user chooses (Lachlan-Cope et al., 2020).

The SW and Gap statistic were calculated using the fviz_nbclust() function (from the "factoextra" package in R) for a maximum of seven clusters. The Gap statistic compares how different is the within-cluster dispersion to a reference null distribution. A large gap statistic means the clustering does not resemble the clustering of a uniform distribution. The reference distribution is averaged from Monte Carlo replicates of a uniform distribution, which gives each number of clusters in the reference distribution a standard deviation. The number of clusters k is chosen as the smallest k for which the Gap statistic is greater than the Gap statistic at k+1 minus 1-standard-deviation. With the SW, the silhouette width of every object is calculated, which is a measure of how similar an object is to its own cluster and how dissimilar it is to other objects. For higher average silhouette width, most points are similar to their own cluster and dissimilar to others, while lower averages indicate there may be too many or too few clusters.

3. Results and Discussion

3.1 Results of the clustering

One of the goals of this project was to evaluate geographical differences in the size distribution of aerosols. To that end, we started by clustering data of the whole cruise with different number of clusters to see the dominant size distributions present in the data and potential patterns in location of each cluster. Similarly, we clustered data separated by leg and latitude band, north of 60 °S and south of 60 °S. It should be noted datapoints south of 60 °S only belong to the southernmost part of leg 2. This was done firstly to explore if there were any regional differences in the clustering, with each leg and latitude band having different sources or influences on aerosols. For example, the SWW, which is strongest at latitudes between 40 and 50 °S (Anderson et al., 2018) or the influence of the Antarctic continent south of 60 °S. Secondly, to determine the optimal number of clusters in each case, which may require different number of clusters to show the main patterns in distribution. In this section we only present the result of the clustering for the whole cruise and for latitude bands. Other clustering results, for example for the legs and different number of clusters, can be found in Appendix B.

We started by calculating the SW and Gap statistic for data of the whole cruise to get an idea of the statistical optimal number of clusters. Figure 3a shows the average SW for one to seven clusters. The average SW is greater for three clusters, indicating that with this number of clusters the observations are on average better matched inside their own cluster and poorly matched to other clusters. For higher number of clusters the differences between clusters become less marked and observations usually start to match with other clusters, bringing the average SW down. Figure 3b shows the gap statistic for one to seven clusters. In this case, we can see that 4 clusters is the smallest number of clusters for which the Gap statistic is higher than the Gap statistic for k+1 minus the standard deviation. The metrics to assess the optimal number of clusters give different results. While these results give us an idea of the statistical optimal, a more subjective decision will be made by observing the PNSDs for different number of clusters and by exploring how the geographical distributions evolve when increasing the number of clusters. This requires previous knowledge of aerosol size distributions and a general idea of what the final clusters should look like, which is usually necessary for k-means clustering, unless a purely exploratory approach guided by statistics is taken.

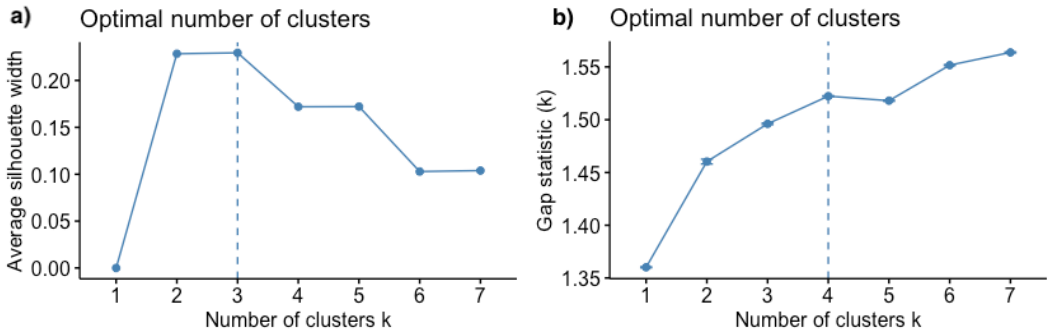

Figure 3: a) Average silhouette width for 1 to 7 clusters using the data from the whole cruise. b) Gap statistic for 1 to 7 clusters

Results of clustering the data for the whole cruise with two to four clusters are given in Figure 4. The lines represent the mean PNSDs of each cluster, with the mean of every datapoint at a given particle size. On the left is the results of the clustering for the normalised data, see the Methods section for the normalisation method. The result of the clustering is applied to the original data, which is shown on the right, but as mentioned before the clustering is not done directly on the original data as to cluster based on the shape of the distributions and not on their intensity. The raw data is shown to get a general sense of the intensity of each group. The clusters have very distinct distributions with modes at different particle sizes. For simplicity, we call each cluster by its main mode (i.e., nucleation cluster, Aitken cluster, accumulation cluster).

When clustering for two groups, we obtain a cluster with an Aitken mode and a cluster with a pronounced bimodal distribution, with a peak in the accumulation mode. The accumulation cluster occurs more often with 65.8 % of the data (representing 5345 data points) while the Aitken cluster has 34.2 % of the data (representing 2774 data points).

When clustering for three groups, the accumulation cluster occurs again more often with 62.6 % of the data. It is still characterised by a pronounced bimodal distribution, with peaks at approximately 45 and 130 nm. The Hoppel minimum is at approximately 72 nm. The Aitken cluster is the 2nd most frequent with 23 % of the data (or 1865 data points). It is characterised by a mode at 34 nm. The nucleation cluster is the smallest cluster with 14.5 % of the data (1174 data points). The left tail of the distribution increases towards smaller particle sizes due to a small number of nucleation events with small particle sizes and with very high particle number concentrations (see Appendix, Figure A.2). We can see the Aitken cluster we found with two clusters separated into a nucleation and Aitken cluster for three clusters, while the accumulation cluster is stable percentagewise.

For four clusters, the Aitken cluster we found with 3 clusters separates again into two new Aitken clusters. The Aitken 1 cluster has a peak at 42 nm and accounts for 18.4 % of the data while the Aitken 2 cluster has a peak at 29 nm and accounts for 15.3 % of the data. The bimodal distribution of the accumulation cluster is less pronounced than with 3 clusters, but a Hoppel minimum is still visible at 70 nm, with the bimodal peaks at 49 and 130 nm. The data percentage is lower than the accumulation cluster for 3 clusters, with 55.2 % of the data. The

nucleation cluster is also stable, with 11.1 % of the data, and shows the same distribution as the nucleation cluster with 3 clusters.

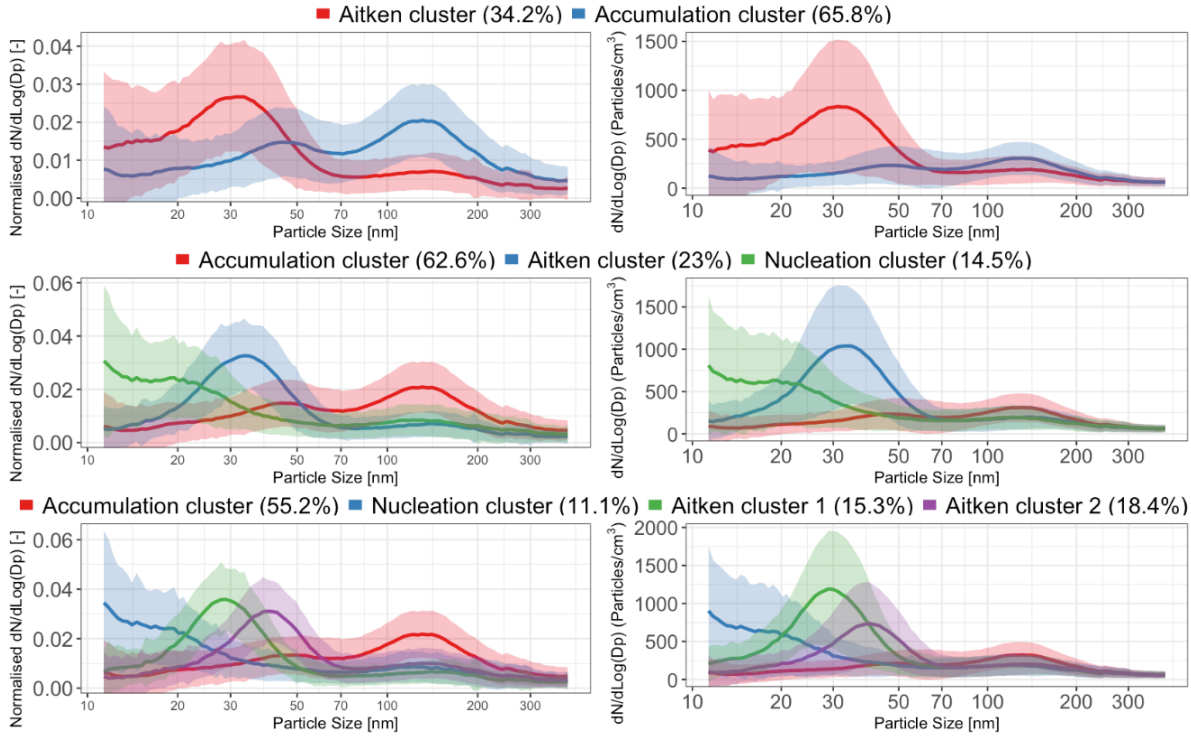

Figure 4: PNSDs of clusters for normalised (on the left) and raw (on the right) data. Clustering was done on the data of the whole cruise, with two to four clusters. Shaded areas represent standard deviation. Colours do not represent the same clusters for the different number of clusters.

Having at least three clusters seems necessary as otherwise we lose information about the nucleation cluster. By only looking at the PNSDs for the different number of clusters, we cannot determine whether 3 or 4 clusters are more appropriate. By looking at the geographical distribution of clusters, we can deduce if the different Aitken clusters are geographically close, representing different stages of particle growth, or if they appear geographically different, meaning they would possibly have different formation processes.

Figure 5 shows the results of the clustering for 3 clusters for the whole cruise. The locations of data points are shown on a map in Figure 5a while the normalised PNSDs for each cluster are shown in Figure 5b. It should be noted that because the SMPS data has a frequency of 5 minutes, the density of points can be high and makes visibility difficult. Therefore, the location of data points on the map are jittered around the original location (both in latitudinal and longitudinal directions) and so do not represent the exact locations.

We can see leg 1 has only sparse data. The beginning of the leg shows mostly a mix of accumulation and Aitken clusters and towards Australia a mix of nucleation and Aitken clusters. Leg 2 shows a high density of accumulation cluster data points all around Antarctica while for latitudes between 50 to 60 °S Aitken cluster data points are mostly present. These latitudes are characterised by the Antarctic circumpolar current, driven by the SWW. A high density of nucleation cluster data points is located around the Mertz glacier on the coast of Antarctica (see Figure 1 in the introduction for the exact location of the Mertz Glacier). These

nucleation datapoints are characterised by very high particle numbers, which are partly responsible for the sharp increase in the left tail of the nucleation cluster. Leg 3 shows successions of all three clusters. The nucleation and Aitken clusters are always geographically close to each other, which is consistent with the growth of the nucleation mode into the Aitken mode.

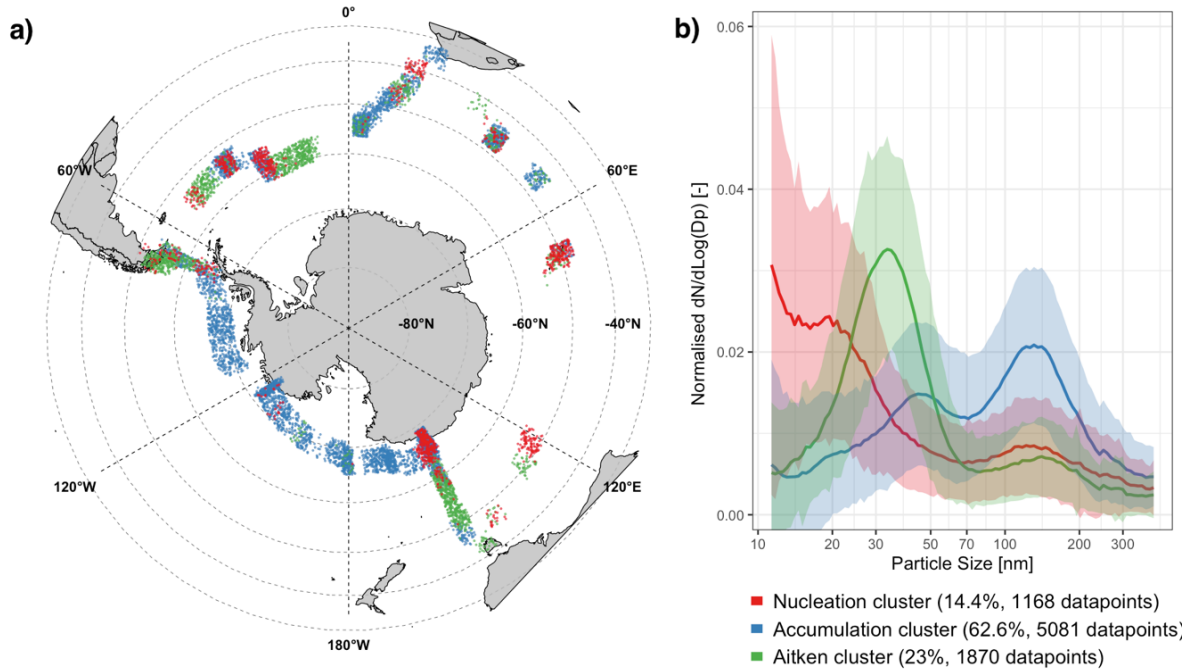

Figure 5: a) Geographical distributions of 3 clusters for the whole cruise. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. b) PNSDs of the 3 clusters for particle size between 11 and 400nm

Figure 6 shows the results of the clustering for 4 clusters for the whole cruise. Note that the colours for the cluster type are not constant. The nucleation and accumulation cluster distributions are very similar to before. We can see more data points south of 60 °S now belong to Aitken cluster 1, where only some Aitken cluster data points were already present with only three clusters. The two Aitken clusters are most of the time very close geographically, either trailing behind each other or mixed. The only exceptions are during leg 1 towards Australia where only the Aitken cluster 2 and the nucleation cluster are found. However, with the missing data, we might have missed the growth of the Aitken cluster 2 into the Aitken cluster 1.

Because the two Aitken clusters do not seem to provide more information than indicating slightly different stages of particle growth, it was decided to use divide the data into only three clusters. This also has the advantage of simplifying the interpretation of further results. Moreover, dividing the data into more groups reduces the statistical robustness of our results, especially when we already partition the data by latitude band or by leg.

