U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-ARM-TR-316

MOSAiC Expedition: Automated Contamination Detection Algorithm for ARM Aerosol Measurements

A Sirna
J Loprete
R Trojanowski
O Mayol-Bracero
J Uin
R Ristow Hadlich
M Zawadowicz
D Assanis

May 2025

ARM

DISCLAIMER

DOE/SC-ARM-TR-316

MOSAiC Expedition: Automated Contamination Detection Algorithm for ARM Aerosol Measurements

A Sirna, Stony Brook University (SBU)
J Uin, Brookhaven National Laboratory (BNL)
J Loprete, SBU
R Ristow Hadlich, SBU
R Trojanowski, BNL
M Zawadowicz, BNL
O Mayol-Bracero, BNL
D Assanis, SBU

May 2025

How to cite this document:

Sirna, A, J Uin, J Loprete, R Ristow Hadlich, R Trojanowski, M Zawadowicz, O Mayol-Bracero, and D Assanis. MOSAiC Expedition: Automated Contamination Detection Algorithm for ARM Aerosol Measurements. U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-TR-316.

Work supported by the U.S. Department of Energy,
Office of Science, Office of Biological and Environmental Research

Abstract

The 2019‑2020 Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition (references: ARM July 2023a, ARM August 2023) recorded measurements in the Arctic, crucial to understanding the extent of climate change and modeling the current global atmosphere dynamics, by deploying the research ship *RV Polarstern* to the Arctic Ocean. As part of the expedition, ARM deployed one of its Aerosol Observing Systems (AOS; Uin et al. 2019) on the ship for in situ measurements of properties of atmospheric aerosols and trace gases. Due to the closeness of ship's infrastructure, some of the AOS measurements were affected by the local emissions from the ship, necessitating an investigation into possible data-filtering techniques to improve the quality of the measured data. Six filtering schemes are evaluated of which four are statistical, one is meteorological, and one is based on the microphysical properties of aerosol particles.

The key conclusions are that one of the statistical filters (removing data beyond three standard deviations of the mean) proved to be the most effective at providing the largest change (35.1%) in reported mean for the least (3.2%) amount of data removed. A filter based on microphysical properties was also capable of achieving a similar (34.7%) change in mean as the best statistical filter method, but it achieved this result with a slightly larger amount of data removed (4.0%). While the meteorological-based filtering scheme, which included both the wind speed and (relative) direction, did provide a trend-wise accurate cleaning of the data set, using only the wind direction was much less effective and may have been producing false positives.

Acknowledgments

This research work was financially supported by Brookhaven Science Associates, LLC Subcontract #417696 under U.S. Department of Energy (DOE) Prime Contract No. DE-SC0012704 for the management and operation of Brookhaven National Laboratory. Data were collected from the Multidisciplinary drifting Observatory for the Study of the Arctic Climate (MOSAiC) by the Atmospheric Radiation Measurement (ARM) user facility, a DOE Office of Science user facility managed by the Biological and Environmental Research Program, under expedition number MOSAiC20192020 and project identifier AWI PS122 00.00.

Acronyms and Abbreviations

AOS	Aerosol Observing System
AOSMET	AOS surface meteorology
ARM	Atmospheric Radiation Measurement
BC	black carbon
BNL	Brookhaven National Laboratory
CO-analyzer	carbon monoxide analyzer
CPC	condensation particle counter
DOE	U.S. Department of Energy
ENA	Eastern North Atlantic
ENA-AM	ENA-aerosol mask
MOSAiC	Multidisciplinary Drifting Observatory for the Study of Arctic Climate
PDA	pollution detection algorithm
QA	quality assurance
QC	quality control
SBU	Stony Brook University
SMPS	scanning mobility particle spectrometer
SP2	single-particle soot photometer

Contents

Figures

Tables

1.1 Introduction and Background

The atmospheric composition and chemical properties in the Arctic are changing, making Earth system modeling increasingly difficult (ARM July 2023a). In an effort to quantify these changes, a year-long international expedition to the Arctic was undertaken.

Figure 1. *RV Polarstern*, with the location of the AOS circled. Photo Credit: Manuel Ernst.

The Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC) expedition deployed a number of instruments and measurement systems aboard the German research vessel *RV Polarstern*. (Figure 1.) Among these was a mobile observatory provided by the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) user facility (ARM July 2023a). The observatory was equipped with several instruments measuring the properties of aerosols and trace gases (ARM July 2023b). The data collected from this expedition has been used to support research since the conclusion of the expedition in October 2020; however, there are concerns regarding measurement contamination from the local emissions from the ship. Even though the ship was moored to the ice floe, its engines and other systems were kept running to provide power and heat. As a result, the ship's stack and many air vents were sources of aerosol particles that mixed with the ambient regional air, contaminating the measurements. Additionally, use of snow mobiles and helicopters close to the ship produced unwanted particle emissions. Several research groups have developed procedures to remove data from periods when local contamination was suspected, but their efforts have been focused on specific research interests and the filtering schemes used were tailored to those specific needs. This work aims to generalize the contamination filtering approach and offer tools to the ARM user community to simplify data QA/QC and facilitate the use of ARM MOSAiC data.

1.2 Prior Related Work

The most prominent technique uses statistical analysis of the recorded data to identify outliers by looking at the mean, median, and standard deviations of the data. When the proximity of an airport to the Eastern North Atlantic (ENA) facility of ARM became a concern, a statistical mask was researched and developed using the standard deviation of the data (Gallo et al. 2020). Gallo et al. confirmed through wind direction analysis that total particle concentration was affected by the local sources; high-concentration events were strongly associated with wind directions aligned with major roads and the nearby airport. To

remove these contaminated data points, an ENA-aerosol mask (ENA-AM) was developed using the standard deviation of the condensation particle counter (CPC) data. The application of the ENA-AM removed about 26% of summer main site data, which resulted in an improved value for the coefficient of determination, R2. When comparing the percentage of data flagged for the ENA-AM and a pure wind direction mask, the ENA-AM was able to flag a smaller percentage of data than the wind direction mask and obtain a higher R2 value. The study concluded that the ENA-AM can be used for filtering long-time-series recorded data at remote locations (Gallo et al. 2020). Bukowiecki et al. also explored statistical-based filtering of aerosols measurements in Switzerland; they looked specifically at locations where tourism was high and introduced a supplementary measuring location to compare spatial influence on the data (Bukowiecki et al. 2021). By looking at the parallel measurements taken at both sites, Bukowiecki et al. were able to associate abnormal measurement behavior with specific wind conditions (Bukowiecki et al. 2021). This filtering scheme is able to capture statistically based outliers by looking at the range of data and removing contaminated points in the upper percentile; however, these filters are not guided by other measurements taken and therefore can generate false negative flags – periods that are flagged as contaminated, yet are known good data.

The other main technique explored to remove contamination periods from aerosols measurement time series involves applying a time derivative filter. A flagging algorithm using this technique was developed by Beck et al. and applied to a particle number concentration data set recorded onboard the MOSAiC expedition. The pollution detection algorithm (PDA) developed by Beck et al. is a multi-stage automated filtering program meant to clean aerosol measurements (Beck et al. 2022). The multi-stage filter includes a few optional filters (neighboring-points-based and median-based) that can be applied; however, the main function of the PDA is the time-derivative-based filter. This filter involves taking the magnitude of the time derivative of the aerosol measurement to determine when large and quick changes in the data occur. These sudden changes in the data magnitude are considered as contaminated and therefore flagged by the algorithm. When the performance of the PDA was tested against visual filtering, it flagged a similar number of points, the two techniques flagging 43% and 41%, respectively. The algorithm can capture the high concentration events coming from the ship stack wind direction. These two techniques have been the main focus for removing high-particle-concentration events from aerosol measurements that have been contaminated by local emission sources. The work presented here explores other data analysis techniques and quantify their performance on the ARM MOSAiC data set.

2.0 Investigation of Contamination Periods

This section will outline how the datastreams central to the development of a filtering scheme were chosen, as well as how these datastreams aided in the confirmation of contamination. Based on the datastreams selected, the automation of the filtering scheme is also outlined.

2.1 Datastream Selection

ARM deployed an extensive list of instruments to be used in the expedition, all supporting one of the following major themes: surface energy budget, aerosol properties, cloud properties and atmospheric boundary layer (ARM July 2023a). As this study looks at how the association with the ship has influenced the data set, the aerosol-based measurements are the primary focus – the main data used being particle concentration (Singh and Kuang 2024), size distribution of particles (Singh and Kuang 2024b), and trace gas concentrations (Springston 2015, Sedlacek 2017), as well as wind direction and speed

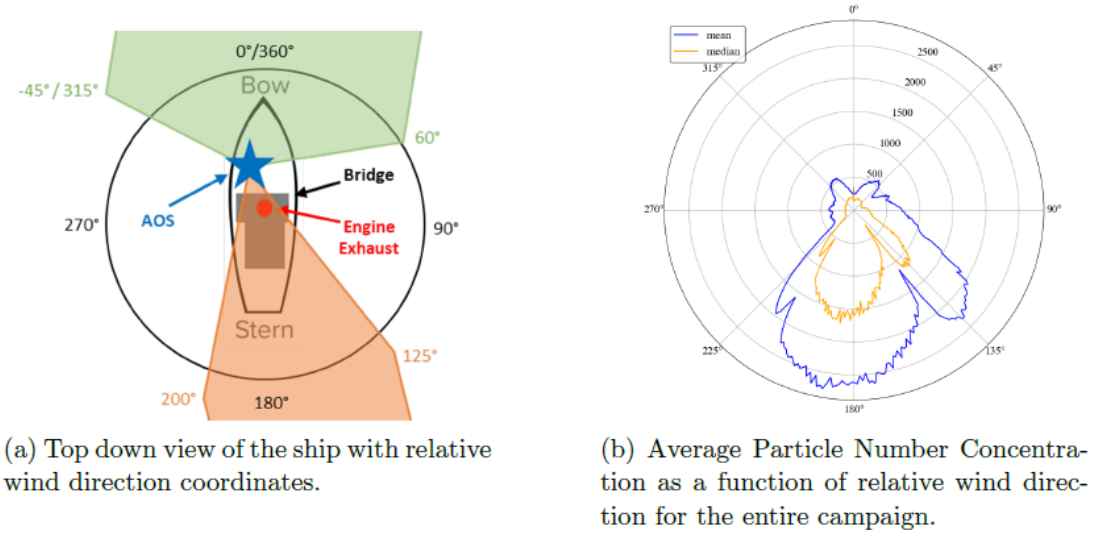
(Kyrouac 2019). These data sets have all been corrected for false zeros caused by intermittent air-intake stack purging (Uin 2024). Table 1 outlines the details for each specific datastream and the relevant instrument used to acquire these measurements. The actual measurement system used onboard during the campaign is listed in the MOSAiC Science Plan (Shupe et al. 2018). All data used in this work are available through the ARM Data Discovery portal that is used to publish all ARM recorded data for public availability (ARM September 2023).

