An unsupervised machine-learning-based classification of aerosol microphysical properties over 10 years at Cabo Verde

Xianda Gong[1], Heike Wex[1], Thomas Müller[1], Silvia Henning[1], Jens Voigtländer[1], Alfred Wiedensohler[1], and Frank Stratmann[1]

[1]Department Experimental Aerosol and Cloud Microphysics, Leibniz-Institute for Tropospheric Research (TROPOS), Leipzig, 04318, Germany

Correspondence: Xianda Gong (gong@tropos.de)

Abstract. The Cape Verde Atmospheric Observatory (CVAO), which is influenced by both, marine and desert dust air masses, has been used for long-term measurements of different properties of the atmospheric aerosol from 2008 to 2017. These properties include particle number size distributions (PNSD), light absorbing carbon (LAC) and concentrations of cloud condensation nuclei (CCN) together with their hygroscopicity. Here we summarize the results obtained for these properties and use an unsu-

5 pervised machine learning algorithm for the classification of aerosol types. Five types of aerosols, i.e., marine, freshly-formed, mixture, moderate dust and heavy dust, were classified. Air masses during marine periods are from the Atlantic Ocean and during dust periods are from the Sahara. Heavy dust was more frequently present during wintertime, whereas the clean marine periods were more frequently present during springtime. It was observed that during the dust periods CCN number concentrations at a supersaturation of 0.30% are roughly 2.5 times higher than during marine periods, but the hygroscopicity (κ) of

10 particles in the size range from ∼30 to ∼175 nm during marine and dust periods are comparable. The long-term data presented here, together with the aerosol classification, can be used as a base to improve our understanding of annual cycles of the atmospheric aerosol in the eastern tropical Atlantic and on aerosol-cloud interactions and it can be used as a base for driving, evaluating and constraining atmospheric model simulations.

1 Introduction

15 Aerosol particles are known to impact the cloud formation, life cycle of clouds and global radiative forcing. These impacts are closely related to aerosol particle physical and chemical properties. Atmospheric aerosol particles originate from a vast variety of anthropogenic and natural sources (Pöschl, 2005). In this study, the focus is on naturally occurring aerosol types. Natural aerosol sources include primary mineral dust, sea spray, and wildfire aerosol, as well as precursors for new particle formation, such as sulfur- and carbon-containing gases from vegetation, ocean environment, and volcanoes, that can subsequently form

20 particles in the atmosphere (Carslaw et al., 2010).

Mineral dust is likely the most abundant aerosol type by mass in the atmosphere (Kok et al., 2017), with estimates for the dust emission ranging from 1000-2150 Tg yr^{-1} (Zender et al., 2004). The direct dust–climate feedback is likely in the range of -0.04 to +0.02 Wm^{-2} K^{-1}, and it may account for a substantial fraction of the total aerosol feedback in the climate system

(Kok et al., 2018). Furthermore, mineral dust can change cloud properties, i.e., serve as cloud condensation nuclei (CCN) (Karydis et al., 2011) or ice-nucleating particles (INPs) (Sassen et al., 2003; DeMott et al., 2003). Based on a global transport model, Karydis et al. (2011) predicted that the annual average contribution of insoluble mineral dust to total particle number concentrations is up to 15% over South American and Australian deserts and up to 40% in the North African and Asian deserts.

5   Wind-generated particles on the surface of the ocean are one of the most important constituents of atmospheric aerosol particles (Prather et al., 2013; Modini et al., 2015; Quinn et al., 2015). Generally, aerosol particles present in marine environments are referred to as marine particles. These particles comprise ocean emitted primary sea spray aerosol which includes particles containing organic components and sea salt, with high fractions of the former in smaller particles and vice versa, together with newly formed particles originating from precursor gases such as dimethylsulphide (DMS), long-range transport

10 of dust, anthropogenic pollution and biomass burning. Marine aerosol in the size range below 10 μm diameter usually featured a trimodal size distribution, showing that the production of marine aerosols is based on different mechanisms (Modini et al., 2015; Wex et al., 2016; Brooks and Thornton, 2018). Marine aerosol particles' hygroscopicity and their ability to act as CCN can be controlled by marine ocean physics, biology and chemistry (Quinn et al., 2015). Sea spray aerosol (SSA, including directly emitted organic matter and inorganic salt) contributed less than 30% of the CCN population on a global basis, with the

15 exception of high latitudes of the Southern Ocean where it may be larger (Quinn et al., 2017). Marine aerosols are expected to be a source of INPs, but less efficient than dust and biological particles (Wilson et al., 2015; DeMott et al., 2016; Gong et al., 2020a).

    The Cabo Verde (a.k.a. Cape Verde) islands are located in the tropical northern Atlantic Ocean. In 2006, the Cape Verde Atmospheric Observatory (CVAO) was set up by British, German and Cabo Verde scientific institutes. CVAO provides a

20 platform for studying remote marine aerosols, as well as long-range transported continental and Saharan-dust aerosols (Fomba et al., 2013; van Pinxteren et al., 2020). CVAO is located at the far edge of the island in the direction of air mass inflow to the island. Therefore, air masses observed at CVAO are free from local anthropogenic pollution. The aerosol composition at CVAO varies spatially and temporally together with meteorological conditions. Generally, CVAO receives clean marine air masses from the northern Atlantic Ocean during late spring and summer, while it receives air masses with high dust load from

25 the Sahara during late fall and winter (Fomba et al., 2014).

    In this study, a detailed overview of long-term particle microphysical data sets obtained at CVAO is provided. Particle number size distributions (PNSDs) in the size range from 20 nm to 10 μm and light absorbing carbon (LAC) mass concentration were characterized from April 2008 to December 2017. CCN number concentration (N_{CCN}) and particle hygroscopicity (expressed as κ) were characterized from October 2015 to March 2016 and from September to November 2017. The corresponding backward

30 trajectories were calculated from the HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) Model from April 2008 to December 2017. It is known that short-term measurements are not representative because the real world has large spatial heterogeneity and temporal variability. The long-term measurement at CVAO provided precious particle microphysical data sets to help understand the particles' nature in a dust-marine environment. Furthermore, long-term measurements can be used for a better constrain of modeling efforts (Reddington et al., 2017).

It is challenging to analyze and interpret the large data sets resulting from long-term measurements. Machine learning algorithms can build powerful models that can quickly and automatically analyze bigger, more complex data sets and deliver faster, accurate results. The choice of unsupervised or supervised machine learning algorithms will depend on the scientific questions to be answered. Each method has its advantages and disadvantages. The different machine learning algorithms can

5   be very sensitive to user-defined parameters and implementation. A number of freely available open-source and easy-to-use packages in popular programming languages, such as Python, have been developed in academia and industry. Machine learning approaches have been increasingly used for particle classification. For example, Christopoulos et al. (2018) implemented a supervised decision tree for aerosol classification for on-line single-particle mass spectrometry. Atwood et al. (2019) classified aerosol population types using an unsupervised K-means cluster analysis. In the present study, an unsupervised machine

10   learning algorithm K-mean is used to help us understand better the particle microphysical properties at CVAO.

2   Experiment and Methods

2.1   Sampling location and measurement setup

The measurement station (CVAO, 16°51′49" N, 24°52′02" W) is located at the northeastern shore of the São Vicente island. It is about 10 m a.s.l and 70 m from the coastline. The São Vicente island itself is in the north Atlantic Ocean, ∼900 km off the

15   African coast. This region constantly experiences northeasterly winds, with an average wind speed of about 7.3 m s^{-1}. The average annual temperature at São Vicente is about 23.6 ± 4.0 °C. Cabo Verde has very little precipitation, with an annual rainfall of about 3 to 10 events and 24-350 mm, mainly between August and October (Fomba et al., 2013).

An aerosol PM$_{10}$ inlet, installed on top of a 32 m tower, was deployed to remove particles larger than 10 μm. The purpose of installing an aerosol inlet on such a high tower is to minimize the influence of sea spray aerosol generated in the surf zone

20   at the coastline. Downstream of the PM$_{10}$ inlet, a vertical stainless-steel sampling pipe (1/2 in. outer diameter), together with a 1-meter long diffusion dryer, was placed on top of a measurement container. All microphysical instruments were placed inside this air-conditioned container. The microphysical instruments included a TROPOS-type MPSS (mobility particle size spectrometer), an APS (aerodynamic particle sizer, model 3321, TSI Inc.), a MAAP (multi-angle absorption photometer, type 5012, Thermo Scientific Inc.) and a CCNC (cloud condensation nuclei counter, Droplet Measurement Technologies). More

25   details concerning the CVAO station setup can be found in Gong et al. (2020b).