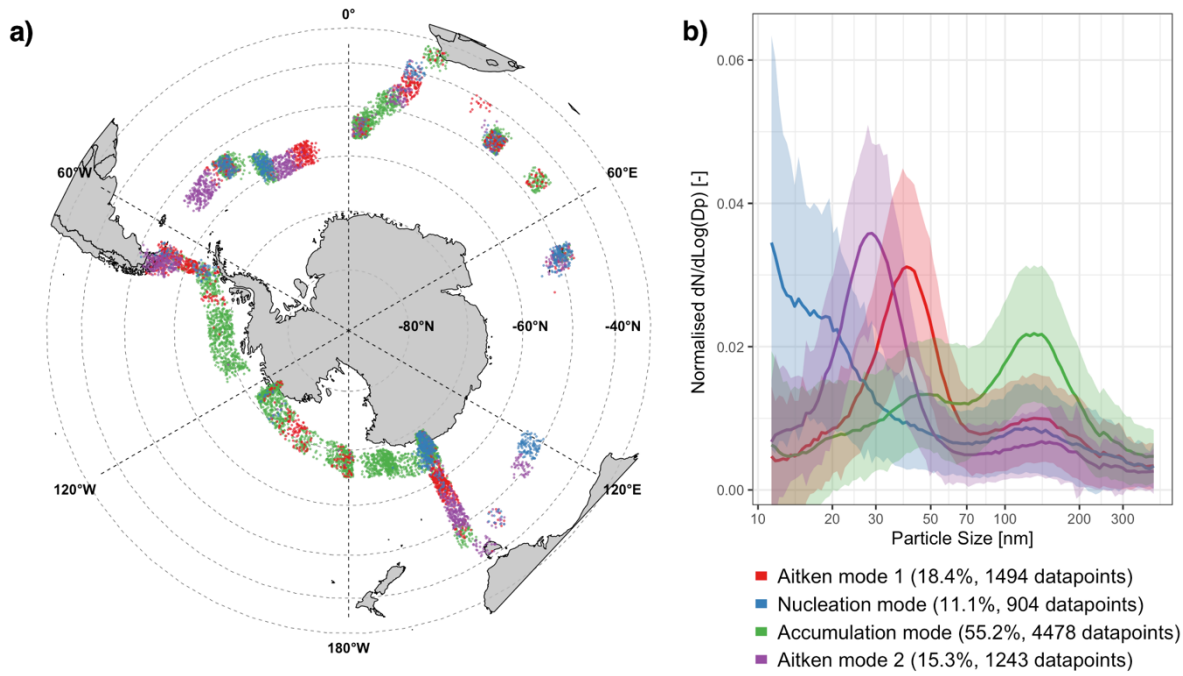

Figure 6: a) Geographical distributions of 4 clusters for the whole cruise. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. b) PNSDs of the 4 clusters for particle size between 11 and 400nm

When clustering only the data south of 60 °S, the distribution of the Aitken cluster has now a pronounced Hoppel minimum, a sign of cloud processing (Figure 7). For simplicity's sake, we will still refer to it as the Aitken cluster as its major mode is the Aitken mode. Clustering on leg 2 was already studied by Schmale et al. (2019), while clustering on the data south of 60 °S doesn't include most of the two Aitken cluster events that happened close to the coasts of Australia and South America. The clustered normalised size distributions they obtained (see Figure 7.a in Schmale et al., 2019) are nearly identical to what we obtained for the clustering of leg 2 (see Appendix, Figure A.2), and are very similar to the clustering of the data south of 60 °S. Using back-trajectory analysis they showed that the accumulation cluster was associated with cold fronts as well as air masses from higher altitudes and more southerly latitudes. They also showed that a cluster with a clear bimodal distribution, equivalent to our Aitken cluster, was associated with MBL air masses and that the cloud height was mostly within the MBL, supporting the hypothesis that cloud processing drive particle size distribution at the surface and not in the FT. Sanchez et al. (2021), also found the MBL cloud fraction was highest for high accumulation mode regimes, again supporting that hypothesis.

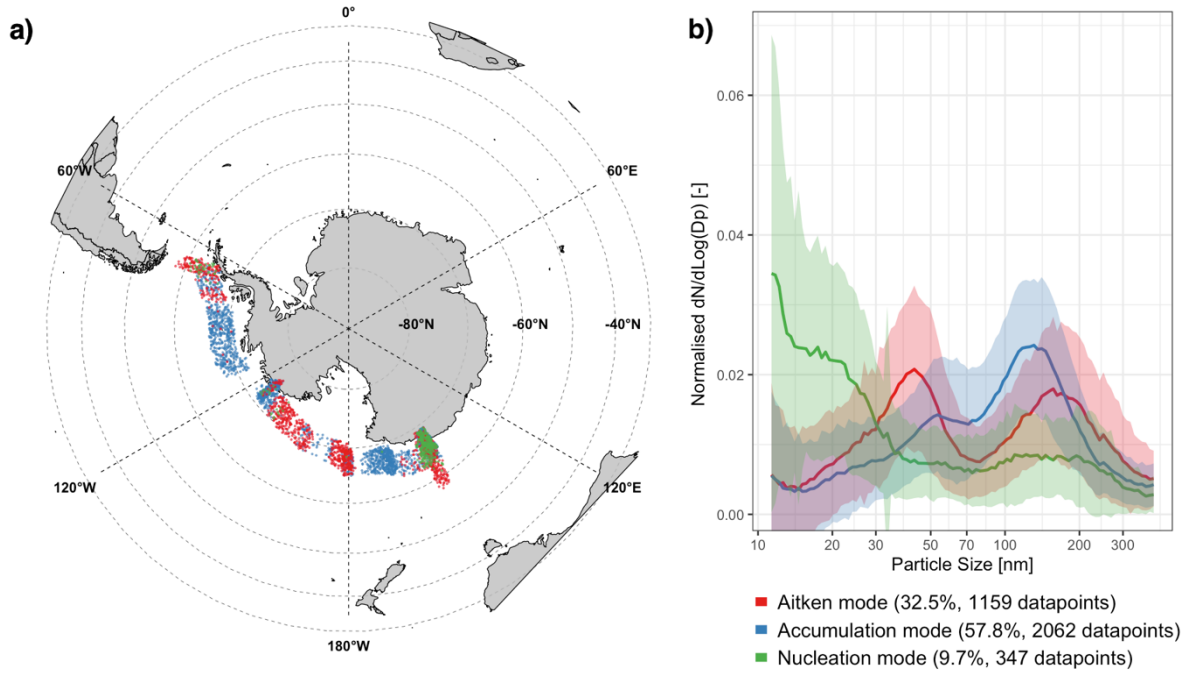

Figure 7: a) Geographical distributions of 3 clusters for datapoints with latitude south of 60 °S. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. b) PNSDs of the 3 clusters for particle size between 11 and 400nm, shaded areas represent standard deviation.

Clustering on the data north of 60 °S, the Aitken cluster shows little sign of cloud processing with a very unimodal distribution. Nearly 20 % of the data belongs to the nucleation cluster and the accumulation cluster is less important with 50 % of the data.

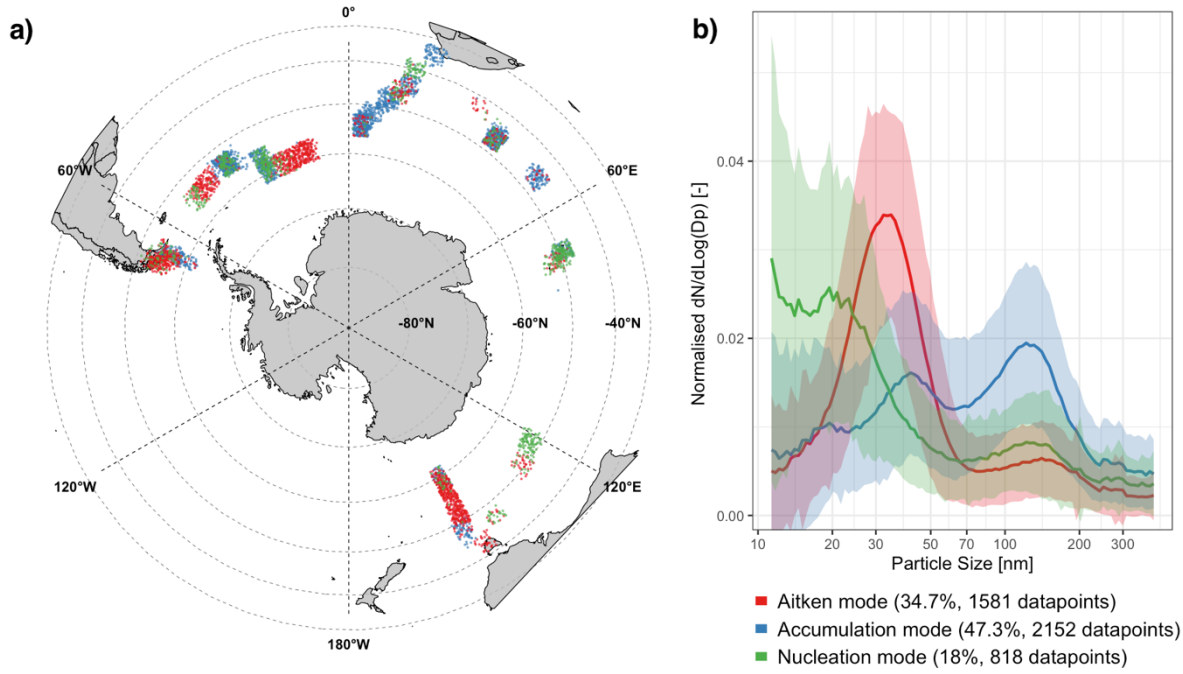

Figure 8: a) Geographical distributions of 3 clusters for datapoints with latitude north of 60 °S. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. b) PNSDs of the 3 clusters for particle size between 11 and 400nm, and shaded areas represent standard deviation.

The results of clustering on the data separated by latitude bands show clear differences, with the nucleation cluster occurring nearly twice as much north of 60 °S than south of 60 °S, and with a much more pronounced bimodal distribution of the Aitken hinting at more prominent cloud processes. To study these regional differences, all following figures will be divided by latitude bands, each using its own clustering results and not the result of the clustering for the whole cruise simply separated by latitude.

3.2 Environmental variables & Back-trajectories

To study the influence of the ancillary variables on the prevalence of each cluster, we first investigated the distributions of each cluster according to environmental variables. Frequency polygons, which are in essence line graphs that join the midpoints of the top of histograms, were calculated. The variables are divided into 30 equal-sized bins and for each bin, the number of data points of each cluster that fall in the bin are counted. Figure 9 shows such frequency polygons for in-situ meteorological variables, in this case, wind speed and solar radiation. We can see the meteorological variables show the same distributions for all clusters, with only the relative importance of each cluster being different. An increase or decrease in an in-situ variable does not influence the importance of only one cluster. Other frequency polygons for in-situ meteorological variables can be found in the Appendix B, while other variables present in the datasets were explored but did not seem to influence cluster type. From this we can say the meteorological variables we explored do not seem to influence the cluster type in-situ.

The fact that in-situ meteorological variables did not seem to influence cluster type prompted the investigation of air mass history. In the back trajectory data, air masses can have collected aerosols from two origins: the FT and the PBL. Fractions of back-trajectories coming from land or coast are very sparse, the 3rd quartiles for such cases are both 0 with mean fractions below 3%, meaning that for statistical purposes we can approximate fractions from the PBL as being from the MBL. This means that for a certain percentage of back-trajectories from the FT, the rest of the trajectories are from the MBL. Figure 10 shows the number of occurrences by latitude band of data points having at least some fraction of back-trajectories from the FT or being entirely from the MBL, for each cluster. Because of the way the dataset is constructed, in the case where aerosols have both origins we cannot know in which order they happened, meaning we don't know for example if air masses have gone through the MBL, then back to the FT before subsiding into the MBL at ship location.

As a reminder, north of 60 °S, approximately 45% of data points belong to the accumulation cluster, 35 % to the Aitken cluster and 20 % to the nucleation cluster while south of 60 °S that distribution is 60 %, 30 % and 10 % respectively. However back-trajectories were not available for all data points so distribution of clusters are not exactly the same in Figure 10 . North of 60 °S, 50 % of air masses are entirely from the MBL, while it is 30 % south of 60 °S.

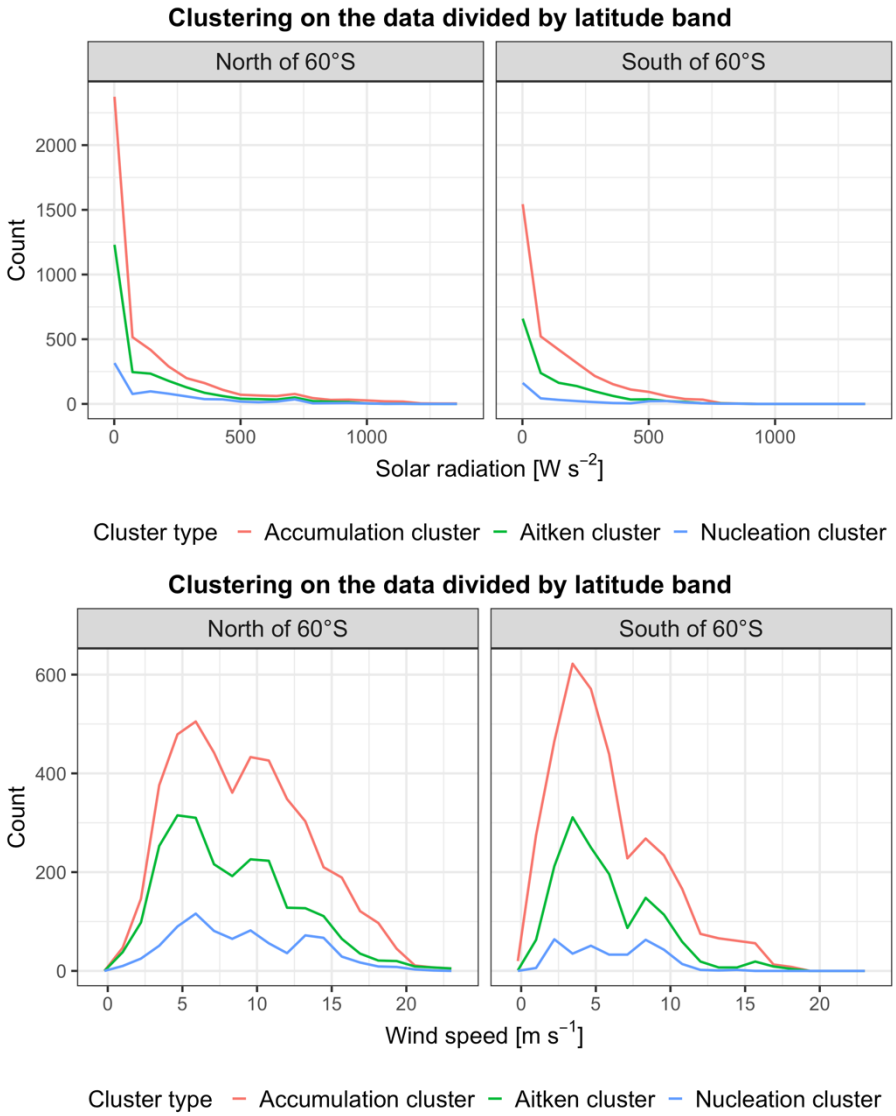

Figure 9: Frequency polygons of each cluster for solar radiation and wind speed. Lines link the top of invisible histograms.