Table 1. Instrumentation names and associated instrumentation (ARM July 2023b).

Data Type	Instrument
Particle concentration (Singh and Kuang 2024)	Condensation particle counter (CPC)
Particle size distribution (Singh and Kuang 2024b)	Scanning mobility particle spectrometer (SMPS)
Wind speed & direction (Kyrouac 2019)	AOS surface meteorology (AOSMET)
Carbon monoxide (Springston 2015)	Carbon monoxide analyzer (CO-analyzer)
Black carbon (Sedlacek 2017)	Single-particle soot photometer (SP2)

2.2 Validation of Suspected Measurement Contamination

Definition of contamination from anthropogenic sources is necessary before proceeding to development of a flagging scheme. Identifying the factors that indicate contamination will help in developing the flagging scheme. Figure 2 (a) shows a top-down view of the ship with the relative positioning of the AOS, the ship's superstructure, and the engine exhaust. The ship is encircled with relative wind direction coordinates (0◦ or "North" direction is aligned to ship's bow). Directions of relatively little contamination concern are highlighted in green and directions of greater potential contamination concern are highlighted in orange. The basis of the schematic relies on the hypothesis that relative wind direction plays a significant role in the contamination taking place; winds coming from behind the ship can carry contamination from various sources on the ship into the AOS sample stack. The wind speed will also play a part, with the possibility that specific speeds can either enhance or diminish this wind direction affect, and that faster winds speeds can dilute any local ship contamination with arctic air. Based on the literature reviewed, particle concentration versus relative wind direction was identified as the data set on which to initially evaluate this hypothesis.



(a) Top down view of the ship with relative wind direction coordinates.

(b) Average Particle Number Concentration as a function of relative wind direction for the entire campaign.

Figure 2. Comparison of ship wind direction orientation and particle concentration.

The scatter plot in Figure 3 shows all the particle concentration data points recorded throughout the campaign with respect to relative recorded wind direction; the points are colored based on the wind speed. As seen in Figure 2, there is a clear and distinct spike in the particle concentration coming from the wind direction of concern. The majority of data points recorded from this direction show high-concentration events at low wind speeds. A polar projection of the average across the year for each degree of wind direction is also determined and presented in Figure 2 (b). These two plots of particle concentration demonstrate that a significant portion of data collected coming from wind directions of 110◦-225◦ was contaminated by local ship pollution, including but not limited to ship engine exhaust and general anthropogenic local activities, confirming the initial hypothesis (Beck et al. 2022).

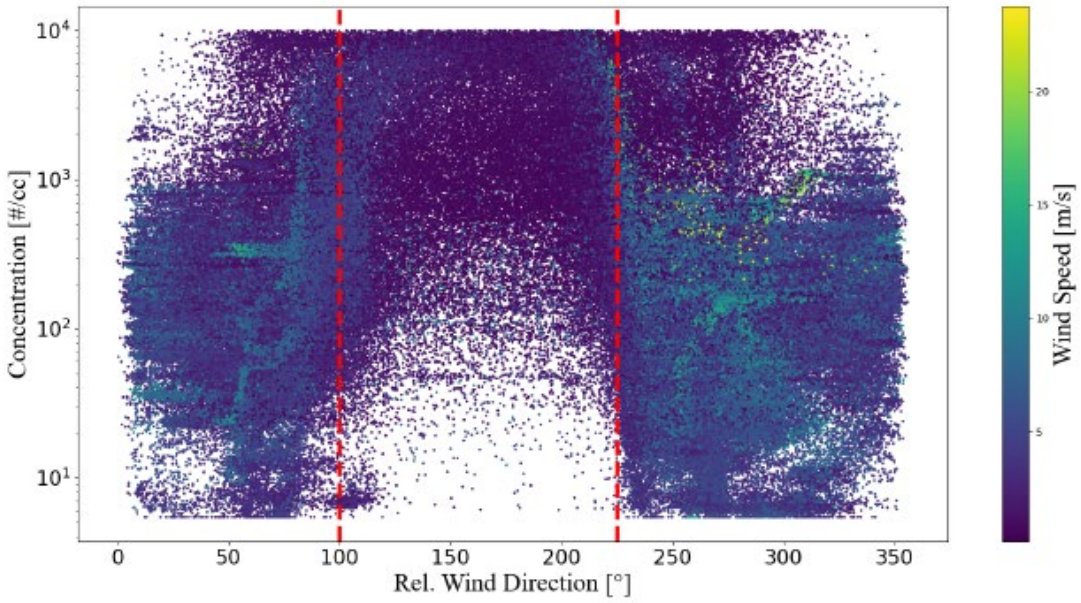


Figure 3. Scatter plot of particle concentration versus relative wind direction colored by wind speed. Particle concentration averaged over one minute. Vertical dashed lines indicate the relative position of the ship's superstructure.

Other measurements, besides particle number concentration, considered as good indicators for ship influence were black carbon and carbon monoxide concentrations, as both are byproducts of combustion. All three of these parameters were evaluated on a monthly polar projection presented in Figures 4, 5, and 6, respectively, to demonstrate how both time and wind direction can serve a role in identifying points of contamination. Weekly scatter plots in polar projection were evaluated, but overall, monthly plots proved to be a better balance between appropriate sample size and refined time scale for identifying localized peak emissions. Examples of these local peaks can be seen in months such as October of 2019 and September of 2020 in Figures 4 and 5, where particle concentration and black carbon both peaked in the direction of concern, giving strong indication of the presence of local contamination. The carbon monoxide data set encountered a much higher level of quality control than other data sets, which is why there is no data for the month of March 2020. From the data remaining in Figure 6, concentric circles can be observed, implying little to no wind direction variation in CO concentrations.

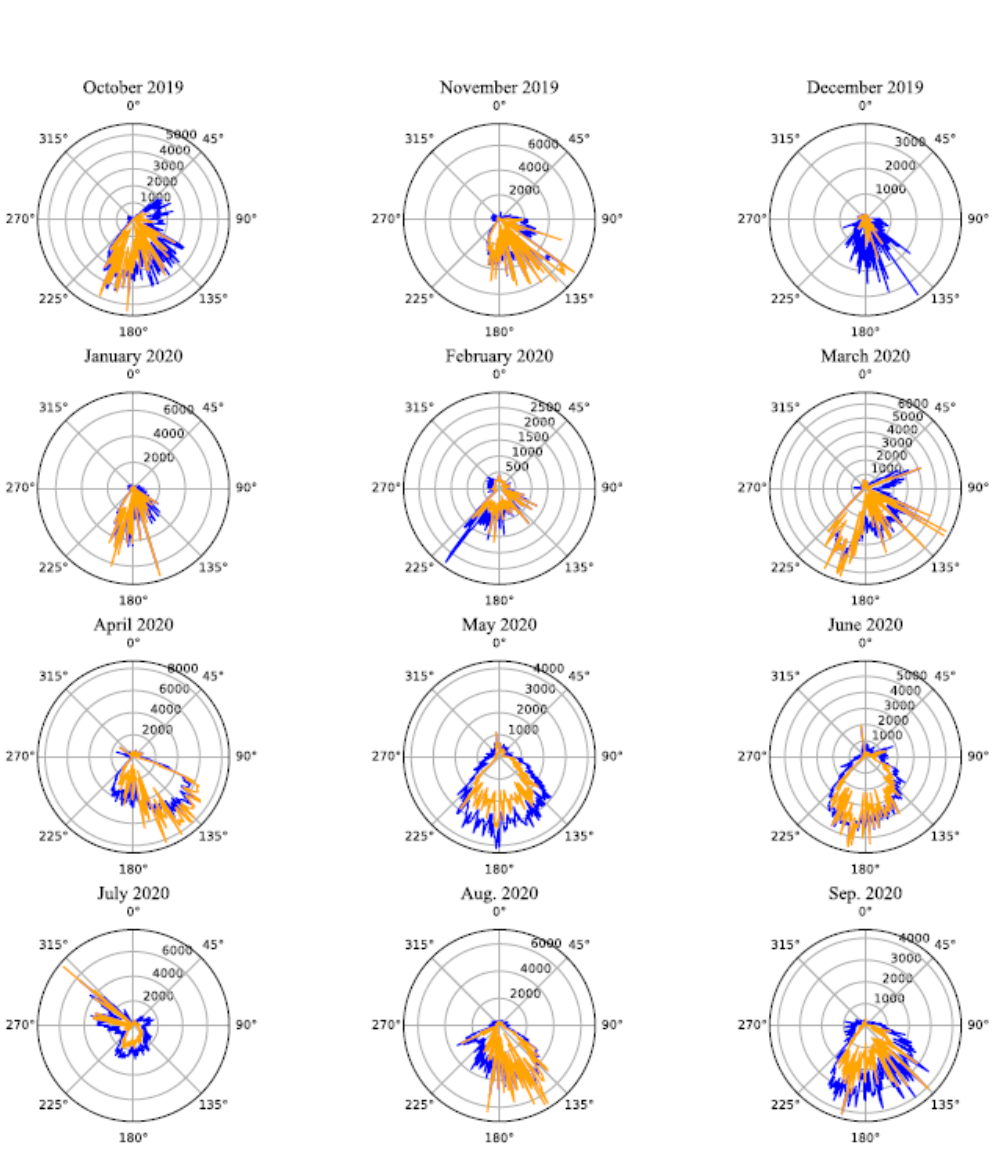


Figure 4. Monthly polar plots of particle concentration in units of #/cm^3.