2.2   Particle size

PNSDs in the size range from 20 nm to 10 μm were measured by using a TROPOS-type MPSS and an APS. The electrical mobility distribution (measured by the MPSS in the size range from 20 to ∼800 nm) is converted to a PNSD by applying an inversion algorithm to correct for multiply charged aerosol particles (Wiedensohler, 1988) and diffusional losses (Wiedensohler

30   et al., 2012). In this study, APS data, i.e., PNSDs for sizes above ∼700 nm, was involved to enable a multiple charge correction of the MPSS data when applying the inversion algorithm (fpr more explanation see Pfeifer et al., 2014). Prior to using APS

data in the inversion, the aerodynamic diameters as measured by APS had to be converted to geometric diameters, for which a dry dynamic shape factor and particle density are needed. The dry dynamic shape factor (χ) and a density (ρ) of sodium chloride is 1.08 and 2160 kg m^{-3} (Kelly and McMurry, 1992; Gysel et al., 2002), respectively, whereas the dynamic shape factor and density of mineral dust is 1.25 (Kaaden et al., 2009) and in a range of 2450-2700 kg m^{-3} (Haywood et al., 2001).

5   An averaged shape factor and density of 1.17 and 2000 kg m^{-3} was used in this study (Schladitz et al., 2011). Overall, the combined PNSD is given based on volume equivalent particle diameters in this study. Size-dependent particle losses within the inlet were corrected (refer to Appendix A). Based on the combined PNSDs, total particle number concentration (N_{total}) and particle number concentrations (PNC) in different size ranges can be calculated. The MPSS and APS were calibrated regularly during the measurements.

10   ## 2.3   Light absorbing carbon

Light absorbing carbon (LAC) concentrations were derived from measurements with a multi-angle absorption photometer (MAAP, Petzold and Schönlinner, 2004). The MAAP measures the optical transmission and reflection at a wavelength of 637 nm of a filter while it is loaded with particles. An internal algorithm calculates an equivalent black carbon concentration (eBC, Petzold et al., 2013) using a mass specific absorption coefficient (MAC) of 6.6 m^{2}g^{-1}. The MAC is a conversion factor between

15   the light absorption coefficient (σ_{abs}), which is the primary measured parameter, and a corresponding eBC, which are related by:

$$\sigma_{abs} = eBC * MAC \tag{1}$$

In Müller et al. (2011), it was shown that the values for eBC and σ_{abs} had to be corrected by up to a factor of 1.05 to correct for an originally wrong assumption of the wavelength. Furthermore, filter-based absorption photometers are affected by an

20   artifact due to light scattering particles. While the multi-angle approach of the MAAP compensates for this to a large extend, it was shown in Müller et al. (2011) that σ_{abs} is still overestimated by about 1% of the value of the light scattering coefficient (σ_{sca}) at the wavelength of 637 nm. The calculation of the fully corrected light absorption coefficient ($\sigma_{abs,corr}$) is shown in the following equation:

$$\sigma_{abs,corr} = 1.05 * \sigma_{abs} - 0.01 * \sigma_{sca} \tag{2}$$

25   Since the σ_{sca} was not measured directly, it was derived from PNSDs using Mie scattering theory (Bohren and Huffman, 2008). The required complex refractive index was assumed to be 1.49-0.01*i. This value is a reasonable assumption for less absorbing marine aerosols (Hess et al., 1998). A comparison of probability density functions (PDFs) between σ_{abs} and $\sigma_{abs,corr}$ is shown in the Appendix B.

2.4   Cloud condensation nuclei and particle hygroscopicity

30   A cloud condensation nuclei counter (CCNC, Roberts and Nenes, 2005) was used to measure CCN number concentration (N_{CCN}). The main part of the CCNC is a cylindrical continuous-flow thermal-gradient diffusion chamber. A constant stream-

wise temperature gradient is established to adjust a quasi constant centerline supersaturation. During the measurement, the sampled aerosol particles are guided through this chamber within a sheath flow along the centerline. Depending on the supersaturation adjusted in the CCNC and the particle's hygroscopic properties, particles can become activated to droplets.

In this study, the calibrated supersaturations were 0.15%, 0.20%, 0.30%, 0.50% and 0.70%. The supersaturation is switched
5 every 12 minutes. The measurement data during the first 5 minutes and the last 30 seconds at each supersaturation were excluded from the data analysis to assure a stable column temperature when switching the supersaturation, and the remaining data points were averaged. To ensure the accuracy of the results, a supersaturation calibration was done repeatedly during the long-term deployment of the instrument at the site (more details concerning calibration method in Gysel and Stratmann, 2013).

Whether or not a particle can be activated to a droplet depends on its chemical composition, its dry size and the maximum
10 supersaturation it encounters (according to Köhler theory, Köhler, 1936). A single parameter κ can be used to describe the relationship between particle hygroscopicity and dry diameter at which the particle will be activated (Petters and Kreidenweis, 2007). Assuming the surface tension of the examined solution droplets ($\sigma_{s/\alpha}$) to be that of pure water, in this study κ was calculated as follows:

$$\kappa = \frac{4A^3}{27d_{\mathrm{crit}}^3 \ln^2(1+SS)} \tag{3}$$

15 with

$$A = \frac{4\sigma_{s/\alpha}M_\omega}{RT\rho_\omega} \tag{4}$$

where SS and d_{crit} are supersaturation in % and the critical diameter above which all particles activate into droplets at this given supersaturation, respectively. M_ω is the molar mass of water; R=8.314 J (K mol)$^{-1}$ is the ideal gas constant; T is the absolute temperature; ρ_ω is the density of water.

20 In this method, aerosol particles are assumed to be internally-mixed. All particles in the neighborhood of a given size are assumed to have a similar κ. A specific particle can be activated to a droplet if its dry diameter is larger than or equal to d_{crit} at a fixed supersaturation. Therefore, d_{crit} is the diameter at which the cumulative particle number concentration from the largest diameter to d_{crit} is equal to N_{CCN}. In this study, the measured N_{CCN} and PNSDs are both used to derive d_{crit}.

2.5 Backward trajectories

25 A backward trajectory calculation was conducted with the HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) Model (Rolph, 2003; Stein et al., 2015). The backward trajectories were calculated for 6 days with a resolution of 6 hours, arriving at 200 m above the measurement site.

2.6 Unsupervised machine learning algorithm

An unsupervised K-means algorithm was used in this study to classify aerosol particle types. The K-means model clusters
30 samples by trying to separate data into k groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares. Given a set of observations ($x_1, x_2, ..., x_n$), where each observation is a d-dimensional real vector, K-means

clustering aims to partition the n observations into k ($\leq n$) sets $C = C_1$, C_2, ..., C_k so as to minimize the within-cluster sum-of-squares. Formally, the objective is to find:

$$\sum_{i=0}^{n} \min_{u_j \in C} (||x_i - u_j||^2) \tag{5}$$

where u_j is the mean of points in C_j.

It is important to start the analysis by looking into the number of clusters, k, to specify the optimal k. In this study, we used an elbow plot and Calinski-Harabasz score (Caliński and Harabasz, 1974) to determine the optimal k, which is explained in more detail in Appendix C.

3 Results and discussion

3.1 Overview of data sets

3.1.1 Time series of particle number size distribution

Particle number size distribution (PNSD) is one of the most important features of atmospheric aerosols, as particles in different size ranges have different formation routes and behaviors. The contour plots for PNSDs of the supermicron and submicron particle are shown in Fig. 1 (a) and (b), respectively. The color scale indicates particle number concentration (dN/dlogD$_p$) in per cubic centimeters. Because of a relatively lower number concentration of supermicron particles, different color bar scales were adopted for submicron and supermicron particles in Fig. 1. Most of the time, Aitken and accumulation modes in the submicron size range and a coarse mode in the supermicron size range are observed. The Aitken mode ranges from ~20 to ~80 nm, and the accumulation mode ranges from ~80 to ~1000 nm. The number concentration of supermicron particles shows a high variation, from 1 to above 100, with a median value of 3.8 cm^{-3}.