We can see that most Aitken and nucleation cluster data points occur more frequently when trajectories are entirely from the MBL, with an exception south of 60 °S for the Aitken clusters which show similar numbers. North of 60 °S, the ratio of Aitken and nucleation cluster data points with a FT influence to data points with an entirely MBL influence are respectively 0.6 and 0.35 while south of 60 °S the ratio for the nucleation cluster is 0.5. For the accumulation cluster, the ratio is 0.6 north of 60 °S and 0.2 south of 60 °S.

Mechanisms driving new particle formation events (NPF) in the SO, which are responsible for generating nucleation and Aitken mode particles, are not yet well understood. The majority of nucleation and Aitken mode particles produced in summer is thought to be due to sulphur-containing gases that are uplifted to the FT, which, due to the lower condensation sink and colder temperatures, allows particles to nucleate and grow to Aitken-mode sizes before they subside into the MBL and grow in size to dominate sulphur-based CCN (McCoy et al., 2021; Quinn et al., 2017; Humphries et al., 2016). However, NPF events have also been observed directly in the summertime MBL (Jokinen et al., 2018; Jung et al., 2019), driven by open water

and melting sea ice regions (Brean et al., 2021). Weller et al. (2015) reported NPF over coastal areas in the MBL but these events were generally restricted to the nucleation mode, meaning particle sizes did not grow to accumulation mode sizes. From our results, it seems the MBL does indeed drive nucleation and Aitken mode particles and might even be more important than FT processes.

South of 60 °S, our observation of the high fraction of accumulation cluster driven by FT sources is concordant with observations from Schmale et al. (2019) for leg 2, which is mostly located south of 60 °S, that airmasses for the accumulation mode came mostly from higher altitudes and latitudes. Simmons et al. (2021) also found higher CCN concentration, which are mostly found in the accumulation cluster, for airmasses coming from the FT.

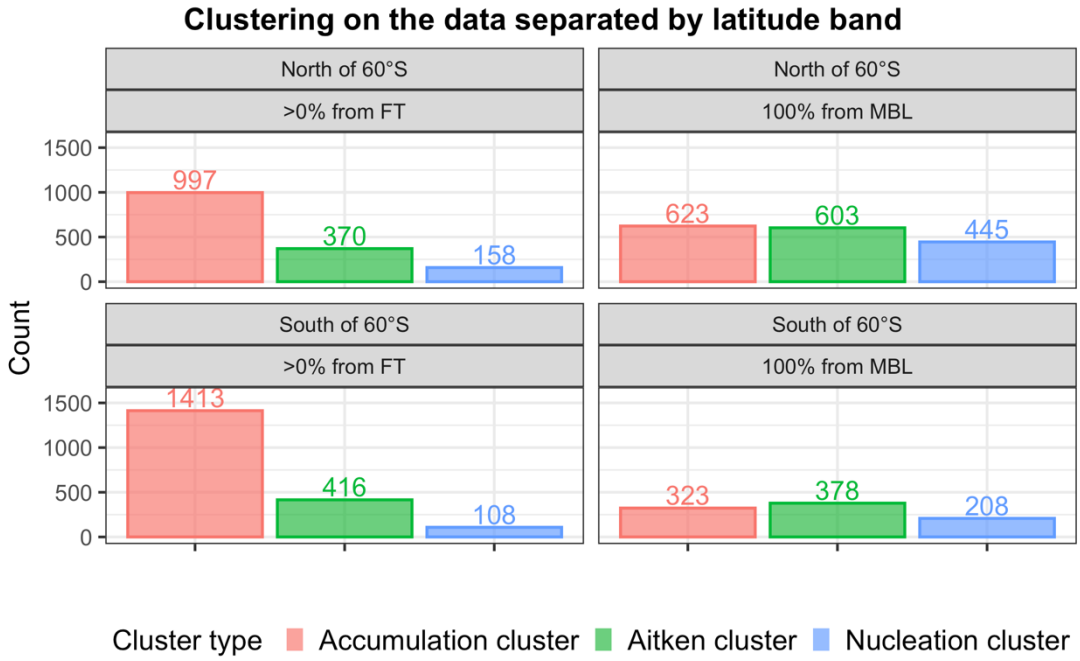

Figure 10: Count of occurrences of data points having a non-zero percentage of back-trajectories from the FT or being entirely from the MBL, north and south of 60 °S

The origins of the nucleation cluster particles in the MBL were studied in more details. Trajectories can come from the open ocean, sea ice zones or marginal sea ice zones depending on the sea ice fraction (see the Methods section for details). While open ocean and sea ice zones did not seem to particularly drive the nucleation cluster, south of 60 °S, back-trajectories from above the marginal sea ice zone seem to predominantly drive the nucleation cluster (Figure 11). Boxplots of all the non-zero percentages of back-trajectories from marginal sea ice zones show that they were mostly associated to the nucleation cluster. The highest concentrations of MSA in the Antarctic are found in the marginal sea ice zone, so this result supports the hypothesis of local NPF in the MBL due to high concentration of precursor gases (Becagli et al., 2022). It must be noted however that south of 60 °S, only 25 % of data points in the nucleation cluster have non-zero percentages of back-trajectories coming from marginal sea ice zones, which means other formation and transport mechanisms, such as free tropospheric NPF, drive the remaining nucleation cluster data points.

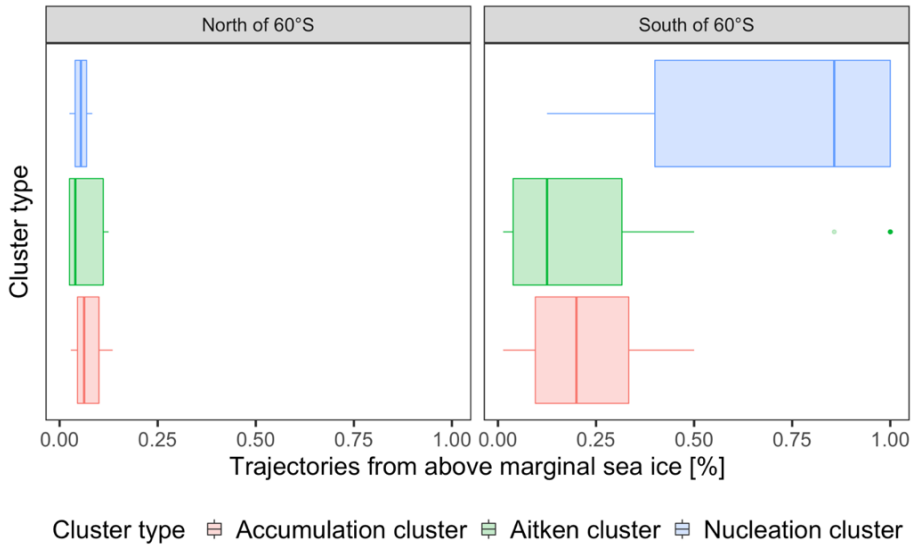

Figure 11: Boxplots of percentages of back-trajectories from marginal sea ice zones for each cluster for data points with a percentage >0%. As only 11% of the data (909 data points) had occurrences of such trajectories, data points with no occurrences are removed for visibility. For the nucleation cluster south of 60°S, only 25% of data points have back-trajectories coming from above the marginal sea ice zone.

We now focus on sources of CCN, which are most important for radiative forcing due to aerosol-cloud interactions. Because few particles in the nucleation and Aitken mode can act as CCN at 0.2 % supersaturation, we focus only on the accumulation cluster. For leg 2, (Schmale et al., 2019) found a median critical particle diameter, the diameter above which particles can act as CCN, of approximately 90 nm for 0.2 % supersaturation.

Conditional density plots show how the value of a variable varies according to intervals of two other variables. In our case, we show how the median N_{CCN} at 0.2 % supersaturation varies according to a categorical variable (i.e., latitude band or advection mask) and a continuous ancillary variable. The ancillary variable is split into quantiles meaning every interval has the same number of cases, giving robustness to the interpretation.

For increased chloride, sulphate and MSA concentrations, we see an increase in median N_{CCN} at 0.2 % supersaturation for the accumulation cluster (Figure 12). This is consistent with (Twohy et al., 2021) who found that, in summer, sulphur-based particles, mostly from biogenically sourced sulphate with some contribution from biogenic MSA, dominate number concentrations of particles 0.1-0.5 μm diameter, with a smaller but significant contribution from sea spray. Sea spray is the direct source of chloride in the atmosphere. Higher MSA concentrations south of 60°S are consistent with higher DMS emissions from summertime microbial activity (Becagli et al., 2022). DMS is the precursor of MSA and the dominant natural source of non-sea-salt sulphate in the atmosphere (Gondwe et al., 2003). Not all pathways of DMS oxidation have been resolved, however MSA is thought to be mainly formed in aerosol and cloud droplets via multiphase oxidation of DMS oxidation products (Chen et al., 2018; Hoffmann et al., 2021). The role of MSA in aerosol formation is debated, with some studies showing it could participate in nucleation (Hodshire et al., 2019; Zhao et al., 2017) while other studies note above all its role in particle growth (Beck et al., 2021).

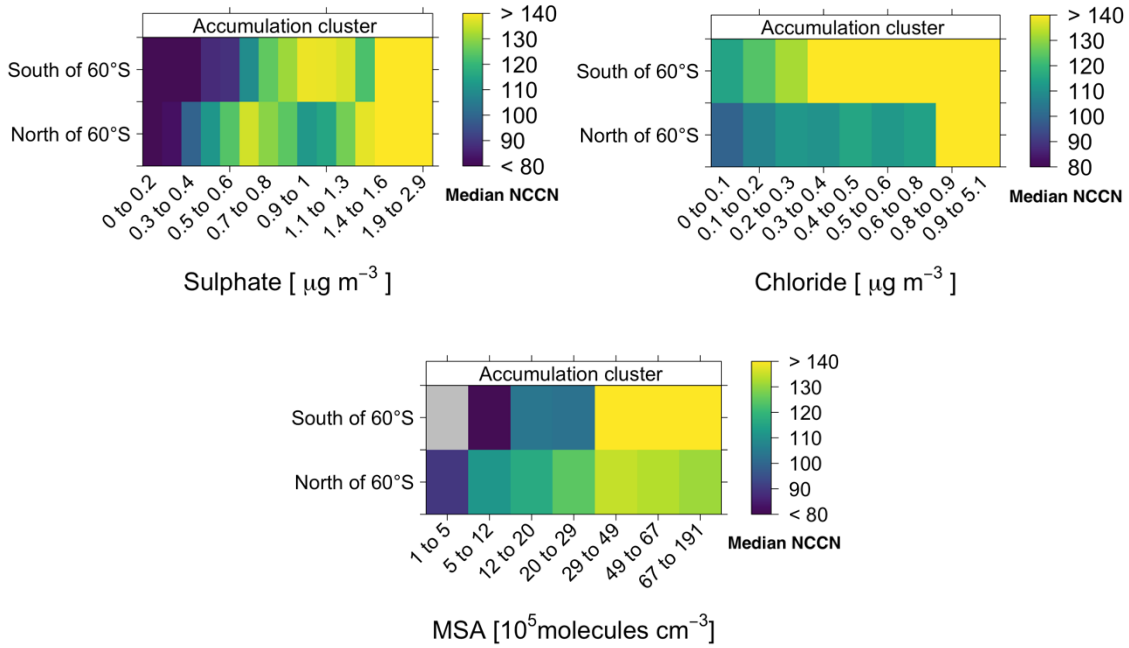

Figure 12: Conditional density plot of N_{CCN} at 0.2% supersaturation according to latitude band and quantiles of concentration of: a) Sulphate b) Chloride and c) MSA

Similarly, we can see the same increase in median N_{CCN} for lower relative humidity and for cold fronts located south of 60 °S, which bring cold and dry air from the FT (Figure 13). Simmons et al. (2021) used absolute humidity, derived from relative humidity and temperature measurements, to categorise aerosol measurements. They related low absolute humidity values to air masses coming from either katabatic flows (which originate from the FT) or from the FT directly and found higher $CCN_{0.55}$ concentration for airmasses with these low absolute humidity values, which is also consistent with findings from Chambers et al. (2018). Moreover, when grouping together particle size distributions by absolute humidity categories (low, mid, high), they saw high humidity conditions, which they related to airmasses having a longer residence time in the MBL, showed a higher magnitude of bimodality than lower absolute humidity conditions. This again points towards the importance of cloud processing in the MBL for aerosols south of 60 °S, which we had seen in the PNSDs of the clusters of Figure 7.