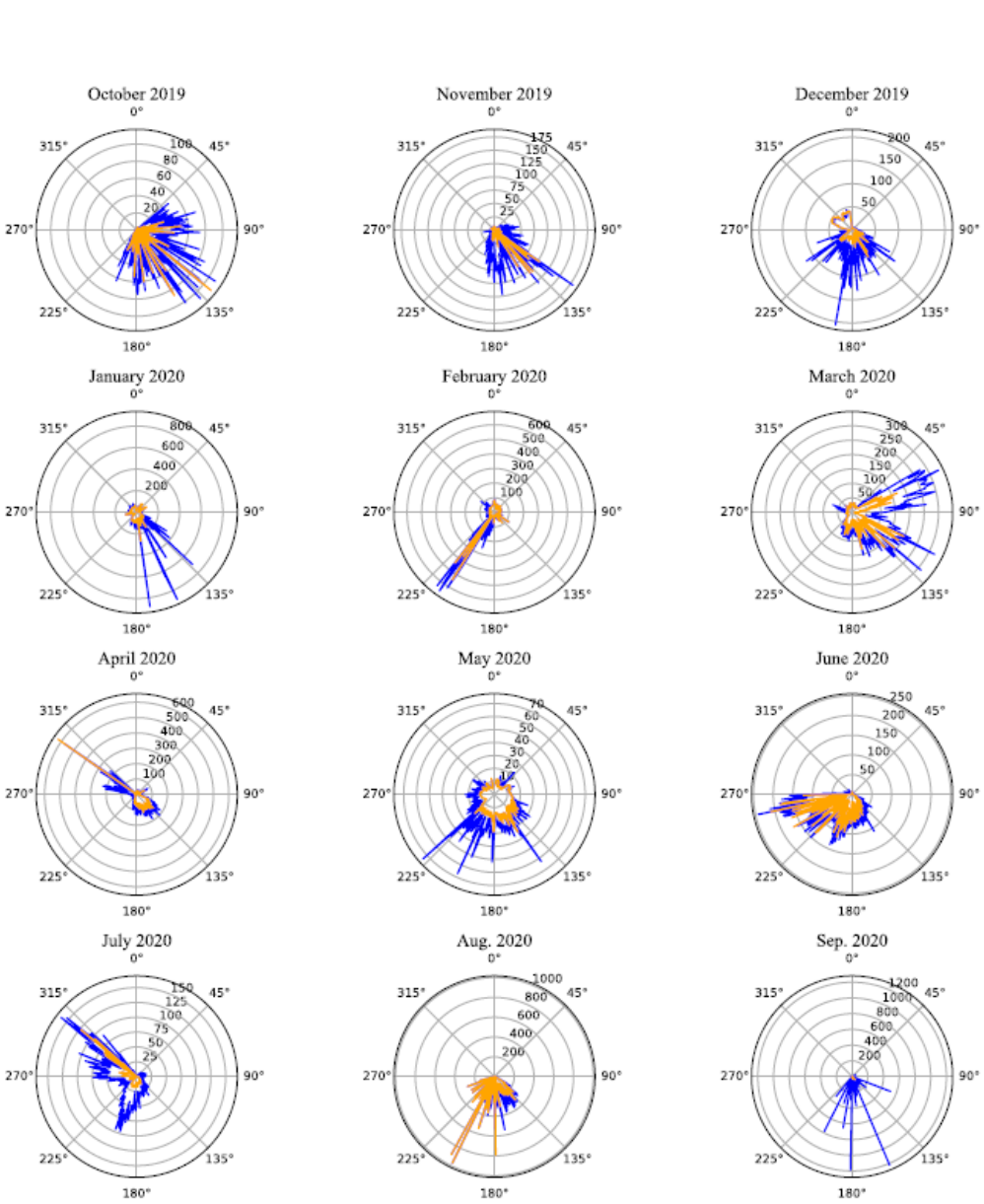


Figure 5. Monthly polar plots of black carbon concentration in units of ng/m^3.

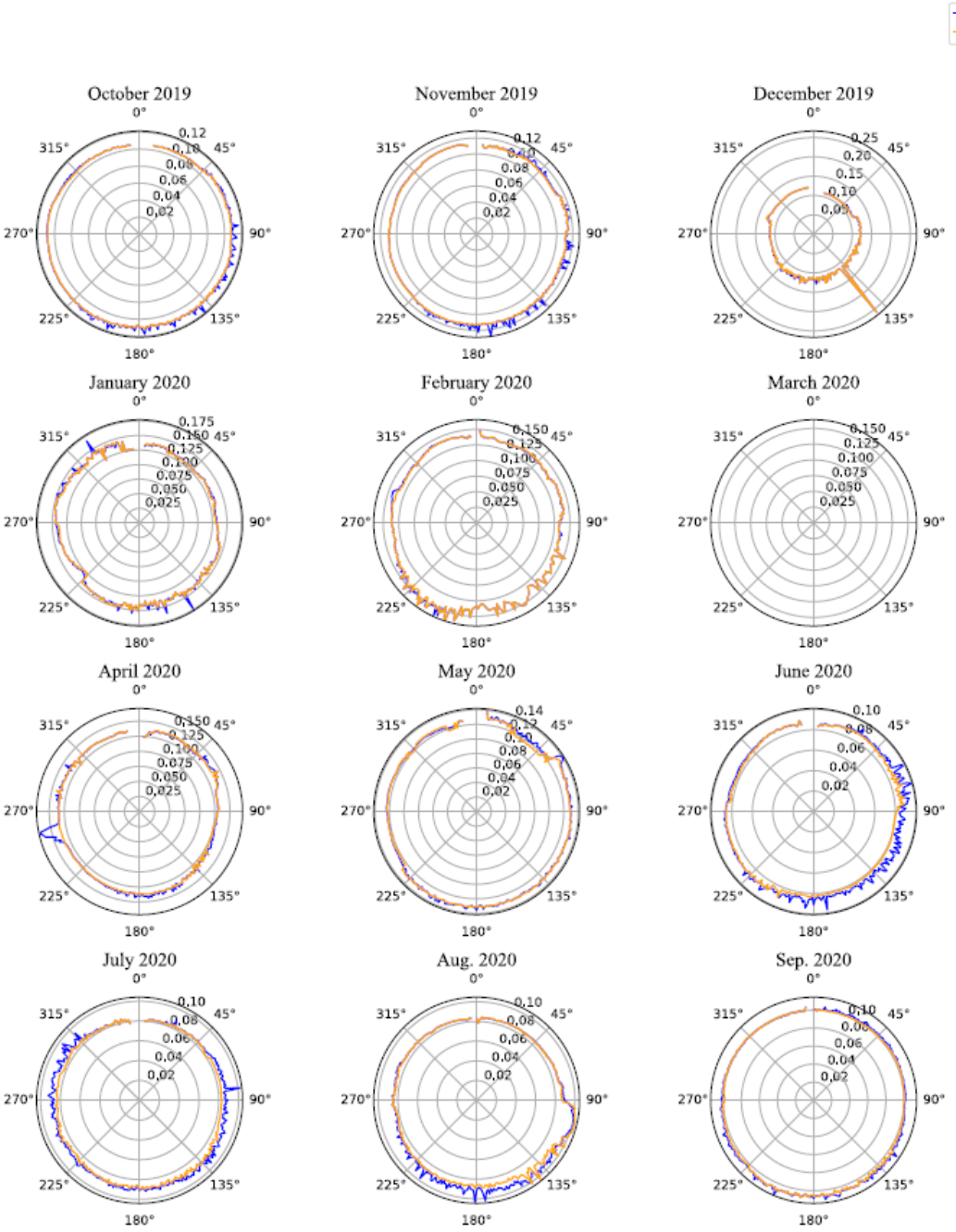


Figure 6. Monthly polar plots of carbon monoxide concentration in units of ppmv.

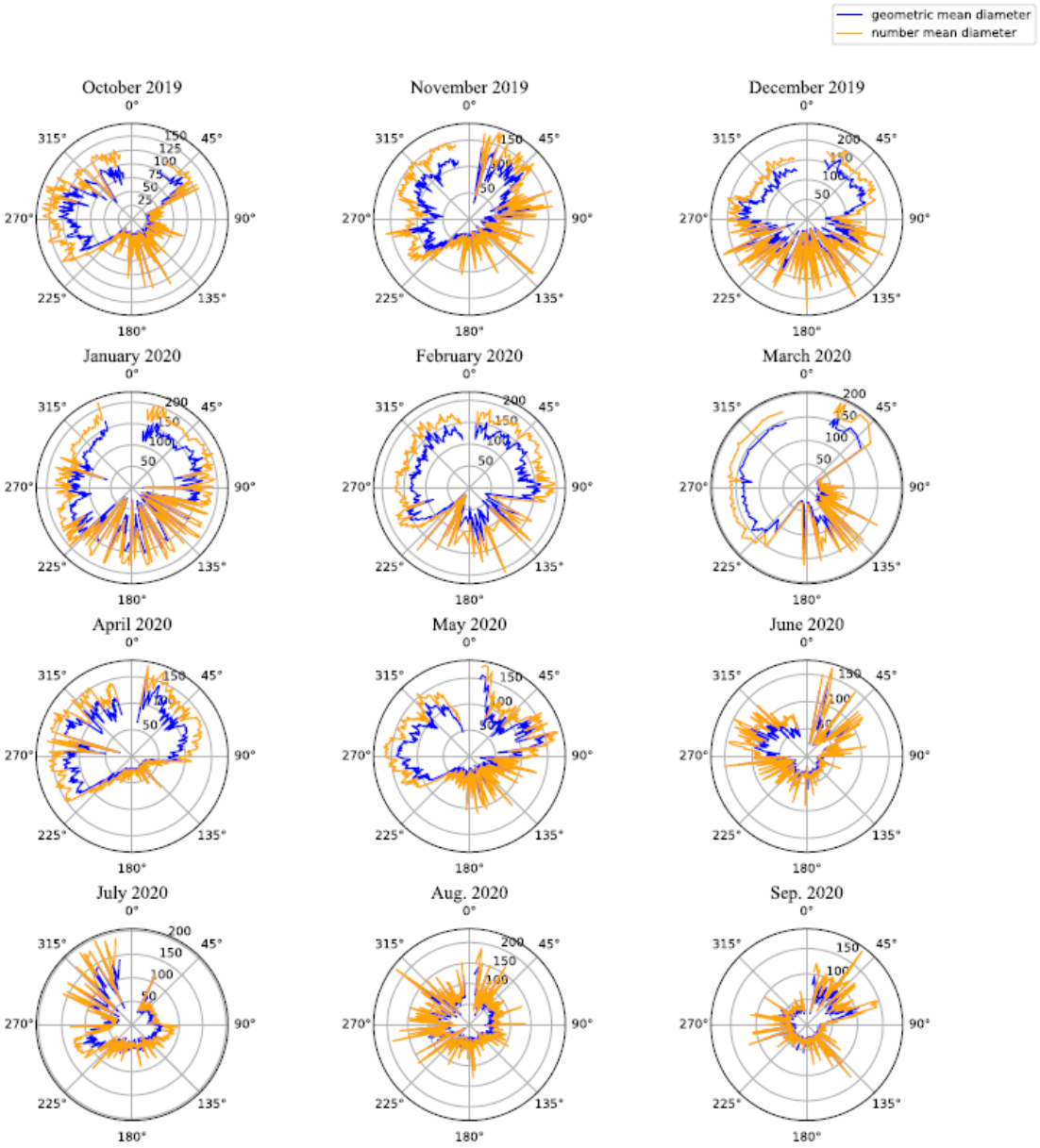


Figure 7. Monthly polar lots of geometric and number mean diameters of particles in nanometers.