3.1.2 Seasonal variation of light absorbing carbon

The monthly variation of σ_{abs} and $\sigma_{abs,corr}$, averaged over all years, is shown in Fig. 2. For $\sigma_{abs,corr}$, concentrations were mostly below $0.5\,\text{Mm}^{-1}$. However, higher values of up to $1\,\text{Mm}^{-1}$ were observed in October, December and January. The light absorbing aerosols usually include black carbon, brown carbon and dust. The CVAO station is free of local anthropogenic pollution because it is located at the northeastern shore and experiences constant northeasterly winds. Therefore it is highly unlikely that black carbon and brown carbon alone could explain the observed variation of the σ_{abs} values. The higher values of σ_{abs} coincide with months during which high dust concentrations were generally observed in this study (Sect. 3.2.2).

3.1.3 Time series and analysis of CCN

Figure 3 shows the time series of CCN number concentration (N_{CCN}) in the upper panel, d_{crit} in the middle panel and κ values in the lower panel, with different colors for different supersaturations. N_{CCN} shows large variability. N_{CCN} at a supersaturation
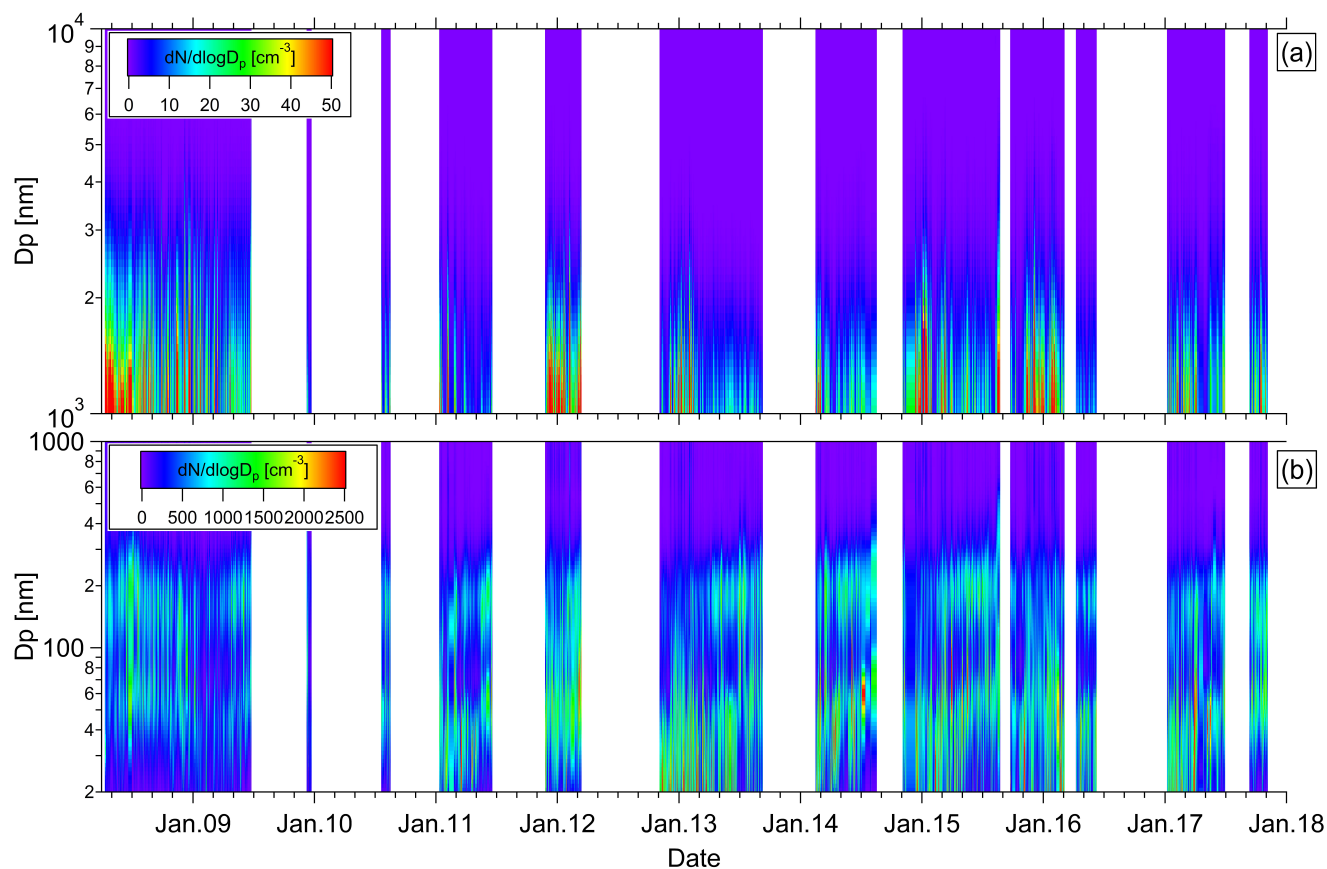
Figure 1. Contour plots for PNSDs of the supermicron size range (1000 nm to 10 μm) in panel (a) and the submicron size range (10 to 1000 nm) in panel (b).

of 0.30 ($N_{\text{CCN, 0.30\%}}$) varied from $\sim$10 to above 1000, with a mean value of 336 cm^{-3}. The high variability of N_{CCN} also can be seen in the wide-spread probability density function of CCN number concentrations (PDF-N_{CCN}), shown in Fig. 4 (a). The PDF-N_{CCN} shows multiple modes at supersaturations of 0.15%, 0.20%, 0.30%, 0.50% and 0.70%, which indicates that particles acting as CCN arriving at CVAO originate from different sources. The boxplot of N_{CCN} at different supersaturations

5 is shown in Fig. 5 (a). It is clear that N_{CCN} values increase towards a higher supersaturation, because increasingly smaller particles can be activated to droplets at a increasingly higher supersaturations. The mean/median values of N_{CCN} also exhibit a large variability, increasing from a mean of 233 to 485 cm^{-3} for a variation in supersaturations from 0.15% to 0.70%.

Time series of d_{crit} values at all supersaturations (0.15%, 0.20%, 0.30%, 0.50% and 0.70%) are shown in Fig. 3 (b). The mean value of d_{crit} decreases from 131 nm at a supersaturation of 0.15% to 40 nm at a supersaturation of 0.70% (Fig. 5). As

10 can be seen from Fig. 4, for the lowest supersaturation of 0.15%, d_{crit} is above the size of 100 nm, located in the accumulation mode, whereas for the supersaturations of 0.70% and 0.50%, d_{crit} is below the size of 70 nm, located in the Aitken mode. The
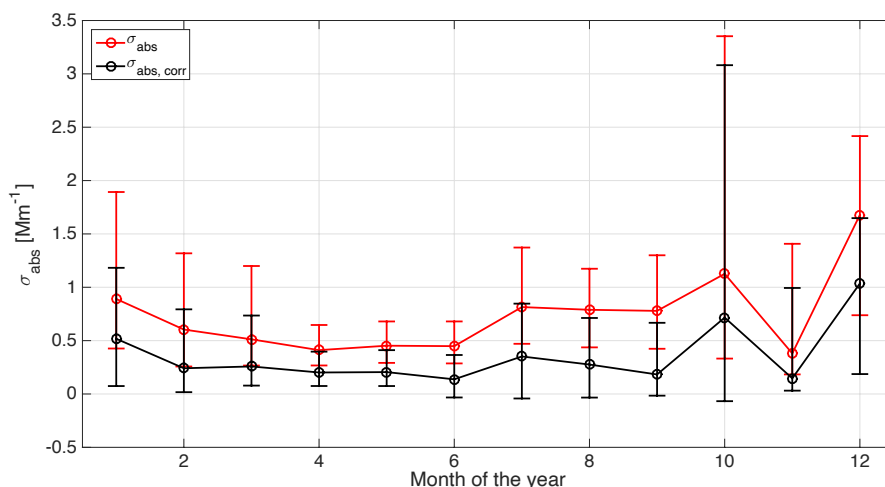
Figure 2. Monthly variation of the LAC mass concentration.

derived κ at a given supersaturation represents the particle averaged hygroscopicity around the corresponding d_{crit}. Therefore, hygroscopicities derived at supersaturations of 0.70% and 0.50% can be assumed to be representative of the Aitken, whereas hygroscopicities derived at a supersaturation of 0.15% can be assumed to be representative of the accumulation mode. d_{crit} at supersaturations of 0.20% and 0.30% are from about 70 to 100 nm. In this size range, marine aerosol particle size distributions

5 typically show a minimum between Aitken and accumulation mode (Modini et al., 2015; Gong et al., 2020b), also known as Hoppel minimum (Hoppel et al., 1986). The hygroscopicity derived at supersaturations of 0.20% to 0.30% represents a mixture of particles from both the accumulation and Aitken mode.