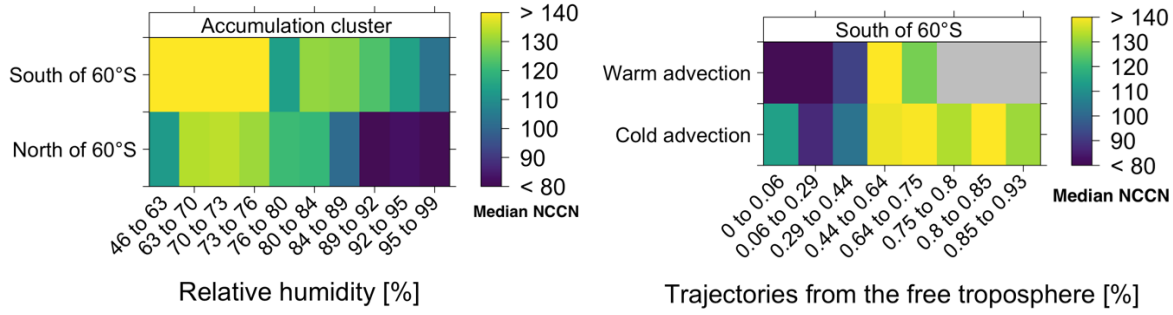

Figure 13: Conditional density plot of N_{CCN} at 0.2 % supersaturation according to: a) latitude band and quantiles of relative humidity and b) advection mask and quantiles of trajectories from the FT south of 60°S

One of the most important sources of natural aerosols is the ocean, through the production of sea spray aerosol (SSA). Sea spray aerosol is not only composed of water and salt but also of a variety of particles with a high organic composition, which gives rise to high uncertainties when modelling its effect on CCN formation (Schiffer et al., 2018). SSA can be produced by several mechanisms, mostly through the action of wind and waves. The main formation mechanism is through the breaking of waves, which entrain air into the water. The air regroups and forms bubbles that burst at the surface, creating film and jet drops that form SSA (Quinn et al., 2015).

The Southern Ocean is one of the stormiest oceans in the world, due in part to the importance of the SWW, and has the potential to create large amounts of SSA. Despite SSA having the highest CCN activation potential, its relative importance to the MBL CCN population is disputed. For summertime in the Southern Ocean, Fossum et al. (2018) report an 8-51 % contribution to marine air CCN with up to 100% contribution when wind speeds are higher than 16 m s^{-1} for 0.2-0.3 % supersaturation while Schmale et al. (2019) reported for the ACE cruise 32, 19 and 30 % SSA mode contribution to CCN at 0.15 % supersaturation for leg 1, 2 and 3 respectively. However, using a hygroscopicity coupled multimodal fitting analysis instead of the free-monomodal approach habitually used, Xu et al. (2022) reported contribution estimates as much as 5 times higher. Here, we see that increased wind speed does not correlate with increased median CCN numbers (Figure 14). North of 60 °S, median CCN numbers were higher for low wind speed conditions while south of 60 °S no trend is visible, with high median NCCN for winds between 9 and 12 m s^{-1} but lower for higher wind speeds. Nair et al. (2005) have found that although increased wind speed increase in-situ produced SSA concentrations, concentration of non-sea-salt aerosols decrease with increased wind speed, which they attribute in part to increased dry deposition. It should also be noted that wind speeds above 4 m s^{-1} are sufficient to form SSA (Seinfeld and Pandis, 2016) and that wind speed is not a direct indicator of SSA formation with Revell et al. (2019) showing that models can overestimate the wind speed dependency of SSA formation (in this case the HadGEM3-GA7.1 chemistry–climate model).

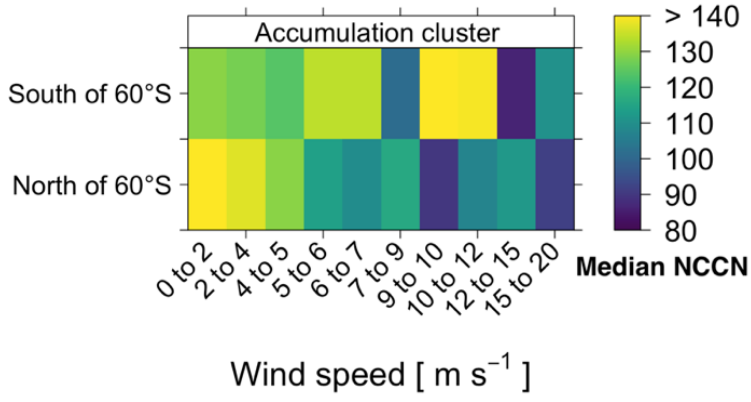

Figure 14: Conditional density plot of N_{CCN} at 0.2 % supersaturation according to latitude band and quantiles of wind speed measured on the ship.

The high transport dependency of particles and the different formation processes of the three main modes of submicron particles explain why in-situ meteorological variables show no effect on the prevalence of a particular cluster. For nucleation particles, the important NPF events that happened near the Mertz glacier would be interesting to study in more detail to see if the hypothesis of local events in the MBL due to high concentrations of biogenic gases could apply to this case. However, due to a malfunction of the mass spectrometer, there is no data about the composition of the chemical composition of the airmass. Another study based on the ACE dataset by Baccarini et al. (2021a) reported that the ship crossed six NPF events, all happening during the day, with the two events happening at the Mertz showing exceptionally cold conditions with high solar irradiance. They also found a marine origin of back-trajectories for the two NPF events, in particular sea-ice regions. NPF events normally create particles that belong to the nucleation and Aitken modes, explaining the Aitken cluster data points found near the Mertz glacier, but not all events found by Baccarini et al. were categorised as such by the clustering algorithm. This can be explained by the two different approaches taken, with Baccarini et al. focusing on events and this paper using statistical analyses, which causes some loss of information during the clustering. As for the importance of meteorological variables found during these two events, other nucleation cluster occurrences during legs 1 and 3 with other meteorological conditions could have obscured the importance of solar irradiance and low temperatures for the occurrence of the nucleation cluster in the MBL, or they could have been due to other processes outside the MBL.

The other hypothesis of nucleation particle formation in the FT is harder to study with the ancillary variables measured on the ship and should be studied by other means, including measurements in the troposphere and more detailed back trajectories. In particular, the altitude history of airmasses might show the entrainment into the FT of biogenic precursor gases from above coastal and marginal sea ice zones and their subsiding into the MBL. North of 60 °S, the Aitken cluster shows a much more pronounced unimodal distribution, which means cloud processing is less prominent there. South of 60 °S, the strong bimodal distribution of particles for both the Aitken and accumulation clusters shows cloud processing is particularly important. As hypothesised in Schmale et al. (2019), the pronounced bimodal distribution might be the result of multiple cycles of cloud processing, due to successive subsidence of cold and dry air, either directly from the FT or through katabatic flows. Conversely, the SWW and its numerous extratropical cyclones could explain the unimodal distribution of the Aitken cluster north of 60°S by bringing warm and humid air from lower latitudes, creating rain events and preventing multiple cycles of cloud processing from taking place.

The accumulation mode south of 60 °S is characterised by the importance of back-trajectories from the FT, with the subsidence of air masses driven by cold fronts. These air masses, which have significant CCN numbers, show an increased concentration of sulphate, chloride and MSA. This again points towards the air masses having gone through cloud processing, with MSA being predominantly produced in the aqueous phase (Baccarini et al., 2021a).

4. Conclusion

To better understand future climates caused by human influences, understanding of atmospheric processes in past climates is necessary. Part of the research during the ACE aimed at improving our understanding of aerosols in a near pristine environment by collecting relevant data and studying the natural processes behind their formation and evolution in the atmosphere. Understanding the drivers of aerosol size distributions in the Southern Ocean is a complex matter, requiring knowledge over a wide range of fields of research, from marine biology to atmospheric physics and chemistry. Relevant time scales for aerosols are diverse, with nucleation events happening in the order of hours and long-range transport happening over days or weeks. The Southern Ocean, which covers an area of more than 20 million km^2, has seasonal events that influence atmospheric processes relevant to aerosols such as summer phytoplankton blooms which increase biogenic precursor gases concentrations or stronger winter winds driving SSA production. As datasets needed to investigate these processes become increasingly large, tools need to be used to reduce to dimensions more comprehensible by humans, without losing essential information.

In this thesis, I investigated in-situ meteorological variables and chemical composition of airmasses to try to determine the environmental drivers of aerosol size distributions and their regional differences. Particle size distribution measured with an SMPS were related to ancillary data measured on ship as well as to hourly 10-day back-trajectories.

Clustering allows us to greatly reduce the dimensionality of a dataset by grouping together similar observations. This is especially useful for SMPS data, which in our case had 100 bins or variables per observation, by reducing it to a single category indicating to which general shape each observation corresponded most. However, an understanding of the data is necessary to choose the optimal number of clusters and interpret them. Statistical methods to aid in that decision exist, but they only point to the optimal statistical solution, which is not always the physically realistic solution when the user knows what they are looking for. Of course, by forcing every observation into a few predefined numbers of groups some information is lost, but patterns remain visible and can be studied.

Particle size distributions were grouped in three clusters, each having a main mode corresponding to one of the three main modes of submicron particles, namely the nucleation, Aitken, and accumulation mode. It was found dividing the data into more clusters would only show different stages of particle growth while having one cluster for each main mode allowed us to investigate processes related to specific parts of the size distribution. Clustering was done on the data for the whole cruise and on the data divided by latitude band or by leg. Three clusters were found to be optimal for every case except for leg 3, where four clusters were necessary to distinguish and separate a nucleation mode from an Aitken mode (see Appendix, Figures A.5 & A.6). Size distributions south of 60°S showed a more pronounced bimodal distribution and north of 60°S a more unimodal distribution for the Aitken cluster. Latitudes between 30 and 60°S are characterised by an open ocean with strong westerly winds unhindered by land called the Southern Ocean westerly wind belt. Latitudes south of 60°S are characterized by the Antarctic continent's coastal waters, with extensive seasonal sea ice.

In-situ meteorological variables, such as wind speed, solar radiation, or temperature, did not seem to influence the cluster type. This prompted the investigation of air mass history to see if it could more readily explain cluster type and CCN concentrations.

We found the MBL was an important source of Aitken clusters and even more of nucleation clusters. While other studies have found that NPF events in the MBL are frequent (Brean et al., 2021; Jokinen et al., 2018; Weller et al., 2015), the main hypothesis is that the majority of nucleation and Aitken mode particles are formed in FT and are subsequently subsided into the MBL (McCoy et al., 2021; Sanchez et al., 2021). Moreover, we found that trajectories coming from the marginal sea ice zone predominantly drive the nucleation cluster in the MBL, confirming their importance in summertime for producing sufficient biogenic gas concentration (Yan et al., 2020) which, coupled with low condensational sink conditions, can be a source for NPF.

When studying the influence of the ancillary variables on CCN, we found increased wind speed did not correlate with increased N_{CCN} at 0.2 % supersaturation, with median values of N_{CCN} being lower for high wind speed (above 12 m s^{-1}) and the highest median values north of 60 °S happening at very low wind speed (<4 m s^{-1}). Wind, and its effect on wave formation, is thought to be a primary driver of SSA production. SSA's importance for marine CCN is disputed with some recent studies showing that contribution could have been underestimated especially for particles in the Aitken mode (Xu et al., 2022), with Fossum et al. (2018) showing contributions of up to 100% of SSA to CCN in high wind conditions (above 16 m s^{-1}). Other studies, however, found similar results to ours over the summertime SO (McCoy et al., 2021; Quinn et al., 2017) and that SSA does not contribute greatly to CCN. They also find a lack of correlation between wind and nucleation and Aitken mode particles and explain the increased Aitken cluster prevalence we observed over the SO to entrainment from the FT with subsequent growth in the MBL.

Datapoints South of 60 °S were characterised by a high prevalence of the accumulation cluster. Increasing median N_{CCN} concentrations were found for higher gas-phase MSA, sulphate and chloride concentrations as well as for lower relative humidity. The high MSA and sulphate concentration are consistent with the hypothesis that in summertime biogenic gases are entrained into the FT where they drive particle nucleation and growth in the FT. Higher median N_{CCN} concentrations were also found for trajectories from the FT, associated with cold fronts. FT airmasses with low relative humidity and high median N_{CCN} are most likely due to Antarctic katabatic winds (Lachlan-Cope et al., 2020), although this was not seen in the back-trajectory data due to low prevalence of data points from the Antarctic continent They are expected to bring colder and drier air masses than those with residence time over the MBL and are often from a free tropospheric origin (Chambers et al., 2018; Simmons et al., 2021). These findings are in accordance with Schmale et al. (2019), who hypothesised that the more pronounced bimodal PNSDs of clusters for leg 2 (which can be approximated by latitudes south of 60 °S) could be due in part to particles growing through condensation or heterogeneous chemistry inside clouds of sulphate and MSA and in part through mass acquisition by repeated cloud processing due to the dry katabatic winds frequently evaporating clouds.

In this thesis, we observed similar findings to the literature regarding the importance of marginal sea ice zones in driving NPF. We found a stronger importance of the MBL as a source of nucleation and Aitken mode particles as opposed to the main hypothesis of the majority of these modes being formed in the FT. Future studies could investigate if and under which conditions particles could grow to accumulation mode sizes directly in the MBL. Accumulation mode particles were shown to have a predominantly FT origin especially south of 60 °S. Observed geographical differences in particle size distributions imply particle formation processes should not be considered to be equivalent throughout the Southern Ocean and support the importance of air mass history analysis as a tool to identify regions driving different processes. Vertical history of back-trajectories could also be used to clarify the importance of the entrainment of air masses in the FT, and their subsequent subsidence, in driving CCN concentration or accumulation mode clusters in the MBL. Studies that can measure localised processes over the entirety of the Southern Ocean are crucial due to its high heterogeneity and are necessary to improve models that can constrain the impact of aerosols on present and future climates.

References

Albrecht, B.A., 1989. Aerosols, Cloud Microphysics, and Fractional Cloudiness. Science 245, 1227–1230. https://doi.org/10.1126/science.245.4923.1227

Anderson, H.J., Moy, C.M., Vandergoes, M.J., Nichols, J.E., Riesselman, C.R., Van Hale, R., 2018. Southern Hemisphere westerly wind influence on southern New Zealand hydrology during the Lateglacial and Holocene. J. Quat. Sci. 33, 689–701. https://doi.org/10.1002/jqs.3045

Arthur, D., Vassilvitskii, S., 2007. K-Means++: The Advantages of Careful Seeding. Presented at the Proc. of the Annu. ACM-SIAM Symp. on Discrete Algorithms, pp. 1027–1035. https://doi.org/10.1145/1283383.1283494

Baccarini, A., Dommen, J., Lehtipalo, K., Henning, S., Modini, R.L., Gysel-Beer, M., Baltensperger, U., Schmale, J., 2021a. Low-Volatility Vapors and New Particle Formation Over the Southern Ocean During the Antarctic Circumnavigation Expedition. J. Geophys. Res. Atmospheres 126. https://doi.org/10.1029/2021JD035126

Baccarini, A., Dommen, J., Lehtipalo, K., Modini, R.L., Gysel-Beer, M., Baltensperger, U., Schmale, J., 2021b. Supporting Information for "Low-volatility vapors and new particle formation over the Southern Ocean during the Antarctic Circumnavigation Expedition" 27.