The spatial variation and elevated concentrations of black carbon are strong indicators that certain measurements were indeed contaminated by combustion byproducts that could include but are not limited to the ship stack and other possible combustion appliances or vehicles operating on or around the nearby ice, hereinafter referred to as non-ship combustion emissions. These non-ship-related, but combustion-related, points of contamination can be seen in Figure 5 in the months of April and July 2020, where there are very high black carbon (BC) concentrations around 300◦. This relative wind direction is not associated with the ship exhaust stack or bridge; however, the order of magnitude of these values

indicates it cannot be arctic air. These values are likely therefore coming from other campaign-related activities such as ice-breakers, helicopters, and generators.

A final variable analyzed for temporal dynamics was the geometric mean diameter of particles detected. The geometric mean diameter is a value outputted by the SMPS based on the size distribution of particles sampled. It is used as a way to better characterize the size distribution of a lognormal sample as opposed to an arithmetic/ numerical mean (Hinds et al. 2022). The equation used to calculate the geometric mean diameter is as follows (TSI 2010):

$$\ln(d_g) = \frac{\sum_{i=m}^{n} \Delta N_i \ln(d_i)}{N} \quad (1)$$

where dg is the geometric mean diameter; di is midpoint diameter for size channel i; N is total concentration; Ni concentration within channel i; m is the first channel; and n is the final channel. The monthly polar projections for the geometric mean and numerical mean are shown in Figure 7. Looking at these polar plots, it can be seen that overall across the year, the geometric and numerical mean diameter of particles being measured from the wind direction of concern (110-225) are an order of magnitude smaller than the rest of the polar projection. Aerosol particles are grouped based on their size into three different modes (nuclei, accumulation, and coarse modes) that are defined by their method of production. For the geometric mean values found in the wind direction of concern, the diameter ranges from around 20 nm to 50 nm, which would be categorized as the nuclei or nucleation mode (Hinds et al, 2022). This mode consists mainly of primary particles produced by combustion processes; they are formed through the gas-to-particle nucleation process. These smaller-diameter events coinciding with higher-particle-concentration events seen in Figure 4 allow for the inference that these particles could have been generated from combustion and therefore come from the ship's exhaust stack.

Overall, Figures 2-7 indicate that contamination of the arctic air measurements through human activities has occurred. The figures demonstrate that overall measurements taken from the wind direction of concern coincide with higher particle concentration values, smaller particle sizes, and elevated levels of black carbon concentration.

2.2.1 Comparison of MOSAiC AOS Measurements to Previous Results from Literature

To better discern measurement results indicating local contamination, measurements from previous campaigns must be evaluated to understand expected results typical of the arctic region.

First, measured black carbon mass concentrations are compared to expected arctic values. Jurányi et al. sought to determine mean and median atmospheric black carbon mass concentration values for spring and summer (Jurányi et al. 2023). The study aimed to determine if there was a seasonal dependence to the black carbon mass concentrations and therefore took measurements across many years. They did find a seasonal dependence to the black carbon mass concentration, with spring resulting in the highest black carbon concentration and summer producing concentrations up to four times lower. The study presented the statistical spread of the data recorded and found that in spring, the concentrations reached a maximum of around 50 ng/m^3. Comparing this value to the degree-averaged black carbon mass concentrations plotted in Figure 5, it can be observed that much higher values of black carbon were recorded for a

significant portion of the campaign. Black carbon is a known byproduct of incomplete combustion, with measurement techniques for marine applications being an active area of research. A study done by Momenimovahed et al. compared a variety of black carbon measuring techniques to determine black carbon mass concentration in the exhaust runner of a 4-stroke marine diesel engine (Momenimovahed et al. 2021). They used different marine-applicable fuels as well as different load conditions to generate their black carbon data set and found in the exhaust runner that black carbon mass concentrations ranged from 10 mg/m^3 to 60 mg/m^3 (Momenimovahed et al. 2021). Since the campaign is not measuring directly from the ship exhaust runner, the values in Figure 5 never reach that high magnitude. However, since the values in Figure 5 are between those measured by Jurányi et al. and Momenimovahed et al., it could be implied that a mixture of exhaust and arctic air was being measured.

Additionally, expected particle number concentrations in the Arctic were looked at and compared to Figure 4. Through multiple years of data collection, Pernov et al. looked for seasonal trends in particle number concentration and plotted an annual cycle of daily median values (Pernov et al. 2022). They did find seasonal variation in particle number concentrations with peaks in total number concentration occurring during the summer. This was speculated to be due to the dominance of new particle formation caused by the ocean's emission of organics into the atmosphere (Pernov et al. 2022). Even with the seasonal variation, particle number concentration peaked at around 700 #/cm^3, which is significantly lower than the values plotted in Figures 3 and 4, where values reach up to the thousands, an order of magnitude higher. Particles can come from several different sources, both natural and anthropogenic, with the size of these particles being indicative of their source of origin and their age. Freshly generated particles are often small in size and large in number and can come from either gas-to-particle nucleation or from incomplete combustion (Seinfeld and Pandis 2006). The new particle formation rate does change in the arctic with season, which is what the paper presents (Pernov et al. 2022), but the large difference in number concentration between values presented by Pernov et al. and those recorded in this campaign do indicate that the sample at times is contaminated with exhaust emissions. This is especially true when looking at Figures 4 and 7 where the high number concentration and small median size are seen to correlate. These comparisons drawn are to provide a possible source of the contamination observed based on other ambient values studied in the Arctic.

2.3 Exploring Filtering Techniques

A variety of flagging algorithm techniques exist from prior literature (Gallo et al. 2020), (Beck et al. 2022), for adaptation and use. Specifically, the following techniques are explored in this work: a) frequency-based filtering, b) statistical-based filtering, c) wind direction and speed filtering and d) aerosol microphysics-based filtering. All techniques were explored at the cursory level and specifically, all of the stated techniques, except for frequency-based filtering, were explored in detail by application to relevant AOS data collected during the MOSAiC campaign. This time-series data originates from the datastreams documented in Table 1. Additional filtering techniques could be explored using rate of change of certain variables and particle chemical composition data, but were excluded from the current analysis to constrain its scope.

2.3.1 Frequency Filtering

All time series data can be processed using signal processing and time series analysis techniques. These techniques are based on Fourier transformations that convert the examined data set from the temporal

domain to the frequency domain where features of the frequency spectrum will highlight any periodicity in the data. Identifying these frequencies of interest (expected and unexpected) can provide a greater understanding of the periodicity of data. Contamination with a specific periodicity could then easily be identified and a filtering scheme could be easily developed to focus on identifying and cleaning specific frequency ranges of interest. The first step in exploring frequency-based filtering is to determine a periodogram, which is an analytical method used to discover the frequencies present in a time series data set (Shumway and Stoffer 2011). Figure 8 shows the periodogram of the particle concentration time series for the whole MOSAiC expedition.

A clear peak in power spectral density can be seen in Figure 8 at a frequency of around zero. Besides this initial peak, which represents the average of the data, no other significant resonance frequencies could be identified. For this reason, no further work to develop a frequency-based filtering scheme was performed.

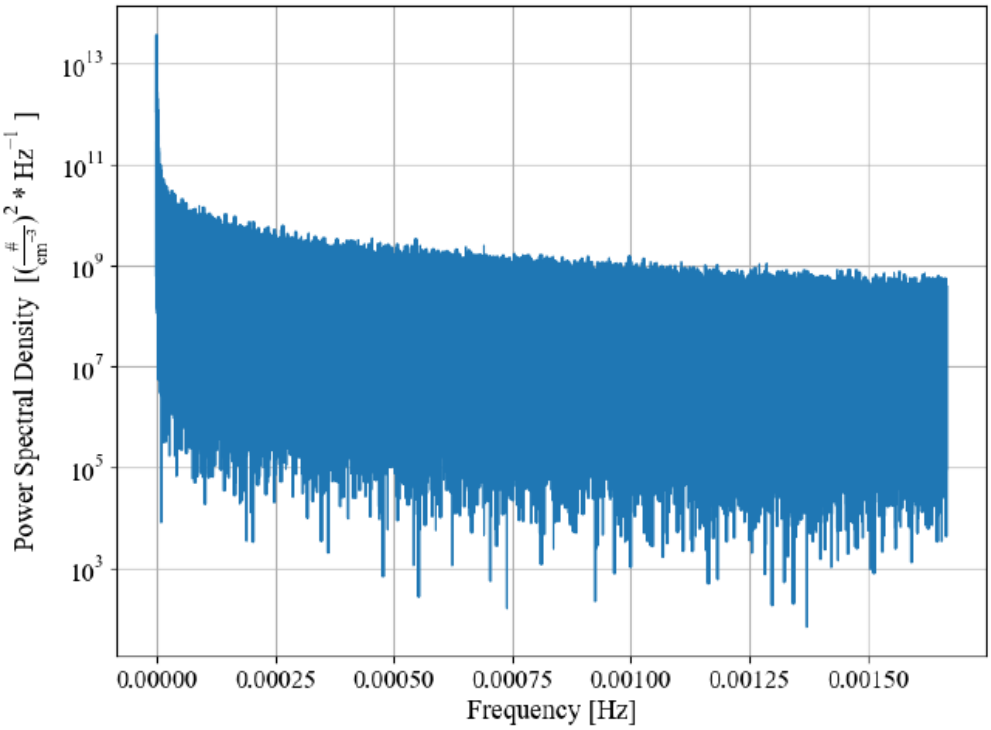


Figure 8. A flat-top periodogram of particle concentration.