Figure 3(c) shows the time series of the particle hygroscopicity parameter κ. The κ values at each supersaturation show a larger variability during the whole campaign compared to a previous study of a one-month intensive measurement campaign at

10 CVAO (Gong et al., 2020b), with one standard deviation varying from 0.10 (SS=0.15%) to 0.16 (SS=0.20%, 0.30%). This also can be seen in the wide-spread PDF of κ at different supersaturations (Fig. 4 (c)). The PDF-κ values at each supersaturation show uni-model distributions, except for the supersaturation of 0.70%. The PDF-κ at a supersaturation of 0.70% is bimodal featuring a small mode with a peak at 0.16 and a large mode with a peak at 0.36. At this supersaturation of 0.70%, the larger κ values are present when CCN number concentration is low, and vice versa. Figure 5 (c) shows the overall κ as a function

15 of supersaturation. A slightly increasing trend of κ was observed from supersaturations 0.15% to 0.20%, whereas a slightly decreasing trend of κ value was observed from 0.20% and above. However, there is no strongly pronounced difference between κ at different supersaturations.

Figure 5 (d) shows κ as a function of d_{crit} and plotted error bars indicate standard and geometric standard deviations for d_{crit} and κ, respectively. In the size range from 50 to 100 nm (Aitken mode), a slightly increasing trend of κ over increasing d_{crit} is

20 observed, indicating that the sulfate produced from DMS oxidation rather than organics might be a major species for particle
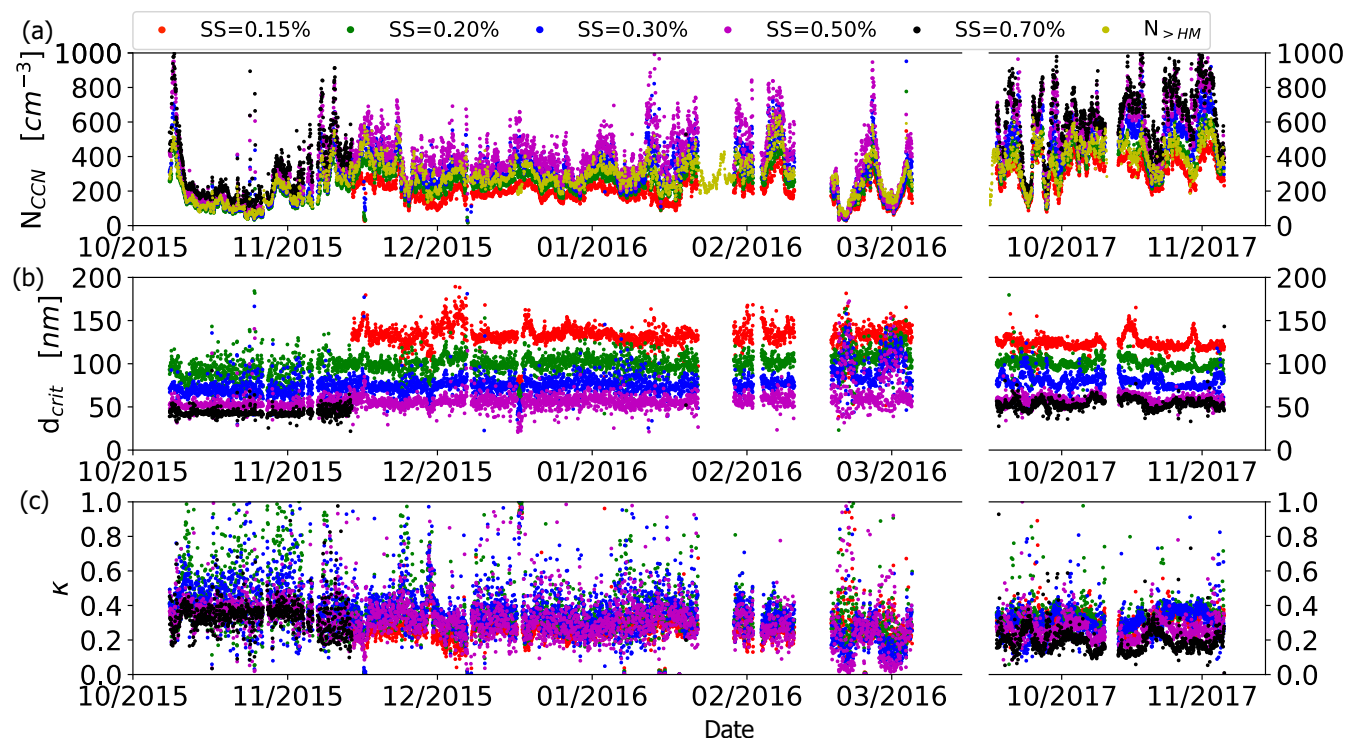
Figure 3. Time series of N_{CCN} in panel (a), d_{crit} in panel (b) and κ in panel (c). Red, green, blue, purple and black dots indicate supersaturations (SS) of 0.15%, 0.20%, 0.30%, 0.50% and 0.70%, respectively. Yellow dots in the upper panel show the time series of particle number concentration for particles larger than the Hoppel minimum diameter ($N_{>HM}$).

condensational growth. In the accumulation mode, κ is getting smaller. But the overall uncertainty in these derived values is rather large so that no further interpretation of the κ values at different sizes will be given. The overall average κ value is 0.32.

However, this value cannot be taken as an annual average, as no data exists for the summer months. In Barbados, both modeling (Pringle et al., 2010) and measurements (Wex et al., 2016) suggested that κ of marine aerosol in the tropical Atlantic

5 can show an annual variation with lowest values of around 0.3 to 0.4 in the summer months. Contrary, for Cabo Verde, Pringle et al. (2010) suggested values for κ of up to 0.6 for the time spanning April until July, while they suggest lower values of roughly 0.3 for fall and winter months, in agreement with our measurements.

The Hoppel minimum divides particles that were not activated to a cloud droplet, yet, from other, larger ones, which were. Therefore, the supersaturation of 0.30% can be seen to roughly represent the averaged maximum supersaturations in the clouds

10 along the path of the sampled air masses. On the other hand, we can use the integrated particle number concentration for particle sizes above the Hoppel minimum ($N_{>HM}$) to estimate atmospheric CCN number concentrations at a supersaturation of 0.30% at CVAO, as shown in Fig. 3 upper panel, where a good agreement between measured $N_{CCN,0.30\%}$ and calculated $N_{>HM}$ values can be seen. When comparing these values in a scatter plot, the coefficient of determination (R) is 0.89. By this, the CCN number
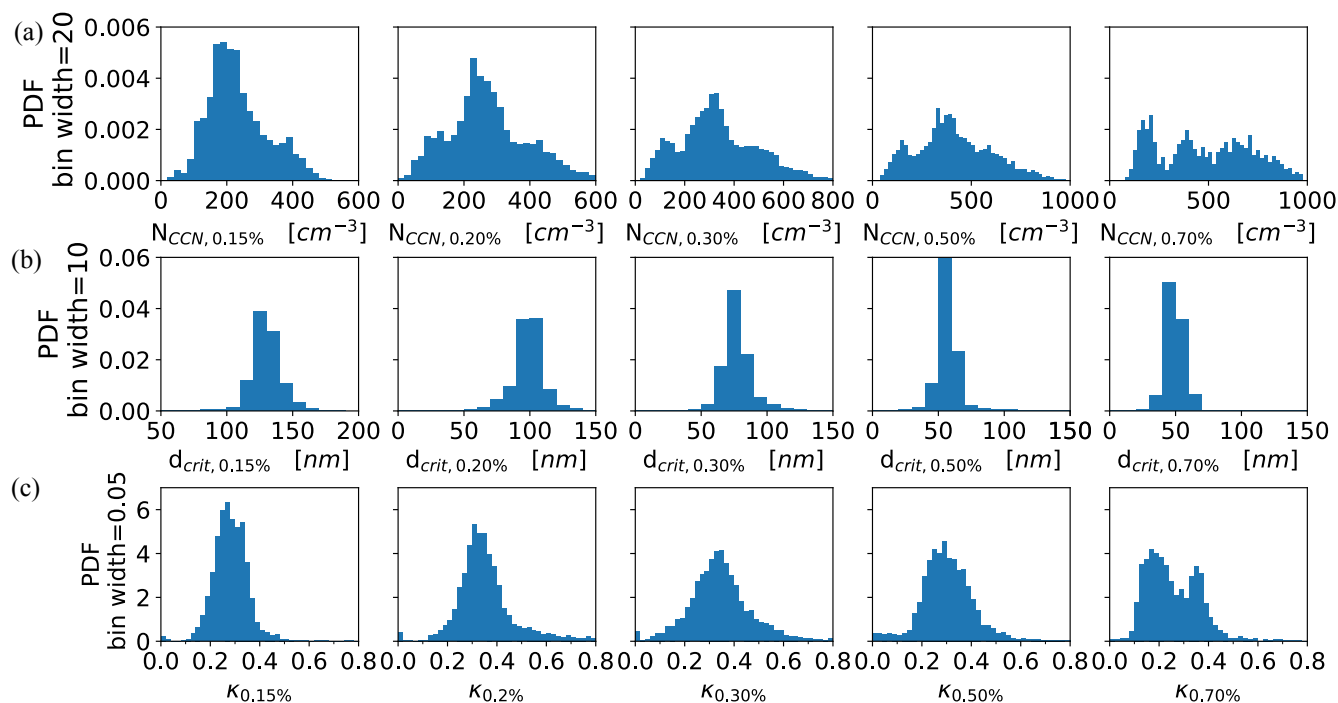
Figure 4. Probability of density functions (PDFs) of N_{CCN} (a), d_{crit} (b) and κ (c) at supersaturations from 0.15% to 0.70%.