Barker, P.F., Filippelli, G.M., Florindo, F., Martin, E.E., Scher, H.D., 2007. Onset and role of the Antarctic Circumpolar Current. Deep Sea Res. Part II Top. Stud. Oceanogr. 54, 2388–2398. https://doi.org/10.1016/j.dsr2.2007.07.028

Becagli, S., Barbaro, E., Bonamano, S., Caiazzo, L., di Sarra, A., Feltracco, M., Grigioni, P., Heintzenberg, J., Lazzara, L., Legrand, M., Madonia, A., Marcelli, M., Melillo, C., Meloni, D., Nuccio, C., Pace, G., Park, K.-T., Preunkert, S., Severi, M., Vecchiato, M., Zangrando, R., Traversi, R., 2022. Factors controlling atmospheric DMS and its oxidation products (MSA and $nssSO_4^{2-}$) in the aerosol at Terra Nova Bay, Antarctica. Atmospheric Chem. Phys. Discuss. 1–29. https://doi.org/10.5194/acp-2022-195

Beck, L.J., Sarnela, N., Junninen, H., Hoppe, C.J.M., Garmash, O., Bianchi, F., Riva, M., Rose, C., Peräkylä, O., Wimmer, D., Kausiala, O., Jokinen, T., Ahonen, L., Mikkilä, J., Hakala, J., He, X.-C., Kontkanen, J., Wolf, K.K.E., Cappelletti, D., Mazzola, M., Traversi, R., Petroselli, C., Viola, A.P., Vitale, V., Lange, R., Massling, A., Nøjgaard, J.K., Krejci, R., Karlsson, L., Zieger, P., Jang, S., Lee, K., Vakkari, V., Lampilahti, J., Thakur, R.C., Leino, K., Kangasluoma, J., Duplissy, E.-M., Siivola, E., Marbouti, M., Tham, Y.J., Saiz-Lopez, A., Petäjä, T., Ehn, M., Worsnop, D.R., Skov, H., Kulmala, M., Kerminen, V.-M., Sipilä, M., 2021. Differing Mechanisms of New Particle Formation at Two Arctic Sites. Geophys. Res. Lett. 48, e2020GL091334. https://doi.org/10.1029/2020GL091334

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

Beddows, D.C.S., Dall'Osto, M., Harrison, R.M., Kulmala, M., Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A.M., Sellegri, K., Birmili, W., Bukowiecki, N., Weingartner, E., Baltensperger, U., Zdimal, V., Zikova, N., Putaud, J.-P., Marinoni, A., Tunved, P., Hansson, H.-C., Fiebig, M., Kivekäs, N., Swietlicki, E., Lihavainen, H., Asmi, E., Ulevicius, V., Aalto, P.P., Mihalopoulos, N., Kalivitis, N., Kalapov, I., Kiss, G., de Leeuw, G., Henzing, B., O'Dowd, C., Jennings, S.G., Flentje, H., Meinhardt, F., Ries, L., Denier van der Gon, H. a. C., Visschedijk, A.J.H., 2014. Variations in tropospheric submicron particle size distributions across the European continent

2008–2009. Atmospheric Chem. Phys. 14, 4327–4348. https://doi.org/10.5194/acp-14-4327-2014

Bezdek, J., Pal, N., 1995. Cluster Validation with Generalized Dunn's Indices. pp. 190–193. https://doi.org/10.1109/ANNES.1995.499469

Boucher, O., Randall, D., Artaxo, P., 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Brean, J., Dall'Osto, M., Simó, R., Shi, Z., Beddows, D.C.S., Harrison, R.M., 2021. Open ocean and coastal new particle formation from sulfuric acid and amines around the Antarctic Peninsula. Nat. Geosci. 14, 383–388. https://doi.org/10.1038/s41561-021-00751-y

Carslaw, K.S., Lee, L.A., Reddington, C.L., Pringle, K.J., Rap, A., Forster, P.M., Mann, G.W., Spracklen, D.V., Woodhouse, M.T., Regayre, L.A., Pierce, J.R., 2013. Large contribution of natural aerosols to uncertainty in indirect forcing. Nature 503, 67–71. https://doi.org/10.1038/nature12674

Chambers, S.D., Preunkert, S., Weller, R., Hong, S.-B., Humphries, R.S., Tositti, L., Angot, H., Legrand, M., Williams, A.G., Griffiths, A.D., Crawford, J., Simmons, J., Choi, T.J., Krummel, P.B., Molloy, S., Loh, Z., Galbally, I., Wilson, S., Magand, O., Sprovieri, F., Pirrone, N., Dommergue, A., 2018. Characterizing Atmospheric Transport Pathways to Antarctica and the Remote Southern Ocean Using Radon-222. Front. Earth Sci. 6.

Chen, Q., Sherwen, T., Evans, M., Alexander, B., 2018. DMS oxidation and sulfur aerosol formation in the marine troposphere: a focus on reactive halogen and multiphase chemistry. Atmospheric Chem. Phys. 18, 13617–13637. https://doi.org/10.5194/acp-18-13617-2018

Curtius, J., 2009. Nucleation of atmospheric particles. Eur. Phys. J. Conf. 1, 199–209. https://doi.org/10.1140/epjconf/e2009-00921-0

Dall'Osto, M., Beddows, D.C.S., Tunved, P., Harrison, R.M., Lupi, A., Vitale, V., Becagli, S., Traversi, R., Park, K.-T., Yoon, Y.J., Massling, A., Skov, H., Lange, R., Strom, J., Krejci, R., 2019. Simultaneous measurements of aerosol size distributions at three sites in the European high Arctic. Atmospheric Chem. Phys. 19, 7377–7395. https://doi.org/10.5194/acp-19-7377-2019

Dinoi, A., Conte, M., Grasso, F.M., Contini, D., 2020. Long-Term Characterization of Submicron Atmospheric Particles in an Urban Background Site in Southern Italy. Atmosphere 11, 334. https://doi.org/10.3390/atmos11040334

Fossum, K.N., Ovadnevaite, J., Ceburnis, D., Dall'Osto, M., Marullo, S., Bellacicco, M., Simó, R., Liu, D., Flynn, M., Zuend, A., O'Dowd, C., 2018. Summertime Primary and Secondary Contributions to Southern Ocean Cloud Condensation Nuclei. Sci. Rep. 8, 13844. https://doi.org/10.1038/s41598-018-32047-4

Glantschnig, W.J., Chen, S.-H., 1981. Light scattering from water droplets in the geometrical optics approximation. Appl. Opt. 20, 2499–2509. https://doi.org/10.1364/AO.20.002499

Gondwe, M., Krol, M., Gieskes, W., Klaassen, W., de Baar, H., 2003. The contribution of ocean-leaving DMS to the global atmospheric burdens of DMS, MSA, SO2, and NSS SO4=. Glob. Biogeochem. Cycles 17. https://doi.org/10.1029/2002GB001937

Hamilton, D.S., Lee, L.A., Pringle, K.J., Reddington, C.L., Spracklen, D.V., Carslaw, K.S., 2014. Occurrence of pristine aerosol environments on a polluted planet. Proc. Natl. Acad. Sci. 111, 18466–18471. https://doi.org/10.1073/pnas.1415440111

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Appl. Stat. 28, 100. https://doi.org/10.2307/2346830

Hodshire, A.L., Campuzano-Jost, P., Kodros, J.K., Croft, B., Nault, B.A., Schroder, J.C., Jimenez, J.L., Pierce, J.R., 2019. The potential role of methanesulfonic acid (MSA) in aerosol formation and growth and the associated radiative forcings. Atmospheric Chem. Phys. 19, 3137–3160. https://doi.org/10.5194/acp-19-3137-2019

Hoffmann, E.H., Heinold, B., Kubin, A., Tegen, I., Herrmann, H., 2021. The Importance of the Representation of DMS Oxidation in Global Chemistry-Climate Simulations. Geophys. Res. Lett. 48, e2021GL094068. https://doi.org/10.1029/2021GL094068

Hoppel, W.A., Frick, G.M., 1990. Submicron aerosol size distributions measured over the tropical and South Pacific. Atmospheric Environ. Part Gen. Top. 24, 645–659. https://doi.org/10.1016/0960-1686(90)90020-N

Huang, Y., Dickinson, R.E., Chameides, W.L., 2006. Impact of aerosol indirect effect on surface temperature over East Asia. Proc. Natl. Acad. Sci. 103, 4371–4376. https://doi.org/10.1073/pnas.0504428103

Humphries, R.S., Klekociuk, A.R., Schofield, R., Keywood, M., Ward, J., Wilson, S.R., 2016. Unexpectedly high ultrafine aerosol concentrations above East Antarctic sea ice. Atmospheric Chem. Phys. 16, 2185–2206. https://doi.org/10.5194/acp-16-2185-2016

Hussein, T., Puustinen, A., Aalto, P.P., Mäkelä, J.M., Hämeri, K., Kulmala, M., 2004. Urban aerosol number size distributions. Atmospheric Chem. Phys. 4, 391–411. https://doi.org/10.5194/acp-4-391-2004

Intra, Assoc.Prof.Dr.P., Tippayawong, N., 2007. An overview of aerosol particle sensors for size distribution measurement. Maejo Int. J. Sci. Technol. 1, 120–136.

Jokinen, T., Sipilä, M., Kontkanen, J., Vakkari, V., Tisler, P., Duplissy, E.-M., Junninen, H., Kangasluoma, J., Manninen, H.E., Petäjä, T., Kulmala, M., Worsnop, D.R., Kirkby, J., Virkkula, A., Kerminen, V.-M., 2018. Ion-induced sulfuric acid–ammonia nucleation drives particle formation in coastal Antarctica. Sci. Adv. 4, eaat9744. https://doi.org/10.1126/sciadv.aat9744

Jung, J., Hong, S.-B., Chen, M., Hur, J., Jiao, L., Lee, Y., Park, K., Hahm, D., Choi, J.-O., Yang, E.J., Park, J., Kim, T.-W., Lee, S., 2019. Characteristics of biogenically-derived aerosols over the Amundsen Sea, Antarctica (preprint). Aerosols/Field Measurements/Troposphere/Chemistry (chemical composition and reactions). https://doi.org/10.5194/acp-2019-133

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., 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., Kulmala, M., 2011. Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation. Nature 476, 429–433. https://doi.org/10.1038/nature10343

Lachlan-Cope, T., Beddows, D.C.S., Brough, N., Jones, A.E., Harrison, R.M., Lupi, A., Yoon, Y.J., Virkkula, A., Dall'Osto, M., 2020. On the annual variability of Antarctic aerosol size distributions at Halley Research Station. Atmospheric Chem. Phys. 20, 4461–4476. https://doi.org/10.5194/acp-20-4461-2020

Lamy, F., Chiang, J.C.H., Martínez-Méndez, G., Thierens, M., Arz, H.W., Bosmans, J., Hebbeln, D., Lambert, F., Lembke-Jene, L., Stuut, J.-B., 2019. Precession modulation of the South Pacific westerly wind belt over the past million years. Proc. Natl. Acad. Sci. 116, 23455–23460. https://doi.org/10.1073/pnas.1905847116

Landwehr, S., Volpi, M., Haumann, F.A., Robinson, C.M., Thurnherr, I., Ferracci, V., Baccarini, A., Thomas, J., Gorodetskaya, I., Tatzelt, C., Henning, S., Modini, R.L., Forrer, H.J., Lin, Y., Cassar, N., Simó, R., Hassler, C., Moallemi, A., Fawcett, S.E., Harris, N., Airs, R., Derkani, M.H., Alberello, A., Toffoli, A., Chen, G., Rodríguez-Ros, P., Zamanillo, M., Cortés-Greus, P., Xue, L., Bolas, C.G., Leonard, K.C., Perez-Cruz, F., Walton, D., Schmale, J., 2021. Exploring the coupled ocean and atmosphere system with a data science approach applied to observations from the Antarctic Circumnavigation Expedition. Earth Syst. Dyn. 12, 1295–1369. https://doi.org/10.5194/esd-12-1295-2021

Lange, R., Dall'Osto, M., Skov, H., Nøjgaard, J.K., Nielsen, I.E., Beddows, D.C.S., Simo, R., Harrison, R.M., Massling, A., 2018. Characterization of distinct Arctic aerosol accumulation modes and their sources. Atmos. Environ. 183, 1–10. https://doi.org/10.1016/j.atmosenv.2018.03.060

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

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, Ö., Yu, R., Zhou, B. (Eds.), 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

McCoy, I.L., Bretherton, C.S., Wood, R., Twohy, C.H., Gettelman, A., Bardeen, C.G., Toohey, D.W., 2021. Influences of Recent Particle Formation on Southern Ocean Aerosol Variability and Low Cloud Properties. J. Geophys. Res. Atmospheres 126, e2020JD033529. https://doi.org/10.1029/2020JD033529

Merikanto, J., Spracklen, D.V., Mann, G.W., Pickering, S.J., Carslaw, K.S., 2009. Impact of nucleation on global CCN. Atmospheric Chem. Phys. 9, 8601–8616. https://doi.org/10.5194/acp-9-8601-2009

Nair, P.R., Parameswaran, K., Abraham, A., Jacob, S., 2005. Wind-dependence of sea-salt and non-sea-salt aerosols over the oceanic environment. J. Atmospheric Sol.-Terr. Phys. 67, 884–898. https://doi.org/10.1016/j.jastp.2005.02.008

Noble, S.R., Hudson, J.G., 2019. Effects of Continental Clouds on Surface Aitken and Accumulation Modes. J. Geophys. Res. Atmospheres 124, 5479–5502. https://doi.org/10.1029/2019JD030297

Quinn, P.K., Coffman, D.J., Johnson, J.E., Upchurch, L.M., Bates, T.S., 2017. Small fraction of marine cloud condensation nuclei made up of sea spray aerosol. Nat. Geosci. 10, 674–679. https://doi.org/10.1038/ngeo3003

Quinn, P.K., Collins, D.B., Grassian, V.H., Prather, K.A., Bates, T.S., 2015. Chemistry and Related Properties of Freshly Emitted Sea Spray Aerosol [WWW Document]. ACS Publ. https://doi.org/10.1021/cr500713g

Radenz, M., Seifert, P., Baars, H., Floutsi, A.A., Yin, Z., Bühl, J., 2021. Automated time–height-resolved air mass source attribution for profiling remote sensing applications. Atmospheric Chem. Phys. 21, 3015–3033. https://doi.org/10.5194/acp-21-3015-2021

Revell, L.E., Kremser, S., Hartery, S., Harvey, M., Mulcahy, J.P., Williams, J., Morgenstern, O., McDonald, A.J., Varma, V., Bird, L., Schuddeboom, A., 2019. The sensitivity of Southern Ocean aerosols and cloud microphysics to sea spray and sulfate aerosol production in the