2.3.2 Statistical Filtering

With the collection of any large data set, there is naturally a variability in measured values due to inherent randomness and range of responses any system or measurement quantities can exhibit. When the range of recorded values becomes large enough with respect to the expected variation in the data, statistical filtering can be used to remove the anomalous data points. In terms of specific measurement quantities of the MOSAiC campaign, the measured values could deviate due to atmospheric composition and dynamics such as weather patterns or localized events. Statistical-based filtering was explored specifically for particle concentration, which is known to have been heavily affected by local ship-related activities, as shown in Figures 3 and 4. The benefit of statistical filtering is that it can be employed on a wide range of

the variables recorded during the expedition, and different statistical criteria can be used. In this type of filtering, the focus is to exploit unexpected deviations from mean and median values to identify affected data points (such as too high concentration) and remove these values to ensure a clean data set is preserved. The Empirical Rule indicates that for a data set that expresses a normal distribution, 97% of the data can be found within three (3) standard deviations from the mean (Peck 2008). Thus, this rule is often used to identify and remove statistical outliers, data points that fall beyond three standard deviations from the mean. This statistical outlier detection technique was used as a starting point for creating a generic statistical filter. Since not all of the data under filtering consideration could be expressed as a normal distribution, flexibility in determining the critical threshold number of standard deviations away from the mean was introduced through the use of a variable parameter. Additionally, while the Empirical Rule would determine all values beyond three standard deviations from the mean to be considered as outliers (in both the positive and negative direction), this rule loses meaning when the rule introduces negative values for physical measurements. Therefore, the empirical rule method is used more as a guiding tool to aid in the design of a statistical-based approach with additional user-defined parameters implemented, in order to decipher the correct number of standard deviations to use in non-normal distributed data sets.

The Empirical Rule and statistical filtering were implemented and relevant quality control and purge flags published by ARM (Singh and Kuang 2024, Uin 2024) were included in the developed software before applying any statistical filtering schemes. Most commonly, data that existed three or more standard deviations beyond (positive direction) the mean were removed as the data were classified as statistical outliers. All data that fell below the mean were intentionally not filtered, regardless if they existed at or beyond three standard deviations less than the mean, as small particle concentrations were expected to be seen in this campaign due to the arctic climate and geographically remote nature of the study.

Classifying such small concentration values as statistical outliers would inherently bias the final data set by removing real and potentially expected values. In addition, the lower value filtration is less of a concern to be contaminated as the working definition of contamination in this work is expected to be from anthropogenic sources, such as the ship stack, which are expected to produce high particle concentration values. The resulting time series and polar projection for the year can be seen in Figures 9 and 10. In Figure 9, the clear cut-off line for three standard deviations above the mean is around 5000 cm^{-3}. When looking at this filtered data on a polar projection, as seen in Figure 10, there is minimal change to data outside of the wind direction of concern and large decrease in concentration magnitude in the direction of the ship stack. This demonstrates that statistically filtering high-concentration events impact the degrees of concern the most. Overall, this filter removed about 3% of the data for a decrease in average particle concentration from 609.37 cm^{-3} to 394.7 cm^{-3}. This exploration into one variation of statistical filtering proved that statistics can be an effective means of removing small amounts of data for big impacts on polar projection plots and average concentrations.

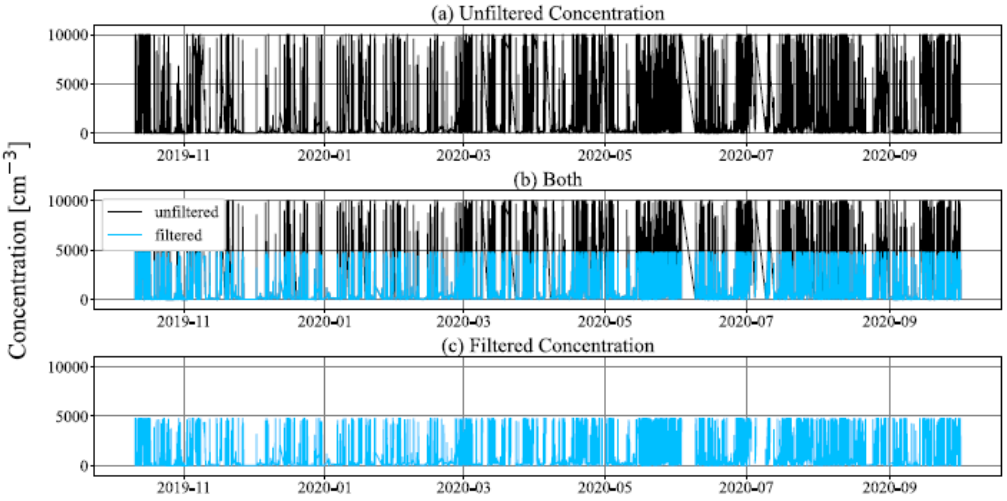


Figure 9. Time series of particle concentration with the three-standard-deviation statistical filter applied.

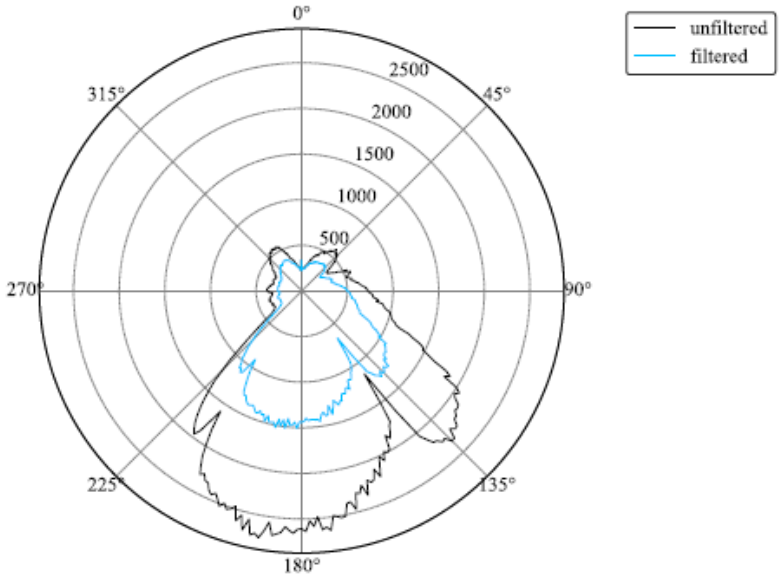


Figure 10. Polar projection of filtered and unfiltered particle concentration where data three standard deviations or higher beyond the mean has been removed.

2.3.3 Meteorological Filtering

Meteorological data, specifically wind direction and speed, varied greatly throughout the expedition as seen in Figure 4. Due to this variation, an exploration into filtering was undertaken to see if using purely meteorological data could be sufficient to make a reliable filter for contamination. In order to understand the effect of variation in meteorological data on aerosol particle number concentration, the concentration data was grouped based on wind data. The two conditions used to group the data are as follows: (1) data associated with wind from the direction of ship's superstructure (110-225◦, Figure 2) versus out of that range and (2) data that was recorded at low wind speeds (< 5 m/s) versus high speeds (≥ 5 m/s). The ranges for these two conditions were selected based on Figures 2 and 3.

With these two conditions, the data was grouped and statistics for each grouping are presented in Table 2. It is important to note that the merging of any two datastreams recorded at different time intervals can affect the statistics of the data due to interpolation. For transparency, statistics both before and after filtering are presented. When looking at the data split by wind direction, an overwhelming majority of the data is recorded outside the range of concern (˜90%). When taking an average of particle concentration outside the direction of concern, there is an almost 150 #/ cc drop in mean particle concentration; indicating that a significant portion of high-concentration events were removed. The correlation between high-concentration events and the wind direction of concern can also be observed by looking at the average particle concentration, which is over 1000 #/ cc higher than the overall average. When splitting the data across wind speed, trends are a little less obvious. There is an almost even split in the data between measurements taken at less than and greater than 5 m/s. When looking at the respective average concentrations, slow wind speeds show a slightly higher average concentration and faster wind speeds show a slightly lower. This confirms a hypothesis presented in the previous section that faster wind speeds help with the diluting of contamination with more fresh arctic air, while at low wind speeds the local contamination can persist even when the wind is from the direction of the ship's superstructure. Based on the percentage splits presented in Table 2, a purely meteorological filter would remove a substantial amount of data. For example, if a pure wind direction filter were deployed, 10% of the data collected would be removed and as can be seen in Figure 3; not all of the data collected from this direction is contamination. Therefore, pure wind direction or wind speed filters were not explored further; a filter that combines wind direction and low wind speeds will be presented in later section for comparison.

Table 1. Meteorological data splits for particle concentration.

Data Type	Data Ranges	Percentage (%)	Mean (#/cc)	Median (#/cc)
All Data	–	100	609.37	167.4
Wind Direction	0-110° and 225-360°	89.25	449.38	154.0
	110-225°	10.75	1965.63	839.2
Wind Speed	< 5 m/s	50.69	894.95	198.60
	≥ 5 m/s	49.31	318.76	148.80

2.3.4 Microphysical Filtering

A final type of filtering explored in this study was microphysical filtering, which is filtering based on the size of the sampled aerosol particles. Figure 7 demonstrates a large variation in the geometric mean diameter of the particles being sampled by the AOS sample inlet stack. This variation tended to coincide with events of high concentration when looking at the polar projections of particle concentration in Figure 4. The nucleation mode, composed of primary particles, tends to be defined as particles with a diameter up to 0.1 μm or 100 nm, and this became the cut-off value for an initial filtering scheme (Hinds et al. 2022). A correlation study was done between geometric mean diameter and particle concentration to further confirm that high-concentration events are associated with these smaller particle sizes. The geometric mean diameter was plotted against total particle concentration as shown in Figure 11. Through this figure it is clearly shown that all high-particle-concentration events are associated with smaller particle sizes. Although the nuclei mode does span up to 100 nm, from observation of this graph, most statistical outliers of particle concentration are associated with much smaller sizes, of around 40 nm. Therefore, though guidance of Figures 7 and 11, a filter was designed to remove data recorded at high particle concentrations (defined as values 3-std or more away from the mean) and when the geometric mean diameter was at or below 40 nm. The resulting filter is presented and compared to the other filtering techniques in the results section.