concentration can be extended from the 9 months during which N_{CCN} was measured with a CCNC to a range of 10 years for which $N_{>\mathrm{HM}}$ values can be derived based on measured PNSDs. If the Hoppel minimum is not present, which means during the moderate and heavy dust periods (discussed in Sect. 3.2.1), we used the cumulative particle number concentration above 100 nm to estimate CCN number concentration. The 100 nm was chosen because the d_{crit} was about 100 nm at a supersaturation of

5   0.30% during dust periods (discussed in Sec. 3.3), and it also coincides with a boundary between two modes in PNSD of the moderate dust.

3.1.4   Air mass analysis

To understand the air mass origin and transport, we analyzed the backward trajectories at CVAO. From previous studies at CVAO, we learned that the marine boundary layer (MBL) was typically well mixed, with boundary layer heights from about

10   550 to 1100 m. An occasionally decoupled layer around 400 to 600 m was also observed (Gong et al., 2020b). Therefore, we used backward trajectories starting at 200 m altitude to be sure to represent aerosol collected at CVAO. During air mass transport, the aerosol in an air mass might be altered due to precipitation events. We disregarded backward trajectories from our analysis (~4.73% of total backward trajectories) when total precipitation exceeded 20 mm in the past 6 days.

Figure 6 shows the relative frequency of backward trajectories over the 10-year measurements. Most of the backward tra-

15   jectories featured paths over the North Atlantic Ocean and the coastal area in Africa and arrived at CVAO from the northwest
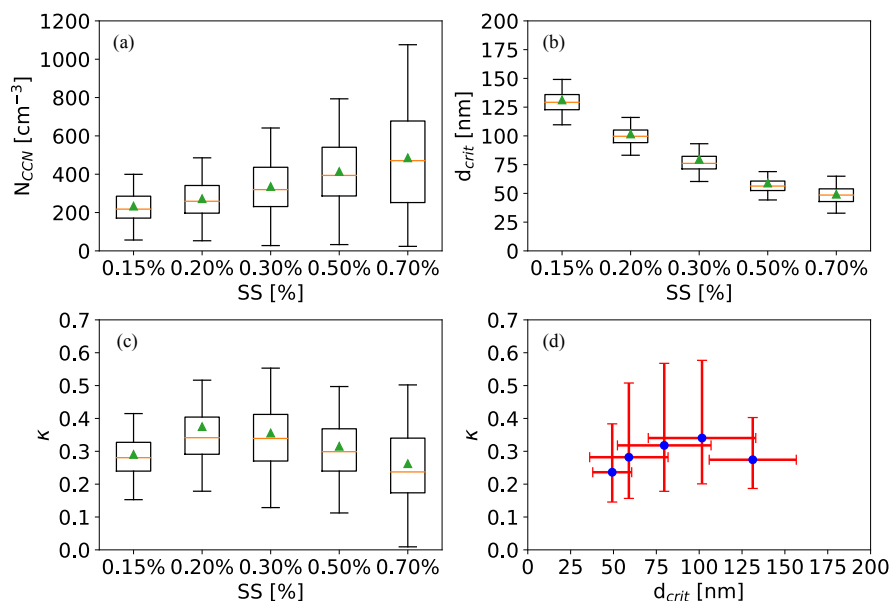
Figure 5. Boxplots of N_{CCN} (a), d_{crit} (b) and κ (c) at supersaturations (SS) from 0.15% to 0.70%. Whiskers show the 90th to 10th percentiles. Green solid triangles show the mean values. (d) κ as a function of d_{crit}. Error bars of d_{crit} show 1 standard deviation. Error bars of κ show 1 geometric standard deviation.

direction. A subgroup of backward trajectories originated in the Sahara, spent a few days above it and arrived at CVAO from a northwest direction. The frequency plot also reveals that backward trajectories featured paths over southern Europe and the Mediterranean region so that anthropogenically emitted particles from these regions might also be carried to CVAO.

3.2 Particle classification, origins and its monthly variation

3.2.1 Particle classification, origins

From the overview over the data sets introduced above, it can be seen that PNSDs, PNC, N_{CCN} and air mass origin show large variability. In the following, we examine how different air mass origins and histories are related to different aerosols being present at CVAO. Therefore, we implement a K-means algorithm to classify aerosol types according to the shape of PNSDs. The input parameter x_i for K-means was defined as the normalized PNC in different size ranges. We used PNC in the size ranges from 20 to 40 nm ($N_{20\_40\text{nm}}$), 40 to 80 nm ($N_{40\_80\text{nm}}$), 80 to 600 nm ($N_{80\_600\text{nm}}$), 600 to 1000 nm ($N_{600\_1000\text{nm}}$) and larger than 1000 nm ($N_{>1000\text{nm}}$). As the absolute numbers in different modes are very different, we normalized PNC in different size ranges to make sure that all size ranges are assessed equally. The normalization method can be found in Appendix C.

These size ranges were chosen based on assuming that particles in different size ranges feature different formation routes and behaviors. For the here presented data, higher $N_{20\_40\text{nm}}$ values indicated the presence of relatively freshly formed particles.
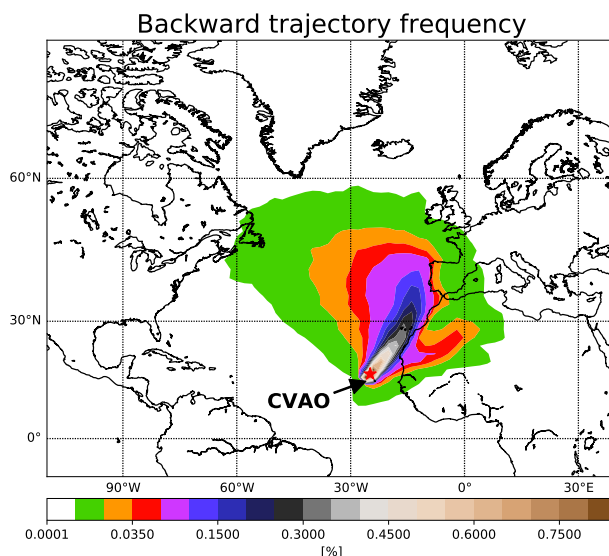
Figure 6. Relative frequency of backward trajectories arriving at CVAO.

Table 1. The input parameter x_i for K-means and the corresponding main aerosol mode and source

Parameter	Main aerosol mode
N_{20_40nm}	Freshly-formed particles
N_{40_80nm}	Aitken mode particles
N_{80_600nm}	Accumulation mode particles
N_{600_1000nm}	Mainly sea spray aerosol with possible dust fraction
$N_{>1000nm}$	Mainly dust, but also large sea spray aerosol

The boundary between Aitken and accumulation is roughly at 80 nm at CVAO. Therefore, N_{40_80nm} and N_{80_600nm} represented Aitken and accumulation mode particle number concentration, respectively. N_{600_1000nm} can be attributed to sea-salt with a possible contribution from dust particles (Schladitz et al., 2011). The largest observed particles ($N_{>1000nm}$) are directly linked to heavy dust plumes at CVAO with a small contribution from large sea spray aerosols (Gong et al., 2020b). A summary of
5 input parameters and the corresponding aerosol mode and source is shown in Tab. 1.