HadGEM3-GA7.1 chemistry–climate model. Atmospheric Chem. Phys. 19, 15447–15466. https://doi.org/10.5194/acp-19-15447-2019

Riipinen, I., Pierce, J.R., Yli-Juuti, T., Nieminen, T., Häkkinen, S., Ehn, M., Junninen, H., Lehtipalo, K., Petäjä, T., Slowik, J., Chang, R., Shantz, N.C., Abbatt, J., Leaitch, W.R., Kerminen, V.-M., Worsnop, D.R., Pandis, S.N., Donahue, N.M., Kulmala, M., 2011. Organic condensation: a vital link connecting aerosol formation to cloud condensation nuclei (CCN) concentrations. Atmospheric Chem. Phys. 11, 3865–3878. https://doi.org/10.5194/acp-11-3865-2011

Roberts, G.C., Nenes, A., 2005. A Continuous-Flow Streamwise Thermal-Gradient CCN Chamber for Atmospheric Measurements. Aerosol Sci. Technol. 39, 206–221. https://doi.org/10.1080/027868290913988

Rousseeuw, P., 1987. Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Comput. Appl. Math. 20, 53-65. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sanchez, K.J., Roberts, G.C., Saliba, G., Russell, L.M., Twohy, C., Reeves, J.M., Humphries, R.S., Keywood, M.D., Ward, J.P., McRobert, I.M., 2021. Measurement report: Cloud processes and the transport of biological emissions affect southern ocean particle and cloud condensation nuclei concentrations. Atmospheric Chem. Phys. 21, 3427–3446. https://doi.org/10.5194/acp-21-3427-2021

Schiffer, J.M., Mael, L.E., Prather, K.A., Amaro, R.E., Grassian, V.H., 2018. Sea Spray Aerosol: Where Marine Biology Meets Atmospheric Chemistry. ACS Cent. Sci. 4, 1617–1623. https://doi.org/10.1021/acscentsci.8b00674

Schmale, J., Baccarini, A., Thurnherr, I., Henning, S., Efraim, A., Regayre, L., Bolas, C., Hartmann, M., Welti, A., Lehtipalo, K., Aemisegger, F., Tatzelt, C., Landwehr, S., Modini, R.L., Tummon, F., Johnson, J.S., Harris, N., Schnaiter, M., Toffoli, A., Derkani, M., Bukowiecki, N., Stratmann, F., Dommen, J., Baltensperger, U., Wernli, H., Rosenfeld, D., Gysel-Beer, M., Carslaw, K.S., 2019. AEROSOLS AND THEIR CLIMATE 25.

Schmale, J., Henning, S., Henzing, B., Keskinen, H., Sellegri, K., Ovadnevaite, J., Bougiatioti, A., Kalivitis, N., Stavroulas, I., Jefferson, A., Park, M., Schlag, P., Kristensson, A., Iwamoto, Y., Pringle, K., Reddington, C., Aalto, P., Äijälä, M., Baltensperger, U., Bialek, J., Birmili, W., Bukowiecki, N., Ehn, M., Fjæraa, A.M., Fiebig, M., Frank, G., Fröhlich, R., Frumau, A., Furuya, M., Hammer, E., Heikkinen, L., Herrmann, E., Holzinger, R., Hyono, H., Kanakidou, M., Kiendler-Scharr, A., Kinouchi, K., Kos, G., Kulmala, M., Mihalopoulos, N., Motos, G., Nenes, A., O'Dowd, C., Paramonov, M., Petäjä, T., Picard, D., Poulain, L., Prévôt, A.S.H., Slowik, J., Sonntag, A., Swietlicki, E., Svenningsson, B., Tsurumaru, H., Wiedensohler, A., Wittbom, C., Ogren, J.A., Matsuki, A., Yum, S.S., Myhre, C.L., Carslaw, K., Stratmann, F., Gysel, M., 2017. Collocated observations of cloud condensation nuclei, particle size distributions, and chemical composition. Sci. Data 4, 170003. https://doi.org/10.1038/sdata.2017.3

Seinfeld, J.H., Pandis, S.N., 2016. Atmospheric Chemistry and Physics 1149.

Simmons, J.B., Humphries, R.S., Wilson, S.R., Chambers, S.D., Williams, A.G., Griffiths, A.D., McRobert, I.M., Ward, J.P., Keywood, M.D., Gribben, S., 2021. Summer aerosol measurements over the East Antarctic seasonal ice zone. Atmospheric Chem. Phys. 21, 9497–9513. https://doi.org/10.5194/acp-21-9497-2021

Slonim, N., Aharoni, E., Crammer, K., 2013. Hartigan's K-means versus Lloyd's K-means: is it time for a change? Presented at the Proceedings of the 23rd International Joint Conference on Artificial Intelligence, pp. 1677–1684.

Sprenger, M., Fragkoulidis, G., Binder, H., Croci-Maspoli, M., Graf, P., Grams, C.M., Knippertz, P., Madonna, E., Schemm, S., Škerlak, B., Wernli, H., 2017. Global Climatologies of Eulerian and Lagrangian Flow Features based on ERA-Interim. Bull. Am. Meteorol. Soc. 98, 1739–1748. https://doi.org/10.1175/BAMS-D-15-00299.1

Sprenger, M., Wernli, H., 2015. The LAGRANTO Lagrangian analysis tool – version 2.0. Geosci. Model Dev. 8, 2569–2586. https://doi.org/10.5194/gmd-8-2569-2015

Tatzelt, C., Henning, S., Tummon, F., Hartmann, M., Baccarini, A., Welti, A., Lehtipalo, K., Schmale, J., Modini, R., 2020. Cloud Condensation Nuclei number concentrations over the Southern Ocean during the austral summer of 2016/2017 on board the Antarctic Circumnavigation Expedition (ACE). https://doi.org/10.5281/zenodo.4415495

Tatzelt, C., Henning, S., Welti, A., Baccarini, A., Hartmann, M., Gysel-Beer, M., van Pinxteren, M., Modini, R.L., Schmale, J., Stratmann, F., 2021. Circum-Antarctic abundance and properties of CCN and INP. Atmos Chem Phys Discuss 2021, 1–35. https://doi.org/10.5194/acp-2021-700

Thurnherr, I., Aemisegger, F., Wernli, H., 2020a. Cold and warm temperature advection mask for the Antarctic Circumnavigation Expedition from December 2016 – March 2017. https://doi.org/10.5281/zenodo.3989318

Thurnherr, I., Wernli, H., 2020. Surface cyclone mask for the Antarctic Circumnavigation Expedition from December 2016 – March 2017. https://doi.org/10.5281/zenodo.3974312

Thurnherr, I., Wernli, H., Aemisegger, F., 2020b. 10-day backward trajectories from ECMWF analysis data along the ship track of the Antarctic Circumnavigation Expedition in austral summer 2016/2017. https://doi.org/10.5281/zenodo.4031705

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B Stat. Methodol. 63, 411–423. https://doi.org/10.1111/1467-9868.00293

Tortell, P.D., Long, M.C., 2009. Spatial and temporal variability of biogenic gases during the Southern Ocean spring bloom. Geophys. Res. Lett. 36, L01603. https://doi.org/10.1029/2008GL035819

Twohy, C.H., DeMott, P.J., Russell, L.M., Toohey, D.W., Rainwater, B., Geiss, R., Sanchez, K.J., Lewis, S., Roberts, G.C., Humphries, R.S., McCluskey, C.S., Moore, K.A., Selleck, P.W., Keywood, M.D., Ward, J.P., McRobert, I.M., 2021. Cloud-Nucleating Particles Over the Southern Ocean in a Changing Climate. Earths Future 9, e2020EF001673. https://doi.org/10.1029/2020EF001673

Twomey, S., 1977. The Influence of Pollution on the Shortwave Albedo of Clouds. J. Atmospheric Sci. 34, 1149–1152. https://doi.org/10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2

Uetake, J., Hill, T.C.J., Moore, K.A., DeMott, P.J., Protat, A., Kreidenweis, S.M., 2020. Airborne bacteria confirm the pristine nature of the Southern Ocean boundary layer. Proc. Natl. Acad. Sci. 117, 13275–13282. https://doi.org/10.1073/pnas.2000134117

Walton, D.W.H., Thomas, J., 2018. Cruise Report - Antarctic Circumnavigation Expedition (ACE) 20th December 2016 - 19th March 2017. Zenodo. https://doi.org/10.5281/ZENODO.1443511

Weller, R., Schmidt, K., Teinilä, K., Hillamo, R., 2015. Natural new particle formation at the coastal Antarctic site Neumayer. Atmospheric Chem. Phys. 15, 11399–11410. https://doi.org/10.5194/acp-15-11399-2015

Wiedensohler, A., 1988. An approximation of the bipolar charge distribution for particles in the submicron size range. J. Aerosol Sci. 19, 387–389. https://doi.org/10.1016/0021-8502(88)90278-9

Yan, J., Jung, J., Lin, Q., Zhang, M., Xu, S., Zhao, S., 2020. Effect of sea ice retreat on marine aerosol emissions in the Southern Ocean, Antarctica. Sci. Total Environ. 745, 140773. https://doi.org/10.1016/j.scitotenv.2020.140773

Xu, W., Ovadnevaite, J., Fossum, K.N., Lin, C., Huang, R.-J., Ceburnis, D., O'Dowd, C., 2022. Sea spray as an obscured source for marine cloud nuclei. Nat. Geosci. 15, 282–286. https://doi.org/10.1038/s41561-022-00917-2

Zhao, H., Jiang, X., Du, L., 2017. Contribution of methane sulfonic acid to new particle formation in the atmosphere. Chemosphere 174, 689–699. https://doi.org/10.1016/j.chemosphere.2017.02.040

Appendix

A. Additional figures for the clustering

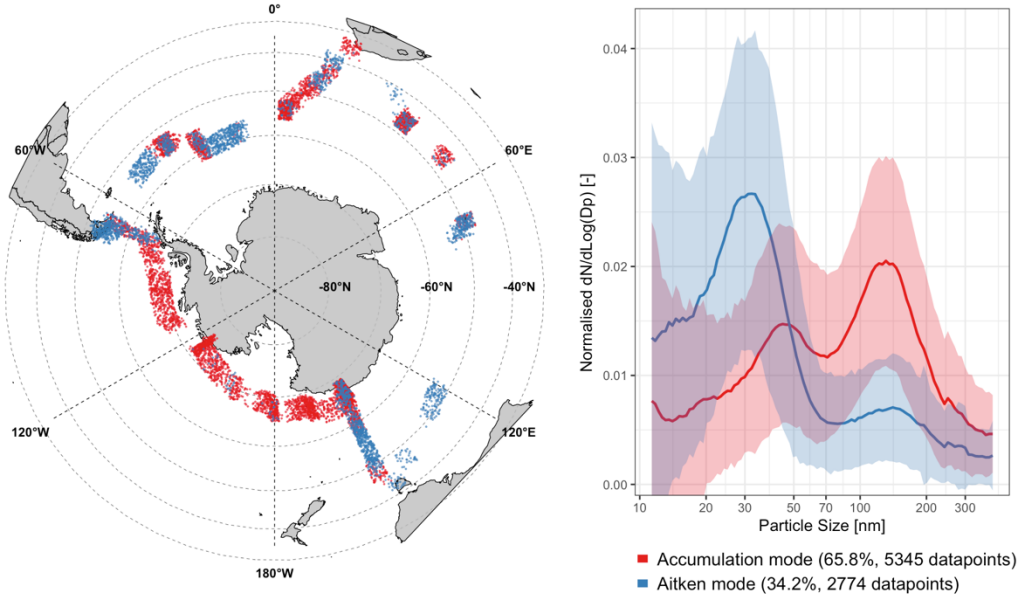

Figure A.1: On the left, the geographical distributions of 2 clusters for the whole cruise. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. On the right, PNSDs of the 3 clusters for particle size between 11 and 400nm

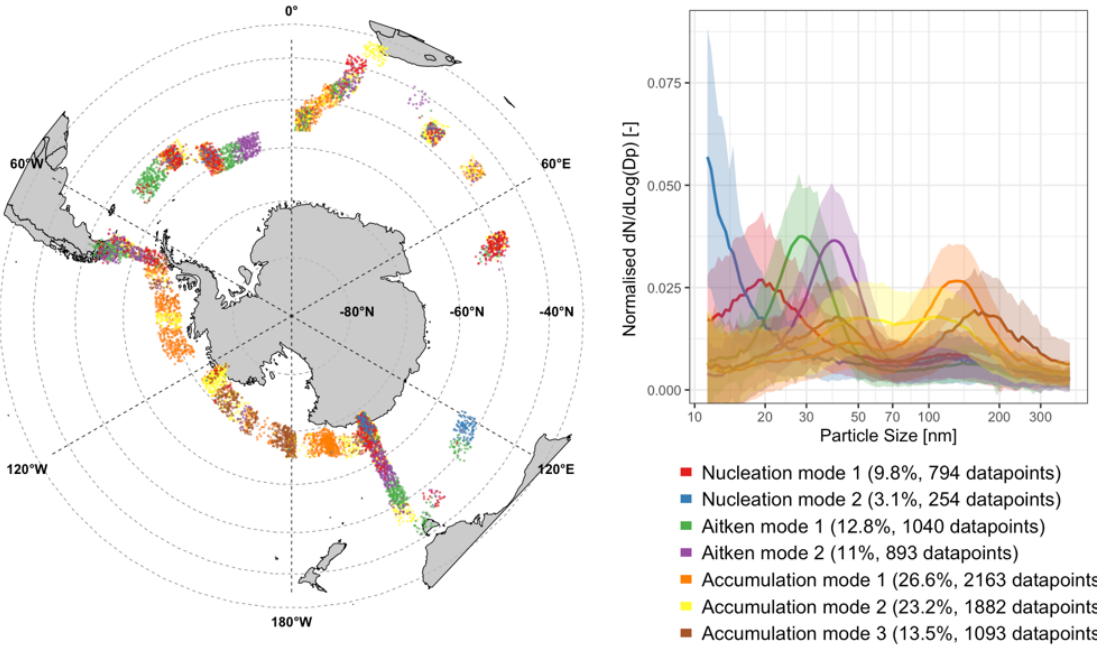

Figure A.2: On the left, the geographical distributions of 7 clusters for the whole cruise. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. On the right, PNSDs of the 3 clusters for particle size between 11 and 400nm. We can see the nucleation mode 2 has very high values for the smallest particle sizes and is probably representative of recent NPF.