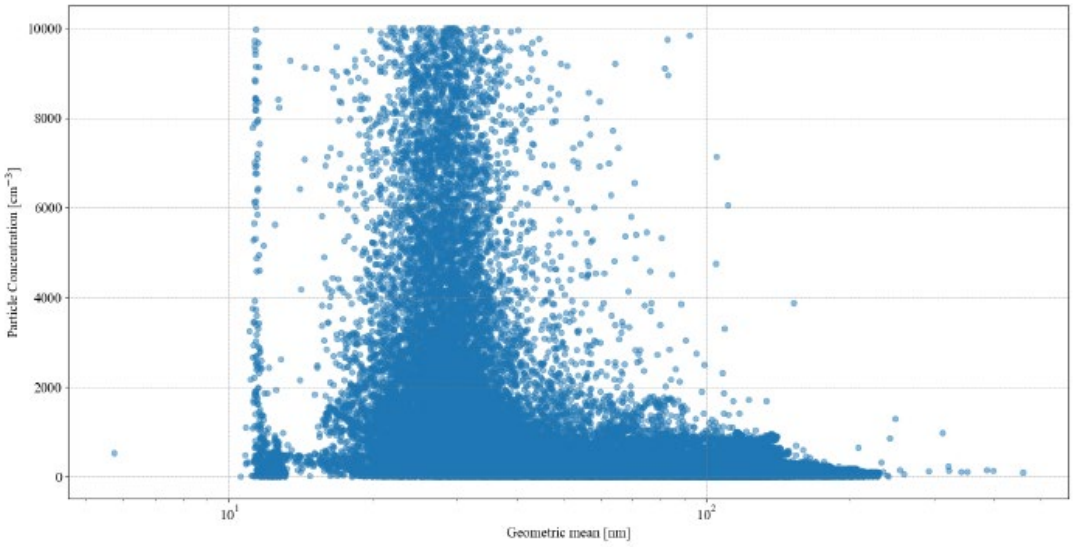


Figure 11. Correlation scatter plot for geometric mean diameter and particle concentration.

To summarize, exploration was done into frequency-, meteorological-, statistical-, and microphysical-based filtering with the result that meteorological, statistical, and microphysical filtering would be pursued further and compared for their respective effects on time series and polar-projected plots.

2.4 Statistical Comparison of Filters

2.4.1 Automation of Filtering Routines and Application to ARM AOS Data

Application of these filtering techniques to the year-long time series required some automation in the code used to process the data. The two general approaches used to automate as much of the code as possible were as follows: (1) building functions within the code to automate common steps and (2) minimizing hard-coded variables by user input.

One crucial step was repeated when testing individual filtering techniques and due to its repetition was converted into a function that can be called within a script when needed. The step repeated is the averaging of data on a wind-direction-degree basis to create the polar project plots that are crucial for comparing filters performance. For example, Figure 4 (b) was generated by taking an average of the particle concentration for each degree of wind direction and then plotting it on a polar projection. The function inputs are the data frame containing the variable to be analyzed and the wind direction as well as a choice of a mean or median trace. The function then manipulates this data frame through the steps outlined and returns a data frame that has an averaged or median value for each wind direction degree.

The other automation effort was to develop some user input to minimize hard-coded decisions. One filter, the statistical filter, was based on this mindset. This user input information includes file path, filename, variables downloaded, and number of standard deviations from mean to be filtered out. The filter uses these inputs to automatically download data, apply filtering based on users specifications, and output a filtered time series and polar projection. These automated functions helped develop the code to run smoother and become more repeatable when applied to the AOS data.

The filtering schemes explored in the previous section were applied to all of the particle concentration data taken over the year-long expedition with results presented below. Prior to applying the explored filtering schemes to recorded data, the data had to go through a two-stage quality control process. First, all data taken during the periods when the AOS clean air purge system was on (see above) were removed. The second stage was to apply each datastream's respective quality control, which is also published by ARM. Once all data had passed through quality control, the filtering scheme was applied and the resulting data analyzed. This section presents results from two statistical filters, a meteorological filter and a microphysical filter.

2.4.2 Filtering Performance

In addition to the three-standard-deviation filter described earlier, two more statistical filters were developed: these are presented in Figures 13 through 16. The first of these two filters is the same structure as the three-standard-deviation filter; however, the cut-off is four standard deviations. This is to demonstrate how a change in the cut-off value will alter the aggressiveness of the filter and vary the degree of change the filter can implement. The second type of statistical filter developed is a rolling-based statistical filter that is presented in Figures 15 and 16. This filter uses the rolling function in Python to take a rolling average of 100 data points at a time. The standard deviation is then calculated on this rolling basis so that the cut-off point changes as the window moves along the x-axis. Upon visual inspection, this filter is clearly less aggressive and shows almost no change in data compared to the previous statistics-based filter. The rolling-based technique was implemented to see if it could be more sensitive to the

natural peaks and valleys of the data, but in doing so the filter removes almost no data. Changing the window size was explored and can help increase the aggressiveness of this technique, but overall, if there is significant contamination a single cut-off value is more effective than one that changes over time. The performance of the four statistical filters is summarized in Table 3. Overall, because of how robust the statistical filtering technique is, the user must decide how much or how little data to be removed.

Table 2. Statistical filter performance.

Filter Type	Conditional	Data Lost [%]	Data Retained [%]	Change in Mean [%]	Change in Median [%]
Statistical #1	>3std	3.22	96.78	35.08	3.46
Statistical #2	>4std	2.13	97.87	25.80	2.27
Statistical #3	>rolling 3std (100 window)	0.88	99.12	1.58	0.48
Statistical #4	>rolling 3std (200 window)	0.89	99.11	2.09	0.59

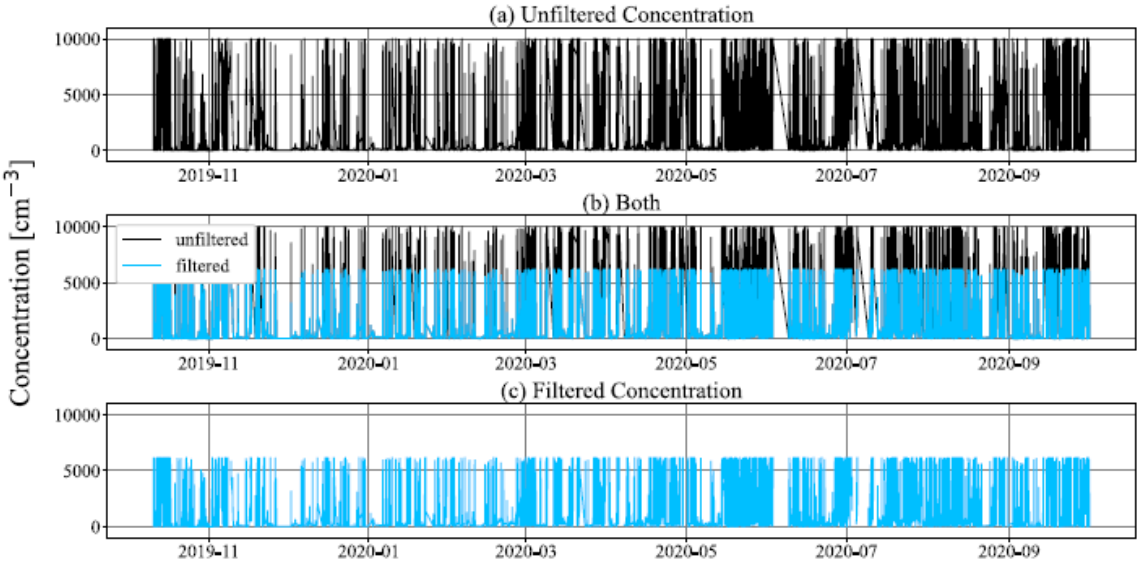


Figure 12. Time series of particle concentration with the four-standard-deviation statistical filter applied.

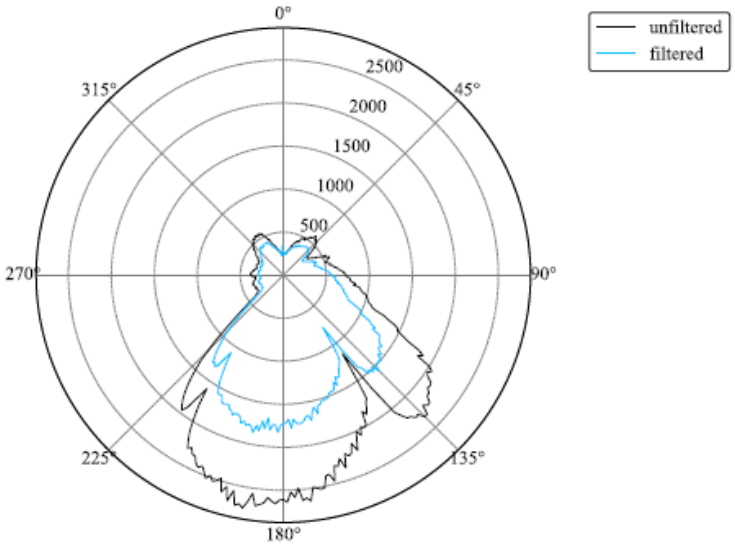


Figure 13. Polar projection of particle concentration with the four-standard-deviation statistical filter applied.