As mentioned above and described in detail in Appendix C, in this study we determined k=5 for the number of clusters. The resulting classification of PNSDs and particle volume size distributions (PVSDs) are shown in Fig. 7. The solid lines show the median of PNSDs and PVSDs of the 5 different aerosol clusters, with a linear scaling on the y-axis in Fig. 7 (a), (c) and a logarithmic in Fig. 7 (b), (d). The error bars indicate the range between 75th and 25th percentiles. Additionally, Figure 8 shows
10 the backward trajectory path frequency and backward trajectory height frequency which are connected to these 5 different clusters of PNSDs.
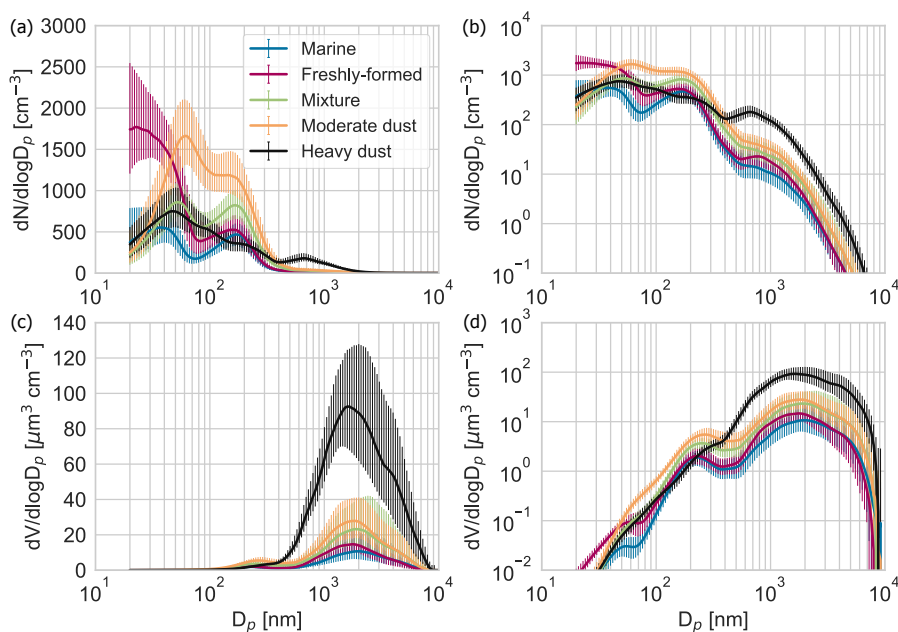
Figure 7. The median of PNSDs (a and b panel) and PVSDs (c and d panel) of marine (blue), freshly-form (purple), mixture (green), moderate dust (brown) and heavy dust (black) clusters with a linear (a and c panel) and a logarithmic (b and d panel) scaling on the y axis. Error bars show the range between 75th and 25th percentiles.

The marine type PNSD features three modes, i.e., Aitken, accumulation and sea spray aerosol, shown as the blue line in Fig. 7. The marine type featured the lowest N_{40_80nm}, N_{80_600nm}, N_{600_1000nm} and $N_{>1000nm}$ values. The Hoppel minimum of marine type PNSDs is at roughly 70 nm. During the marine period, the sea spray aerosol mode can be assumed to be primary marine aerosols, as discussed in Modini et al. (2015) and Gong et al. (2020b). The PNSD representing the freshly-formed

5   cluster is shown as a purple line in Fig. 7. This cluster featured the highest N_{20_40nm}. While we did not measure PNSDs below 20 nm, it can still be safely assumed that such high concentrations of small particles indicate new particle formation events which must have happened recently in the respective air masses. It is interesting to note that this is related to a generally low aerosol particle volume concentration (see Fig. 7 (c) and (d)), as well as mass concentration, as only comparably small numbers of larger particles are present, similar to the above described marine type, so new particle formation seems to happen

10 predominantly in marine air masses with low particle number concentrations, an observation which has also been made for Barbados in Wex et al. (2016).

The frequency of backward trajectory paths that were observed during times when PNSDs from marine and freshly-formed clusters were present a similar pattern, as shown in Fig. 8 (a1) and (b1). The majority of these backward trajectories spend the whole 6-day period over the North or western North Atlantic Ocean and arrived at CVAO from the northeast direction. Some

15 of these backward trajectories touched southwest Europe and crossed the coastal areas in northwest Africa.
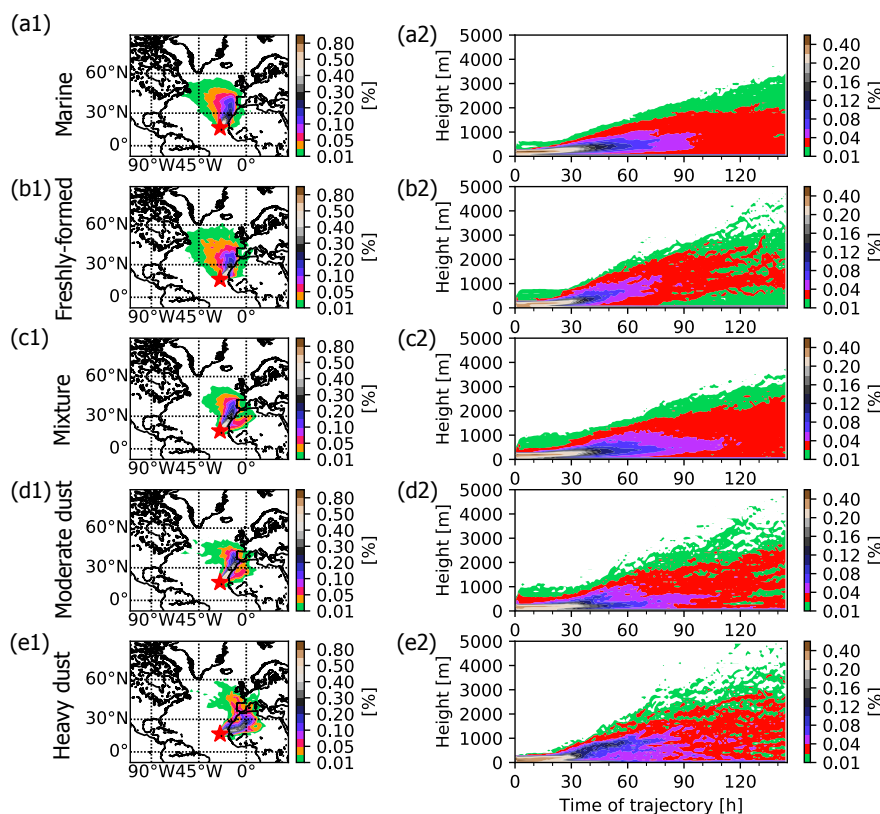
Figure 8. Relative frequency of backward trajectory paths arriving at CVAO during marine (a1), freshly-formed (b1), mixture (c1), moderate (d1) and heavy dust (e1) periods and the corresponding relative frequency of backward trajectory height in panel (a2) to (e2).

No clearly visible Hoppel minimum is present for the moderate dust cluster, as shown in a brown line in Fig. 7 (a) and (b). The moderate dust cluster featured the second highest N_{coarse}. We also observed a very high concentration of Aitken and accumulation mode particles for the moderate dust cluster. The heavy dust cluster is shown as the black line in Fig. 7. No clearly visible Hoppel minimum is present for this cluster, either. The heavy dust cluster featured the highest N_{coarse}. It is worth

5   to be mentioned that similar PNSDs were found in previous field measurements at the Sahara (Weinzierl et al., 2009; Kandler et al., 2009; Kaaden et al., 2009). The highest N_{coarse} also means that the highest volume and mass concentrations (Fig. 7 (c) and (d) black line) are present in the heavy dust cluster, which was already reported in a previous study (Fomba et al., 2014).

Finally, the mixture cluster PNSD is shown as green lines in Fig. 7, with three clearly distinguished Aitken, accumulation and coarse modes. $N_{>1000nm}$ are between the marine and dust clusters. The Hoppel minimum of the mixture cluster is at ∼90 nm.

10   This may indicate that for this cluster the maximum supersaturation reached in the trade wind clouds is lower than the marine and freshly-formed clusters, for which, as said above, it was at 70 nm. These two clusters have lower number concentrations for larger particles. Therefore it does not surprise that PVSDs of the mixture cluster are between the marine and dust clusters. In the presence of higher particle concentrations for larger particles, an overall higher amount of water condenses onto particles

at cloud base, which can be expected to lead to an overall lower supersaturation (Reutter et al., 2009), enabling only larger particles to be activated to cloud droplets, as observed in our data sets.