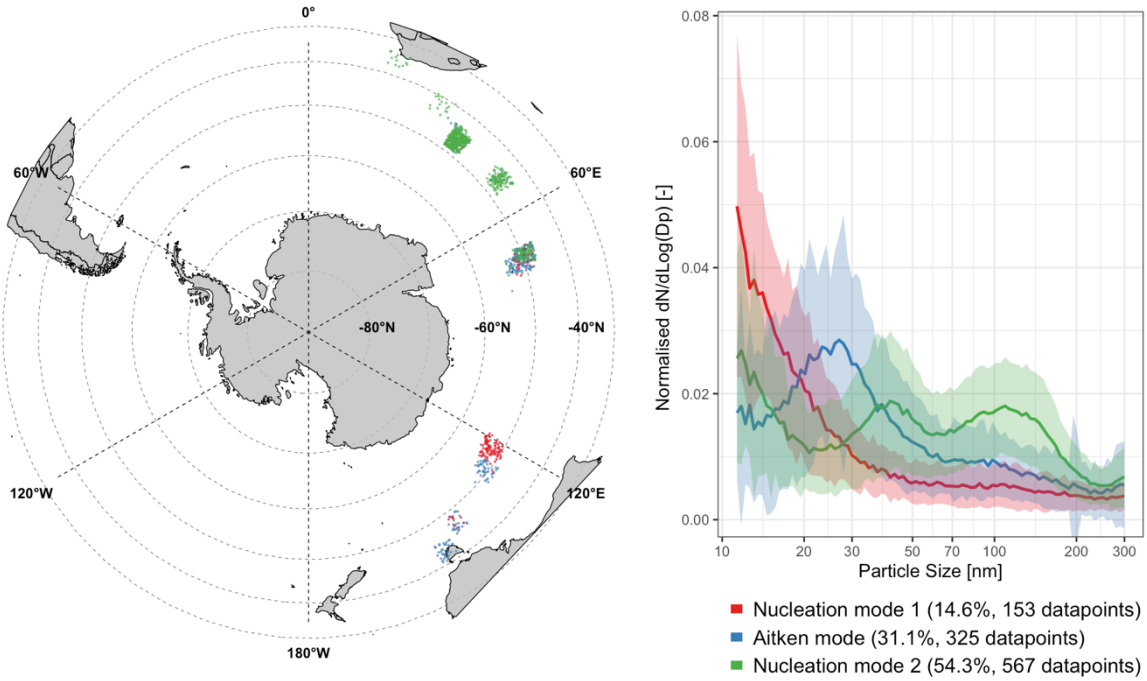

Figure A.3: On the left, the geographical distributions of 3 clusters for leg 1. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. On the right, PNSDs of the 3 clusters for particle size between 11 and 400nm

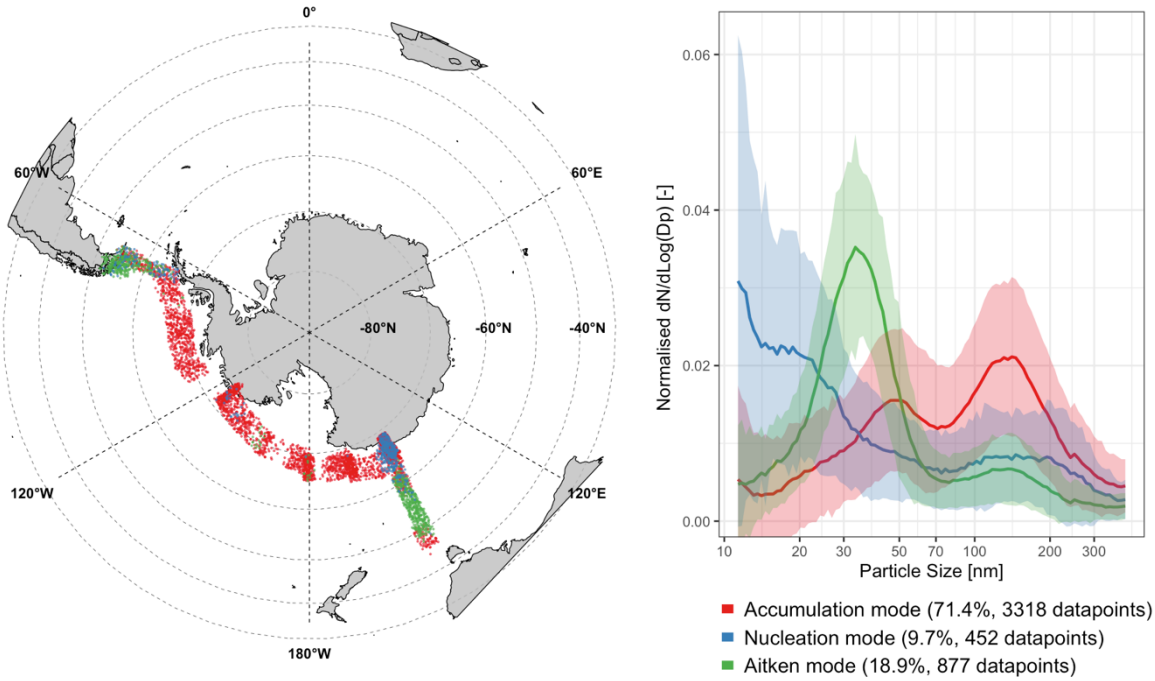

Figure A.4: On the left, the geographical distributions of 3 clusters for leg 2. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. On the right, PNSDs of the 3 clusters for particle size between 11 and 400nm.

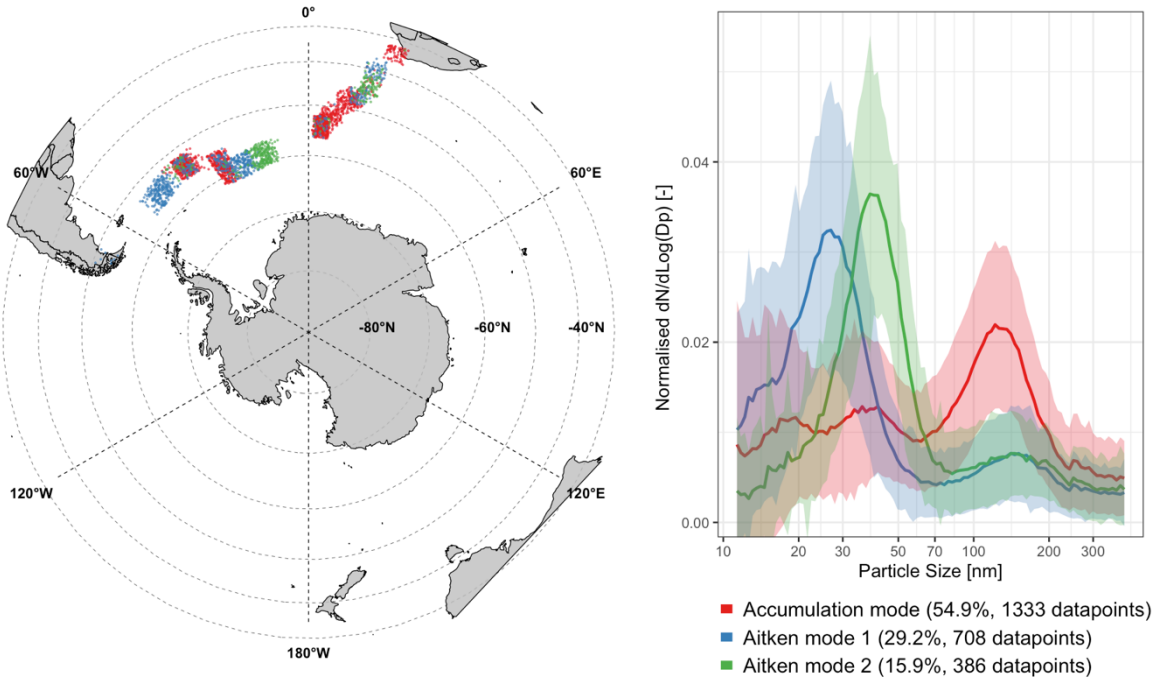

Figure A.5: On the left, the geographical distributions of 3 clusters for leg 3. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. On the right, PNSDs of the 3 clusters for particle size between 11 and 400nm.

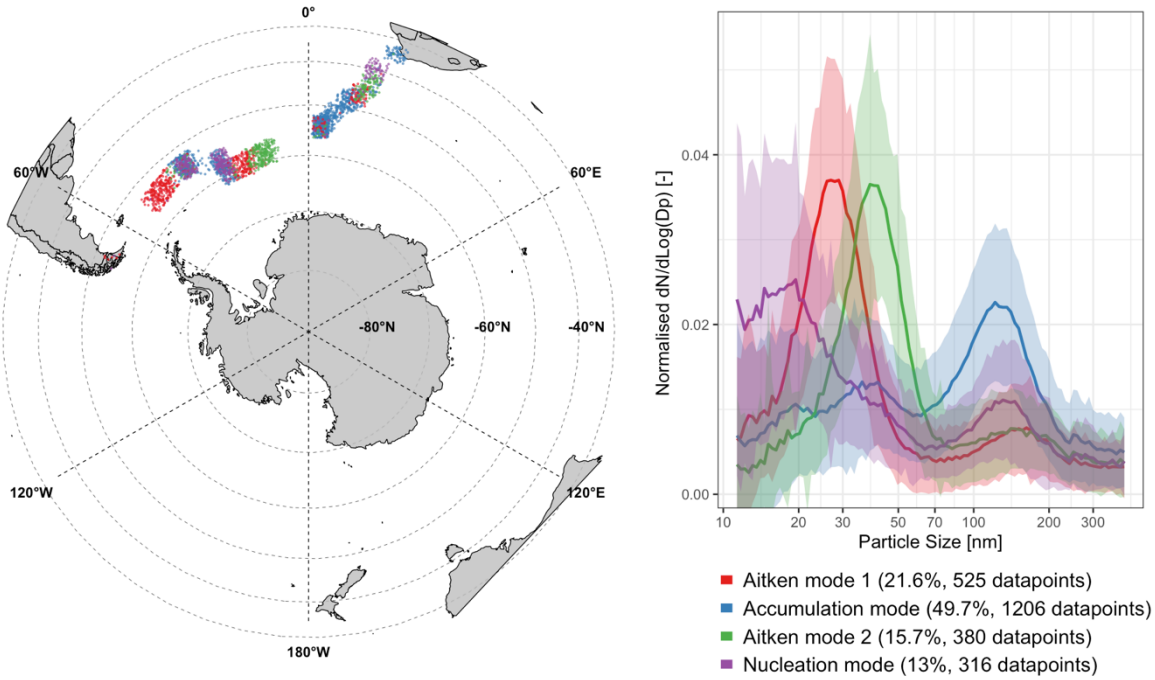

Figure A.6: On the left, the geographical distributions of 4 clusters for leg 3. The points are jittered around the original locations for visibility, hence do not represent the exact cruise tracks. On the right, PNSDs of the 3 clusters for particle size between 11 and 400nm.

B. Additional figures for ancillary variables

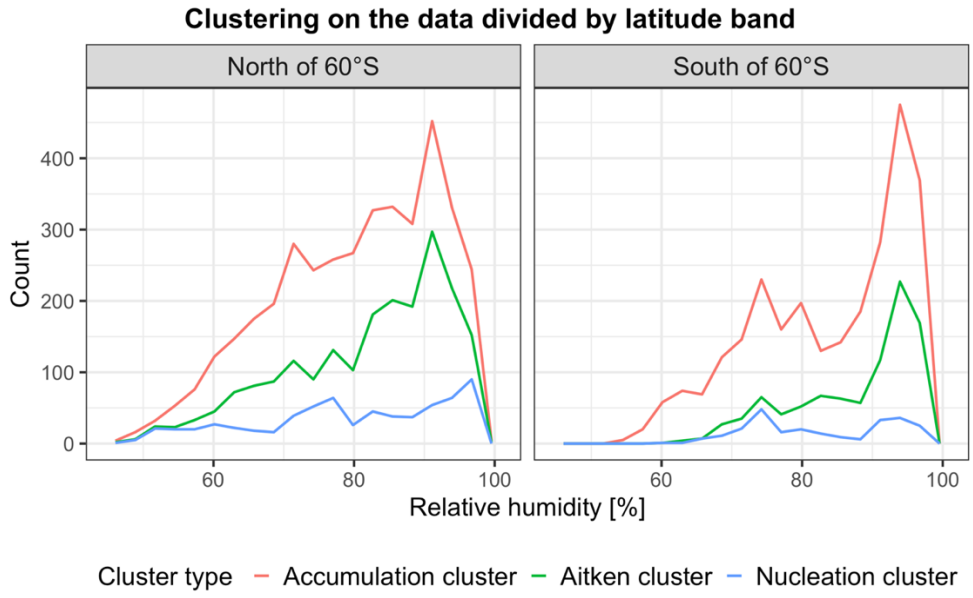

Figure B.1: Frequency polygons of each cluster for relative humidity. Lines link the top of invisible histograms.

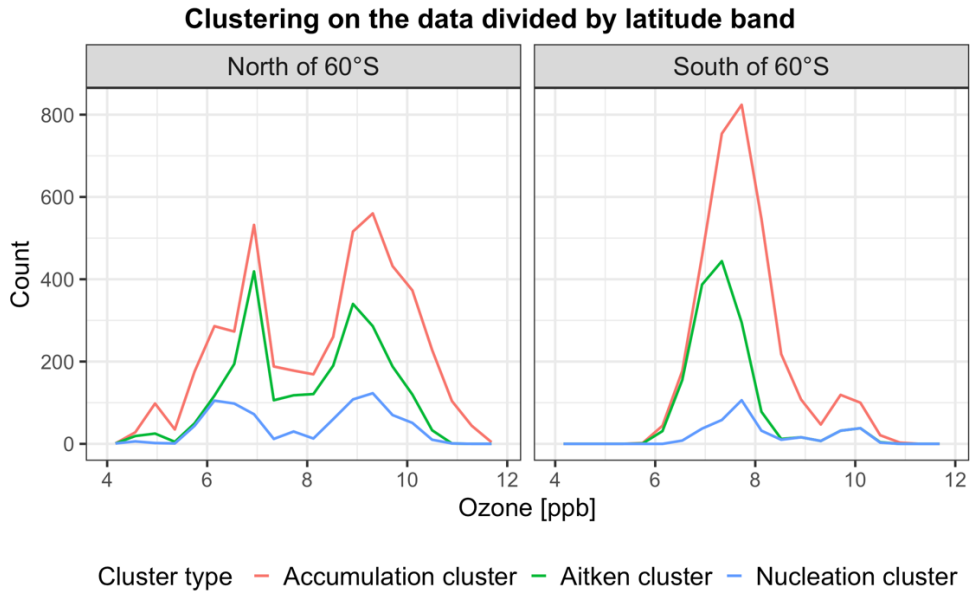

Figure B.2: Frequency polygons of each cluster for ozone. Lines link the top of invisible histograms.