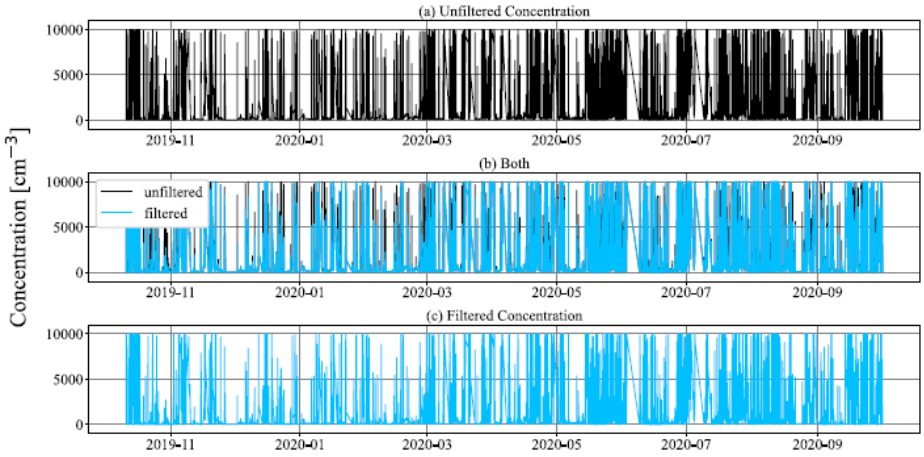


Figure 14. Time series of particle concentration with the rolling statistical filter applied.

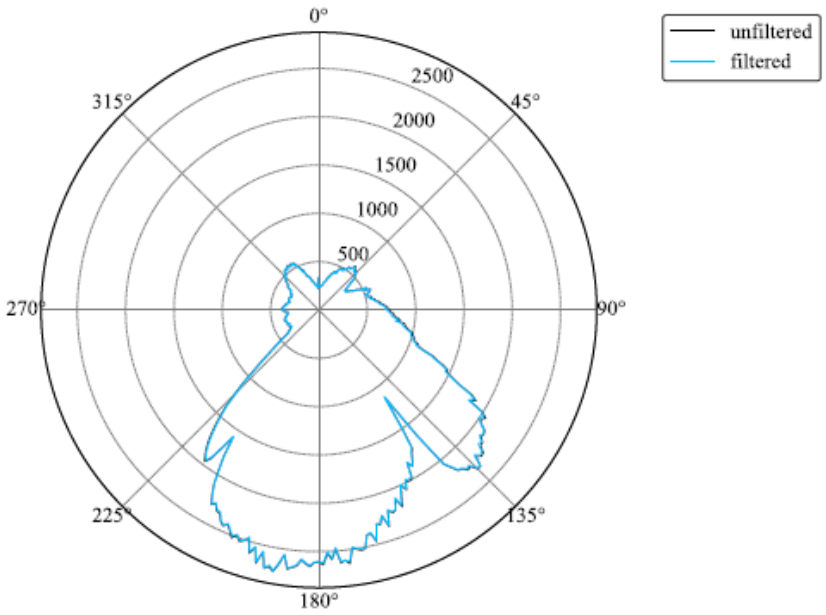


Figure 15. Polar projection of particle concentration with the rolling statistical filter applied.

As stated in the previous section, a pure wind direction or wind speed filter would be too aggressive and could potentially remove good data, so a combination of speed and direction as a filter was applied to the AOS data as presented in Figures 17 and 18. The filter removes data contained within the direction range of 110-225◦ and recorded at a wind speed of less than 5 m/s. Figure 18 shows the polar projection of this filtering scheme and, interestingly, the filter has mixed results in terms of effectiveness. For certain small wind direction windows, the filter is effective at bringing the average particle concentration down, such as between 135-150 ◦. However, there are other wind directions, such as from 180-225◦, where the filter appears to remove good data points and causes the average to actually increase. The net effect of these two opposite trends on the year average value of particle concentration is still a decrease, which can be observed in Table 4; however, the polar projection demonstrates that this filter creates false flags and removes good data for some wind directions. While the filter overall does remove contaminated data points, it does not do so in a consistent or reliable manner, meaning a meteorological-based filter, regardless of conditions of filtering, is not effective.

Table 3. Meteorological filter performance.

Filter Type	Conditional	Data Lost [%]	Data Retained [%]	Change in Mean [%]	Change in Median [%]
Meteorological	within 110-225° and <5 m/s	9.06	90.94	21.16	6.81

The final filtering scheme applied was a microphysical-based one. One important note to make about the microphysical-based filter is that the particle concentration data had to be averaged on a five-minute basis to match the frequency at which the microphysical data was collected. This means that the time scale for this filtering scheme is slightly different than the previous filters which work on a five-second frequency. The filtering scheme removes data points collected three standard deviations away from the mean particle concentration and at a geometric mean diameter of 40 nm or less. Figures 17 and 18 show the time series and polar projection of this filter and Table 5 shows the filter's performance. As stated, the unfiltered trace is slightly different in shape due to the change in frequency sampling. In the polar projection, it can be observed that the microphysical filter is able to reduce the particle concentration across all degrees of concern. This consistency is what the meteorological filter was lacking when comparing filtered traces between Figures 19 and 17.

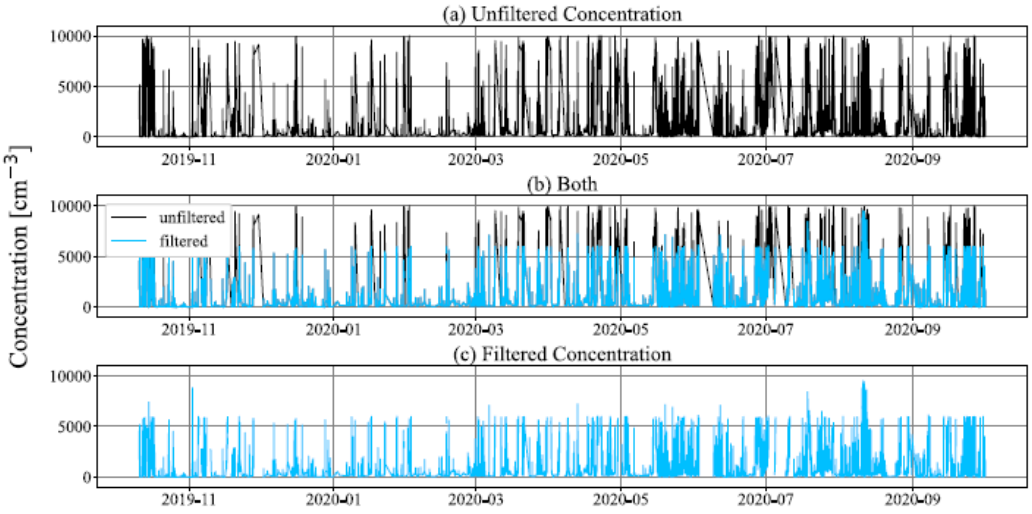


Figure 16. Time series of particle concentration data with the microphysical filter applied.

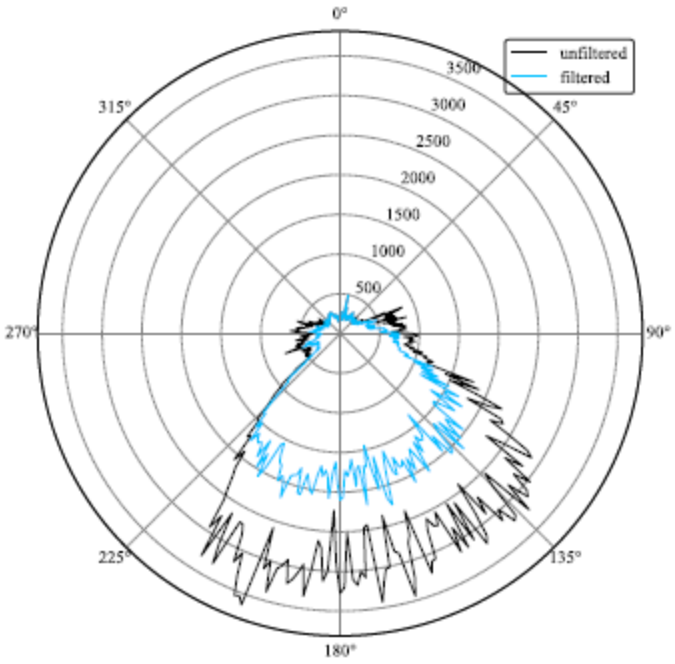


Figure 17. Polar projection of particle concentration data with the microphysical filter applied.

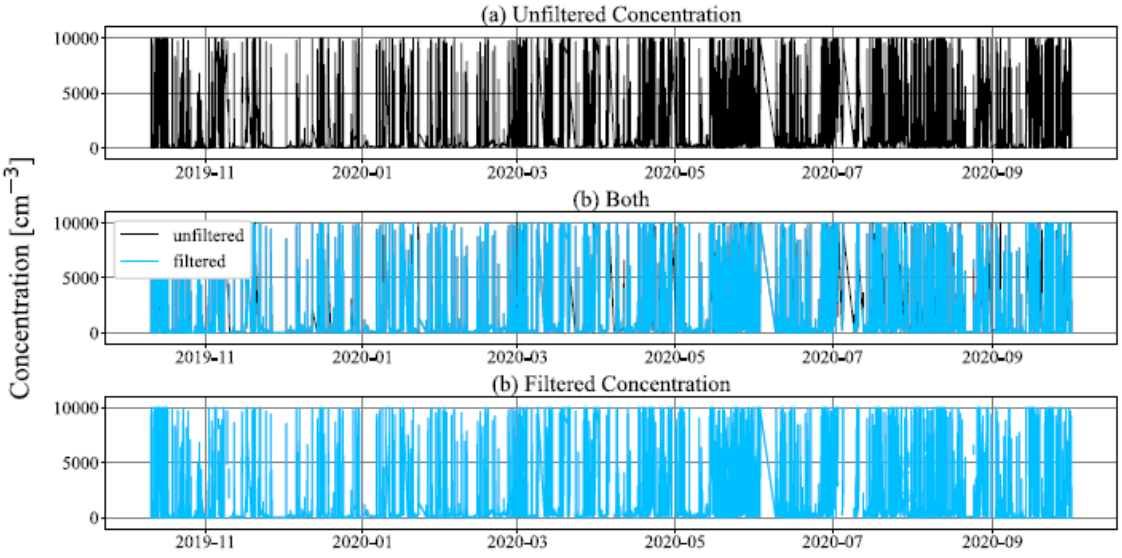


Figure 18. Time series of particle concentration data with meteorological filter applied.