The frequency of backward trajectories during the moderate dust cluster featured a similar pattern as during the mixture period (as shown in Fig. 8 (c1) and (d1)). There are two major paths during mixture and dust periods. One class of back-
5   ward trajectory was from the North Atlantic Ocean and crossed the coastal areas in northwest Africa. One class of backward trajectories spend some days over the Sahara. So overall, for PNSDs which were characterized to belong to the mixture or moderate dust clusters, also sea spray particles contributed overall, and likely also to the particles at larger sizes. The backward trajectories during the heavy dust period were mainly from the Sahara, which justifies assuming that the related particles in this highest N_{coarse} came from Saharan dust particles.

10   The relative frequency of backward trajectory heights is shown in the right panels of Fig. 8. Gong et al. (2020b) measured that marine boundary height at CVAO varied roughly from 550 to 1100 m. From Fig. 8 it can be seen that a large fraction of air masses came from or had contact with the free troposphere. In this context, it is worth noticing that it has been described in the past, that new particle formation in the marine environment often takes place in the free troposphere, followed by growth of these particles into the CCN size range (Korhonen et al., 2008; Merikanto et al., 2009).

15   ### 3.2.2   Monthly variation

In order to understand the large temporal variability of aerosol types on a monthly basis, the monthly variation of the relative frequency of occurrence, number fraction and volume fraction of the five different aerosol clusters are shown in Fig. 9 (a), (b) and (c), respectively. It can clearly be seen that relative frequency of occurrence, number fraction and volume fraction feature different behaviors. For example, the heavy dust cluster represented 14.6% of the cases encountered in January, contributed
20   only 11.7% to particle number, however about 40.4% to particle volume/mass. This needs to be considered when relating number-based properties such as CCN and INP concentrations to e.g. the mass-based chemical composition.

The heavy dust cluster is mainly present during wintertime. This is because, during wintertime, dust is transported westward from the Sahara in the lower troposphere. The deposition of dust particles takes place over the eastern tropical Atlantic and Cabo Verde region. During springtime, the marine aerosol is more often observed. This is because, during springtime, the
25   CVAO mostly receives North Atlantic marine air parcels along with the NNE trade winds.

The above discussed marine, mixture and moderate dust aerosol clusters were recognized from Gong et al. (2020b), who characterized the sources of the particles at CVAO during September and October 2017. The freshly-formed and heavy dust clusters did not show up during this previous study and are described here for the first time for CVAO. It is worth mentioning that even in the relative simple dust-marine intersect location at Cabo Verde, five different aerosol clusters were classified, and
30   these 5 clusters featured different characteristics such as size distributions and origins.

3.3   Evaluate marine and dust contribution to CCN

One important purpose of this study is to evaluate CCN population and particle hygroscopicity for marine or dust aerosol dominated environments. Therefore, we compared κ and N_{CCN} during the marine and the dust periods, where for the latter,
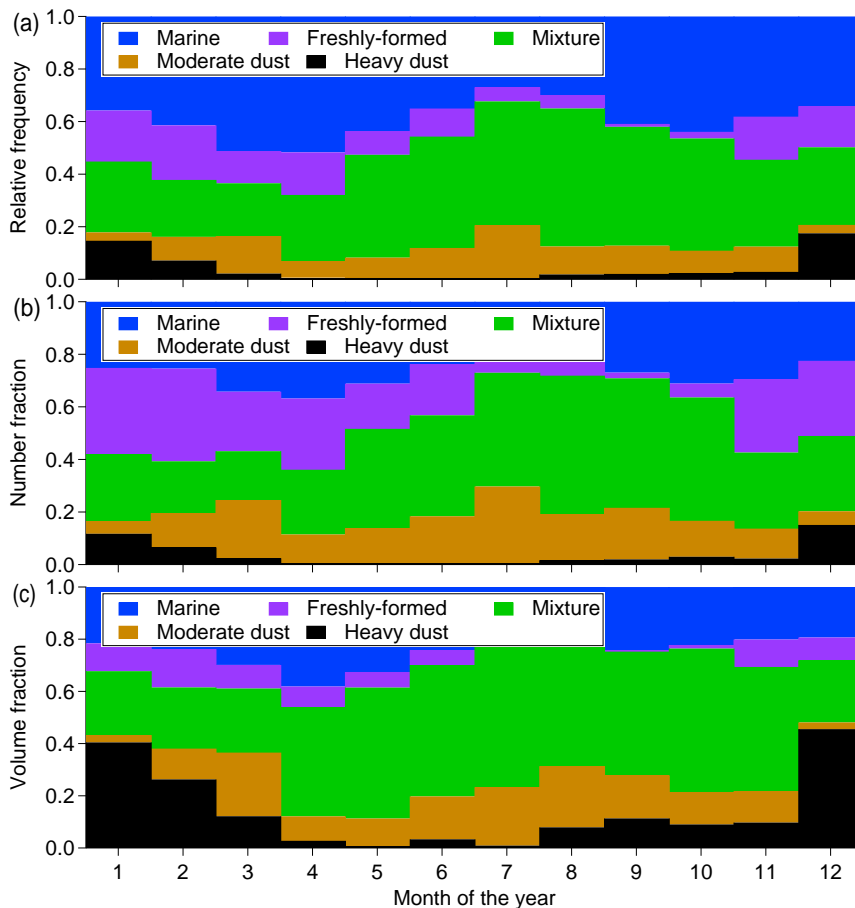
Figure 9. Monthly variation of relative frequency (a), number fraction (b) and volume fraction (c) of different aerosol clusters.

moderate and heavy dust were combined. Figure 10 (a) shows a boxplot of N_{CCN} at different supersaturations during 9 months with CCNC measurement. N_{CCN} at a supersaturation of 0.30% (as a proxy for the average maximum supersaturation encountered in clouds) during the dust period is about 2.5 times higher than that during marine periods. N_{HM} can be used to estimate atmospheric N_{CCN} at a supersaturation of 0.30%. Therefore, to extend atmospheric N_{CCN} from 9 months to 10 years, we also

5 derived N_{HM} for marine and dust periods from 2008 to 2018, and also show these values in Fig. 10 (b). N_{CCN} at a supersaturation of 0.30% in Fig. 10 (a), derived from only 9 months of measurements, and N_{HM} in Fig. 10 (b) derived for 10 years agree well, giving confidence in the extension of the CCN data set. Also, it is clear that the atmospheric N_{HM} (i.e., $N_{CCN,0.30\%}$), meaning also particles being able to act as CCN, is about 2.5 times higher during dust periods than during marine periods at CVAO.

10   Besides, N_{CCN} during dust periods is generally higher than that during marine periods, increasingly so with increasing supersaturation or decreasing d_{crit}. For example, at a supersaturation of 0.70%, corresponding to a d_{crit} of about 50 nm, N_{CCN}
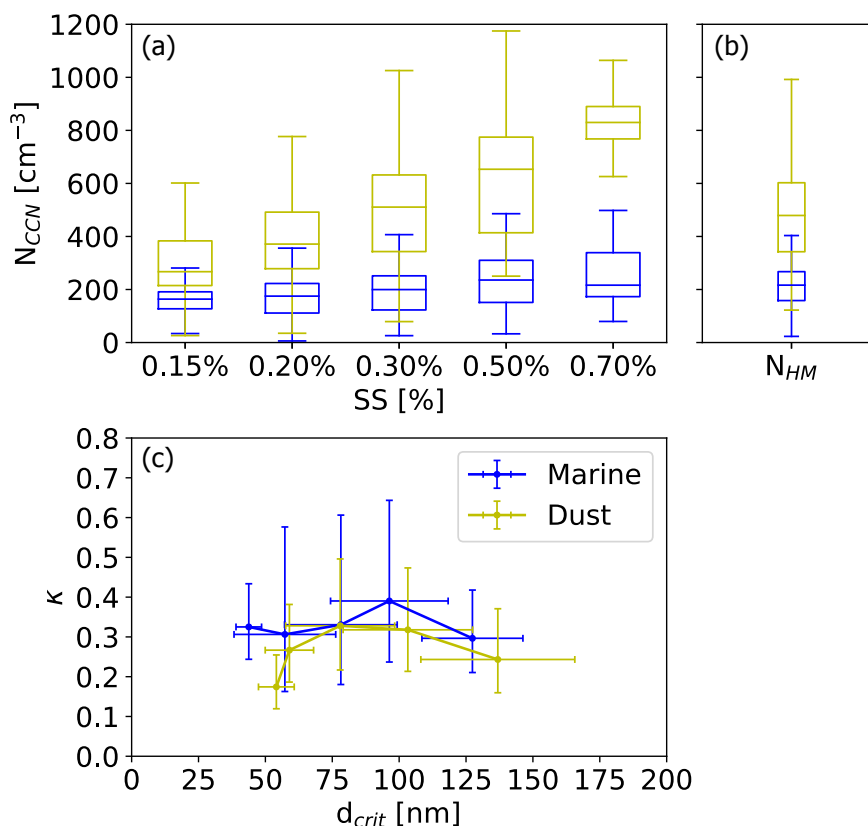
Figure 10. (a) Boxplot of measured N_{CCN} during marine and dust periods at supersaturations (SS) from 0.15% to 0.70%; (b) Boxplot of N_{HM} during marine and dust periods; (c) κ as a function of d_{crit} during dust and marine periods. Dust and marine periods are shown in yellow and blue, respectively.