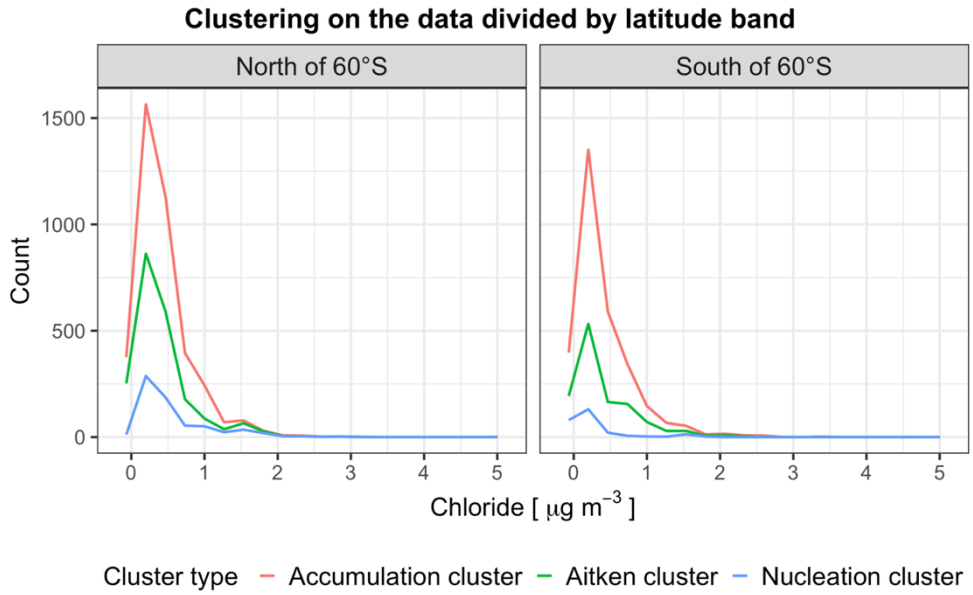

Figure B.3: Frequency polygons of each cluster for chloride. Lines link the top of invisible histograms.

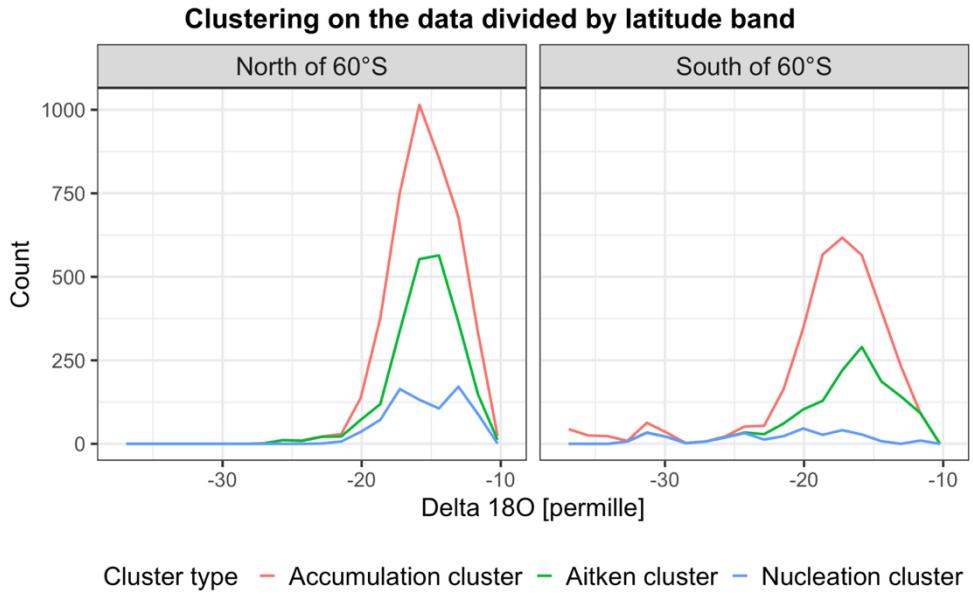

Figure B.4: Frequency polygons of each cluster for $\delta^{18}O$. Lines link the top of invisible histograms.

C. Quality control of the data

Quality checks were performed on the data when it was noticed that some observations showed very high dN/dlogDp values for only one particle size bin in the whole size distribution (Figure C.1). Those high values where equally distributed across all size bins, with usually only one bin size having high values for each observation, which points towards an instrument error and not a pollution event. The observations were present in all legs and were seemed randomly geographically distributed (Figure C.2). Because there were too many instances of such high values to handle them case by case, cut-off values were applied to the data. These cut-off values were chosen after visual appreciation of the distribution, and relatively crude. All dN/dlogDp values above 6000 for leg 1 and 2 and above 4000 for leg 3 were removed. Additionally, values above 2500 for particle sizes above 100 nm, above 4000 for particle size between 20 and 80 nm and above 2000 for particles sizes between 80 and 100 nm were also removed. Additionally, two observations were removed completely due to probably pollution events. The results of the quality check can be seen on Figure C.3.

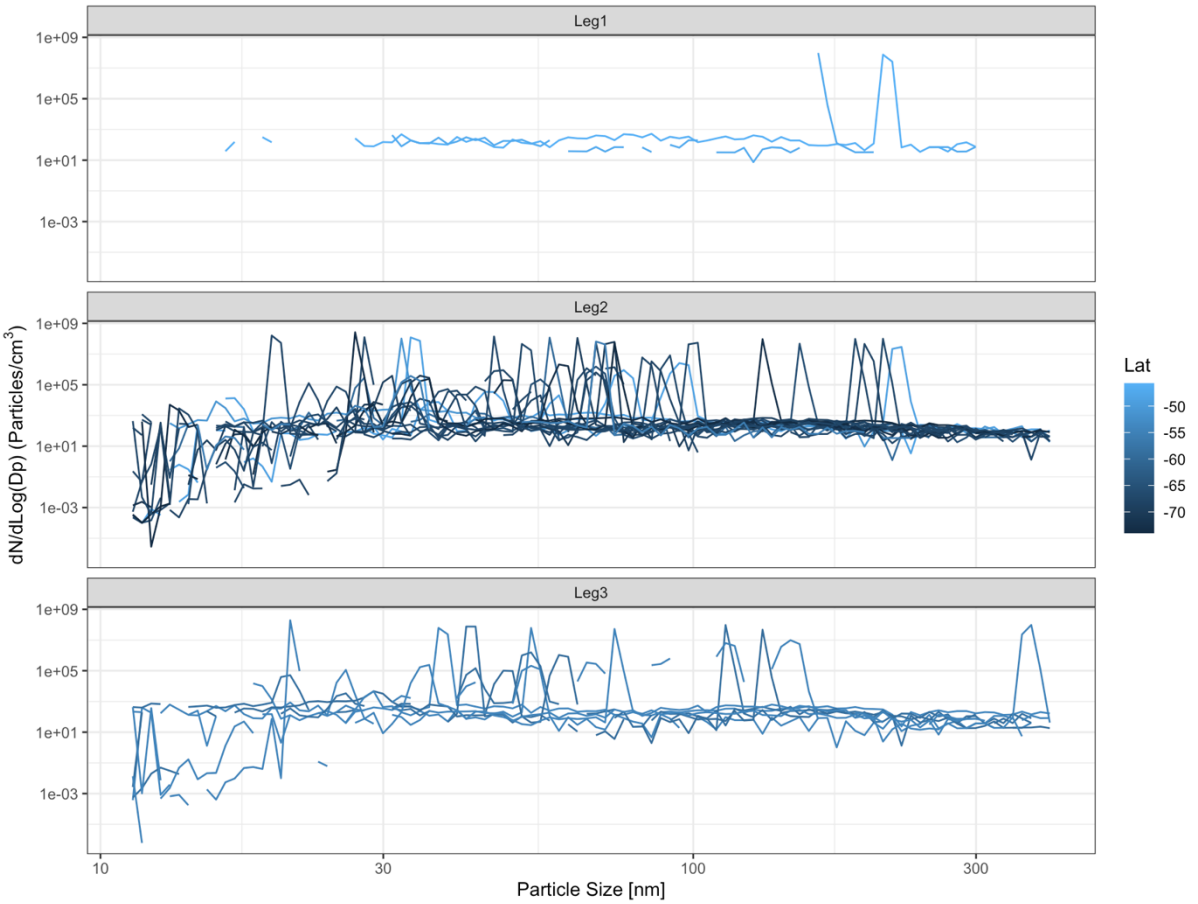

Figure C.1: Particle size distribution where at least one particle size bin had values above 10^5 dN/dlogDp, separated by leg and coloured by latitude.

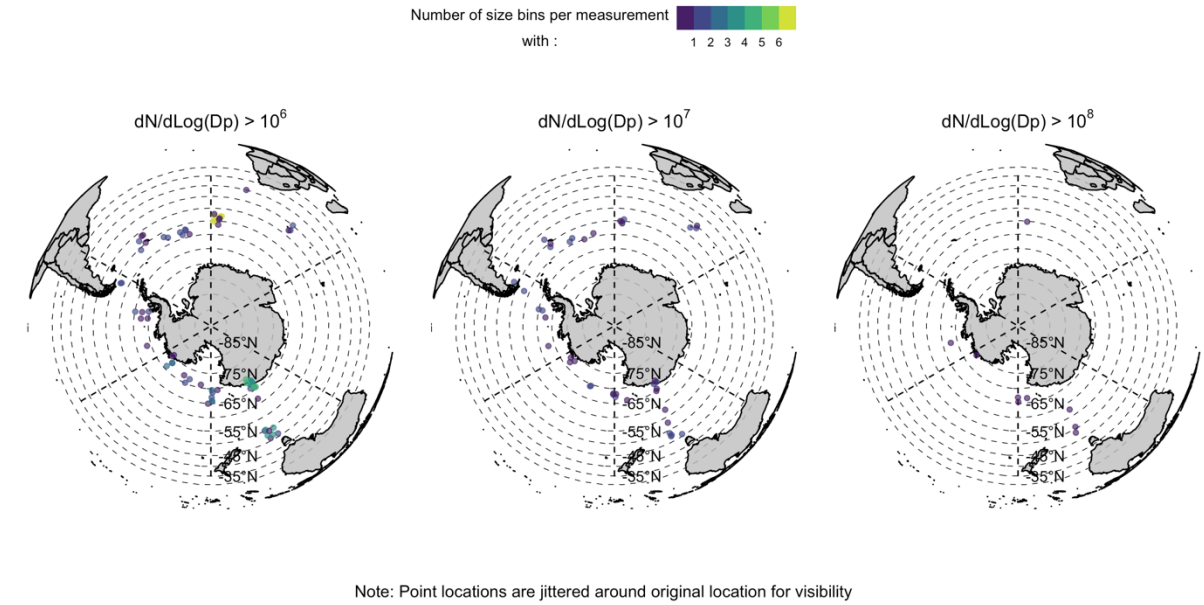

Figure C.2: Map of the number of particle size bins per observation with values above 10^6, 10^7 and 10^8 dN/dlogDp.

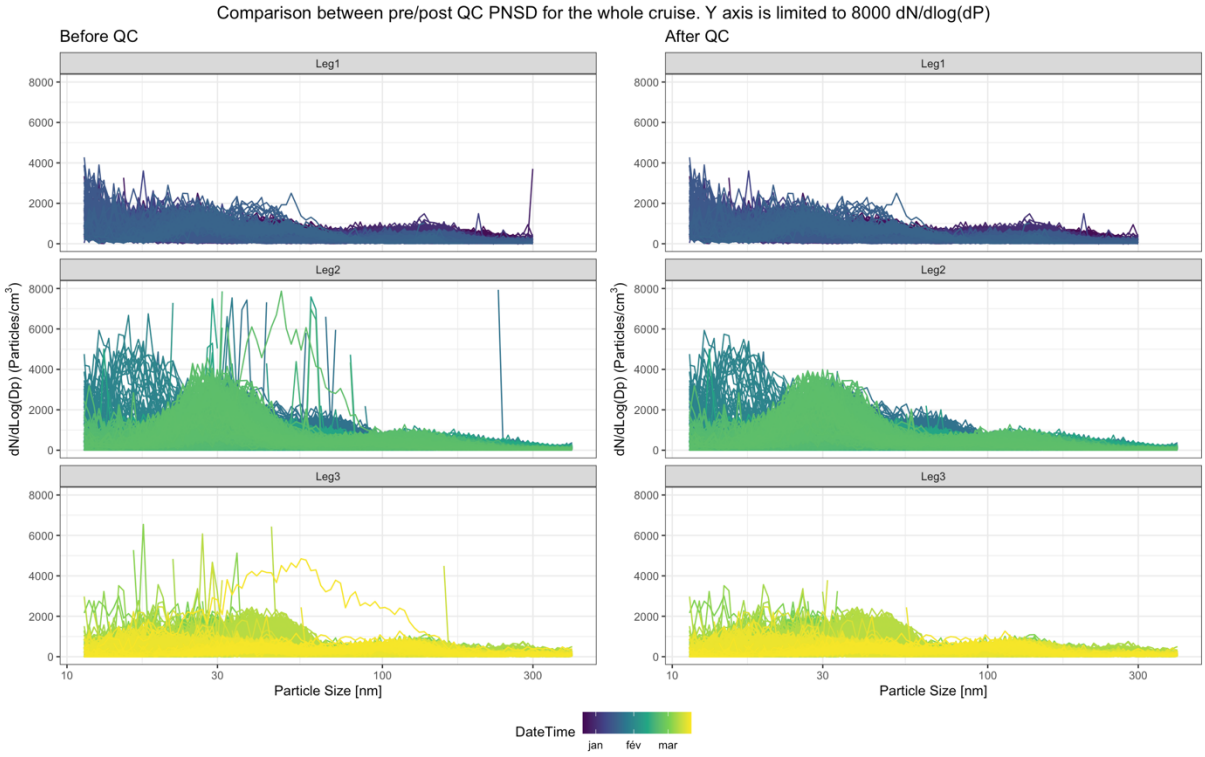

Figure C.3: Particle number size distributions before and after QC for each leg.