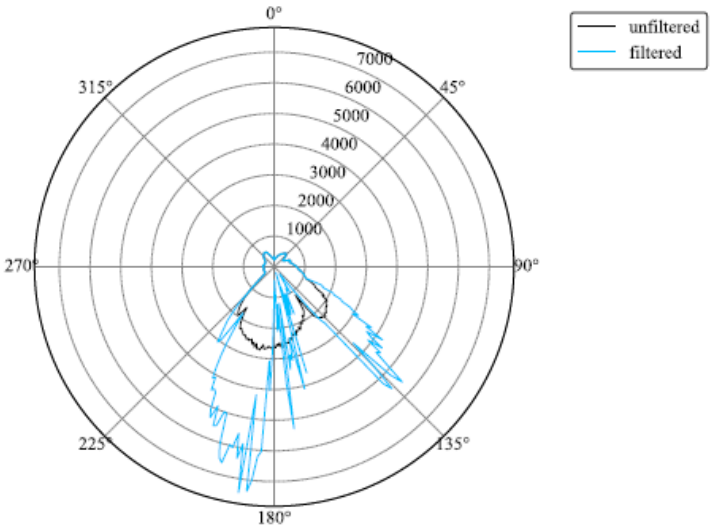


Figure 19. Polar projection of particle concentration data with meteorological filter applied.

2.4.3 Summary of Results

All of the above filtering techniques are now compared for the statistical change they were able to make to average and median particle concentration values as shown in Table 5. The properties analyzed are the amount of data lost versus retained, and percent change in mean and median particle concentrations. In the ideal case, a filter will provide the largest decrease in particle concentration for the smallest amount of data lost, thus minimizing any false positives. All filters tested resulted in a net decrease in particle concentration; however, the effectiveness of these filters varies. The two filters to make the largest impact on the mean concentration for the smallest loss of data were the three-standard-deviation statistical filter and the microphysical filter. They produced a 35 and 35% decrease in mean for only a 3.20 and 3.98% loss of data respectively. This means these two filtering schemes were the most consistent in removing only contaminated data points. The least effective filter was the meteorological-based filter. There was still a decrease in mean concentration, indicating a removal of contaminated points, but the loss of data was triple that of the most effective filter, clearly indicating that good data was removed with contaminated data. Of all filtering schemes explored, the meteorological-based one is the only one making false positives and is therefore not an effective filter. The other statistical filters presented all had varying degrees of impact on the mean value for ranging amounts of data loss. The choice between which statistical filter to use is truly up to users depending on how aggressive they would like to be.

Table 4. Filtering scheme performance and evaluation.

Filter Type	Conditional	Data Lost [%]	Data Retained [%]	Change in Mean [%]	Change in Median [%]
Statistical #1	>3std	3.22	96.78	35.08	3.46
Statistical #2	>4std	2.13	97.87	25.80	2.27
Statistical #3	>rolling 3std (100 window)	0.88	99.12	1.58	0.48
Statistical #4	>rolling 3std (200 window)	0.89	99.11	2.09	0.59
Meteorological	within 110-225° and <5 m/s	9.06	90.94	21.16	6.81
SMPS	>3std and <40nm geometric diameter	3.98	96.02	34.68	4.91

Overall, microphysical filtering proved to be both consistent and reliable for identifying and removing contamination. Overall, all the filters that were explored and deemed viable for application to a full data set were able to be built and successfully run. The consistency of some filters were better than others. The overall statistical performance will be analyzed in the next section.

3.0 Summary and Conclusions

This work aimed to parameterize conditions of contamination and explore a variety possible filtering techniques for the MOSAiC expedition aerosol measurements. Three methods of filtering were deemed feasible for applying to the MOSAiC data (microphysical, statistical, and meteorological); that application was done with the following results:

- The study validated that contamination of arctic air measurements did occur during MOSAiC.
- The tell-tale sign of this contamination was the very high particle concentration that coincided with peaks in black carbon concentration and a decrease in geometric mean diameter of the particles; a lot of small, freshly generated particles were detected from the direction of the ship superstructure.
- Three filtering techniques were developed: meteorological, microphysical, and statistical.
- A total of six filters were built using these techniques and applied to the MOSAiC data.
- Of the three filtering techniques tested, only the meteorological-based filter yielded false positives (i.e., removing good data points), and increased the average particle concentration in some wind directions.
- The three-standard-deviation filter removed the least amount of data for the biggest decrease in mean, removing about 3% of the data for a 35% decrease in mean.
- Filters based on size distribution and statistics proved to be the most effective, removing high-concentration events and decreasing the particle concentration average.

While certain filtering techniques seemed to be more efficient in flagging contamination events in the AOS aerosol data, a choice of a filter and its parameters may depend on the nature of research being done. Also, a combination of multiple filtering techniques may be useful in certain cases. Thus, the procedures

developed in this work are designed to be flexible and user-configurable to simplify data QA/QC and facilitate the use ARM aerosol measurements from MOSAiC. The procedures developed in this work will be made available by ARM.

4.0 References

ARM. "AOS Instruments." [Online]. Available: https://www.arm.gov/capabilities/instruments/aos. (accessed: 07.26.2023).

ARM. "Data search." [Online]. Available: https://adc.arm.gov/discovery/#/results/. (accessed: 09.17.2023).

ARM. "MOSAIC." [Online]. Available: https://www.arm.gov/research/campaigns/amf2019mosaic. (accessed: 08.09.2023).

ARM. "Multidisciplinary drifting observatory for the study of arctic climate." [Online]. Available: https://www.arm.gov/publications/backgrounders/docs/doe-sc-arm-18-009.pdf. (accessed: 07.26.2023).

Beck, I, H Angot, A Baccarini, L Dada, L Quéléver, T Jokinen, T Laruila, M Lampimäki, N Bukowiecki, M Boyer, X Gong, M Gysel-Beer, T Petäjä, J Wang, and J Schmale. 2022. "Automated identification of local contamination in remote atmospheric composition time series." *Atmospheric Measurement Techniques* 15(14): 4195–4224, https://doi.org/10.5194/amt-15-4195-2022

Bukowiecki, N, BT Brem, G Wehrle, G Močnik, S Affolter, M Leuenberger, MC Coen, M Hervo, U Baltensperger, and M Gysel-Beer. 2021. "Elucidating local pollution and site representativeness at the Jungfraujoch, Switzerland through parallel aerosol measurements at an adjacent mountain ridge." *Environmental Research Communications* 3(2): 021001, https://dx.doi.org/10.1088/2515-7620/abe987

Gallo, F, J Uin, S Springston, J Wang, G Zheng, C Kuang, R Wood EB Azevedo, A McComiskey, F Mei A Theisen, J Kyrouac, and AC Aiken. 2020. "Identifying a regional aerosol baseline in the eastern North Atlantic using collocated measurements and a mathematical algorithm to mask high-submicron-number-concentration aerosol events." *Atmospheric Chemistry and Physics* 20(12): 7553–7573, https://doi.org/10.5194/acp-20-7553-2020

Hinds, W, and Y Zhu. 2022. "Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles." John Wiley & Sons, ISBN: 9781119494065. [Online]. Available: https://books.google.com/books?id=TS1sEAAAQBAJ

Jurányi, Z, M Zanatta, MT Lund, BH Samset, RB Skeie, S Sharma, M Wendisch, and A Herber. 2023. "Atmospheric concentrations of black carbon are substantially higher in spring than summer in the Arctic." *Communications Earth & Environment* 4: 91, https://doi.org/10.1038/s43247-023-00749-x

Kyrouac, J. 2019. "Aerosol Observing System Surface Meteorology (AOSMET) Instrument Handbook." U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-TR-184.

Momenimovahed, A, S Gagné, P Martens, G Jakobi, H Czech, V Wichmann, B Buchholz, R Zimmermann, B Behrends, and KA Thomson. 2022. "Comparison of black carbon measurement techniques for marine engine emissions using three marine fuel types," *Aerosol Science and Technology*, 56(1), 46–62, https://doi.org/10.1080/02786826.2021.1967281

Peck, R, C Olsen, and J Devore. 2008. *Introduction to Statistics & Data Analysis.* Thomson Higher Education, ISBN: 9780495118732.

Pernov, JB, D Beddows, DC Thomas, M Dall´Osto, RM Harrison, J Schmale, H Skov, and A Massling. 2022. "Increased aerosol concentrations in the high arctic attributable to changing atmospheric transport patterns." *npj Climate and Atmospheric Science* 5: 62, https://doi.org/10.1038/s41612-022-00286-y

Sedlacek, AJ. 2017. "Single-Particle Soot Photometer (SP2) Instrument Handbook." U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SCARM-TR-169.

Seinfeld, JH, and SN Pandis. 2006. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change.* John Wiley & Sons, ISBN: 9780471720171.

Shumway, R, and D Stoffer. 2011. *Time Series Analysis and Its Applications With R Examples.* Springer Texts in Statistics, vol. 9, ISBN: 978-1-4419-7864-6. https://doi.org/10.1007/978-1-4419-7865-3

Shupe, M, G de Boer, K Dethloff, E Hunke, W Maslowski, A McComiskey, O Perrson, D Randall, M Tjernstrom, D Turner, and J Verlinde. 2028. "The Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAIC) Atmosphere Science Plan." U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-18-005. https://www.arm.gov/publications/programdocs/doe-sc-arm-18-005.pdf

Singh, A, and C Kuang. 2024. "Condensation Particle Counter (CPC) Instrument Handbook." U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-TR-145.

Singh, A, and C Kuang. 2024. Scanning Mobility Particle Sizer (SMPS) Instrument Handbook. U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-TR-147.

Springston, SR. 2015. "Carbon Monoxide Analyzer (CO-Analyzer) Instrument Handbook." U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SCARM-TR-159.

TSI Incorporated. 2010. Model 3936 scanning mobility particle sizer™ (smps™) spectrometer.

Uin, J, AC Aiken, MK Dubey, C Kuang, M Pekour, C Salwen, AJ Sedlacek, G Senum, S Smith, J Wang, TB Watson, and SR Springston. 2019. "Atmospheric Radiation Measurement (ARM) Aerosol Observing Systems (AOS) for Surface-Based In Situ Atmospheric Aerosol and Trace Gas Measurements." *Journal of Atmospheric and Oceanic Technology* 36(12): 2429–2447, https://doi.org/10.1175/JTECH-D-19-0077.1

Uin, J. 2024. Nephelometer Instrument Handbook. U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-TR-165.

www.arm.gov

U.S. DEPARTMENT OF
ENERGY

Office of Science