during dust periods is four times higher than during marine periods. This is interesting, as in the size range around 50 nm dust particles are not expected to be present in significant numbers. An enhanced Aitken mode particle number concentration was observed during moderate dust periods, showing up clearly separated from one large mode observed for particles > 100 nm. We interpret these particles to stem from new particle formation in air mass associated with the dust, an phenomenon which has been observed in a previous field study (Dupart et al., 2012). However, this is highly speculative and proving or disproving this hypothesis is beyond the scope of this paper.

Figure 10 (c) shows κ as a function of d_{crit}. The error bars of d_{crit} show one standard deviation and error bars of κ show one geometric standard deviation. κ during the marine period generally agreed with that of the dust period within uncertainty when considering the error bars. Therefore, the CCN-derived hygroscopicity for dust and marine periods in the size range between 40 and 140 nm shows no significant difference. This is a surprising but also important result, as this may suggest the use of generalized values for κ for related cases. As stated above, these κ values are all representative of the months from October

until March. They are comparably small and in agreement with the modeling study by Pringle et al. (2010). The majority of particles in marine air masses may originate from new particle formation, not from sea spray (Korhonen et al., 2008; Merikanto et al., 2009), explaining the derived κ which is lower than for sea salt particles. Dust particles, on the other hand, were reported to show hygroscopic behavior in the past (Karydis et al., 2011) and may also get more hygroscopic during aging. However, for

5 dust type aerosols the nature of particles in the size range for which κ was determined, particularly for the size range < 100 nm, is not fully clear, as mentioned above. But it did become clear here that the observed particles do not differ in hygroscopicity from those observed in marine type aerosols, which strengthens our suggestion that these particles also originate from new particle formation.

4 Synopsis

10 In the following, we summarize aerosol related results from this study and extend a bit on some of them. As mentioned before, long-term data sets as this one can yield valuable insights for our general understanding but also for improving the performance of models (Reddington et al., 2017). This section is meant to present a broader picture of processes connected to the atmospheric aerosol without going into too much detail.

 In this study, after classification based on machine learning, we found five different aerosol clusters for the 10 year-long data

15 set obtained at CVAO – marine, freshly-formed, mixture, moderate dust and heavy dust.

 When examining monthly variations, we showed that heavy dust was mainly observed during wintertime, where it was present roughly 10% of the time, and contributed, on a monthly base, roughly 10% of particle number but up to $> 40\%$ of the particle volume and mass. Moderate dust events, however, occurred throughout the year, typically for more than 10% of the time during each month, contributing up to more than 20% of all particles in number. Fall and winter months were those during

20 which also the aerosol absorption was highest, and in the absence of local anthropogenic pollution sources upwind of CVAO, the observed absorption can be assumed to originate from dust particles.

 A special feature of the moderate dust type aerosols was a pronounced Aitken mode, which, as said above occurred frequently. While the determination of the nature of these particles is beyond the scope of this study, we speculate if new particle formation events may also occur in moderately dust laden air masses, based on not yet known processes, and propose this topic

25 for further research.

 Clean marine periods were most frequently present during springtime, occurring roughly 50% of the time in March and April. Periods during which the freshly-formed type aerosol prevailed had similar air mass origin as marine aerosol, indicating that the typical and well-known way of new particle formation mainly occurs in marine aerosols with low particle number concentration and low particulate surface area. Together, clean marine and freshly-formed air masses occurred for more than

30 50% of the time from October until May, while the mixed type aerosol was the most often occurring type in the remaining months.

 Dust and marine type aerosols featured clearly different PNSDs. Enhanced particle number concentrations and CCN number concentrations were observed when air masses came from the Sahara. For air masses originating from the Atlantic Ocean,

PNSDs featured the lowest number concentrations in Aitken, accumulation and sea spray mode. CCN number concentration at a supersaturation of 0.30% (a proxy for the average maximum supersaturation encountered in clouds) during the marine period was about 2.5 times lower than during the dust period. However, the CCN-derived particle hygroscopicity (in the size range from 40 to 140 nm) for marine and dust periods showed no significant difference, independent of the air mass. This latter result, however, was based solely on data from the months of October until March. Averaging all κ values obtained in this study, an average value of 0.32 was found, again for the months from October until March, and omitting small size-dependent differences which, however, all were within uncertainty.

The Hoppel minimum, a dip between the Aitken and accumulation mode indicating the division between particles that had been activated to cloud droplets from those that had not, was found to be at roughly 80 nm. More precisely, the Hoppel minimum was at 70 nm for marine and freshly-formed type aerosols and at 90 nm for the mixture type aerosol, where the latter had higher number concentrations for larger particles. This larger size for the Hoppel minimum likely is indicative of these particles being a larger sink for water vapor during cloud droplet formation, lowering the effective supersaturation for air masses with mixture type aerosol.

5 Conclusions and future work

In this study, we summarized measurements of the aerosol microphysical properties in a representative marine-dust aerosol intersect environment at Cape Verde Atmospheric Observatory (CVAO). In-situ aerosol measurements included light absorbing carbon (LAC), particle number size distribution (PNSD), cloud condensation nuclei (CCN) and corresponding backward trajectories. An unsupervised machine learning algorithm, K-means, was implemented to classify aerosol clusters, which were then the base for the classification of particle sources according to the shape of particle number size distributions. Based on the classification, we studied the air mass origins for different aerosol clusters and parsed out the marine and dust contributions to CCN and their hygroscopicity.

We proved that an unsupervised machine learning algorithm K-means can be a good tool to classify aerosol types. It benefits from the long-term measurements with extremely large data sets. We refrain from repeating the key results here, as they were summarized in the previous section. But we point out that the summarized data sets can also help to shape new research questions towards future studies at CVAO. The freshly-formed particles and the corresponding backward trajectories indicated one type of new particle formation which happened at CVAO mainly in air masses of marine origin. The height of backward trajectories indicated that new particle formation might have occurred in the free troposphere and particles were then transported down to the marine boundary layer. While new particle formation in the marine environment was assumed to occur in the free troposphere before, we still lack the knowledge of the overall mechanism, such as the precursors, growth rate, temperature and humidity effects of new particle formation, and also the fraction of CCN which originate from new particle formation. Further, the moderate dust period featured high concentrations of Aitken mode particles, while the heavy dust period did not. More investigations are required to understand the reason behind that.

The data sets collected at CVAO are invaluable as, due to their long duration, they can help to better understand the aerosol microphysical properties, which are at the base of aerosol-cloud interactions. Furthermore, the data sets represent highly valuable information for driving, evaluating and constraining atmospheric model simulations. Long-term measurements of aerosol size, light absorbing carbon and CCN are important in the future because the real world has a large spatial heterogeneity and temporal variability.

Data availability. All data described and analyzed in this study are published in the Data Publisher for Earth & Environmental Science (PANGAEA) (https://doi.org/10.1594/PANGAEA.921321, Gong et al., 2020).

Appendix A: Particle loss

We used the Particle Loss Calculator (written in IGOR Pro, von der Weiden et al., 2009) to calculate the particle losses within the inlet tube system. We considered size-dependent particle losses due to sedimentation, diffusion, turbulent inertial deposition, inertial deposition in a contraction and in a bend. Note that the losses inside the diffusion dryer in this study are equivalent to the losses in a 15 m long tube (Wiedensohler et al., 2012). Figure A1 shows the resulting particle losses (in %) as a function of particle size.
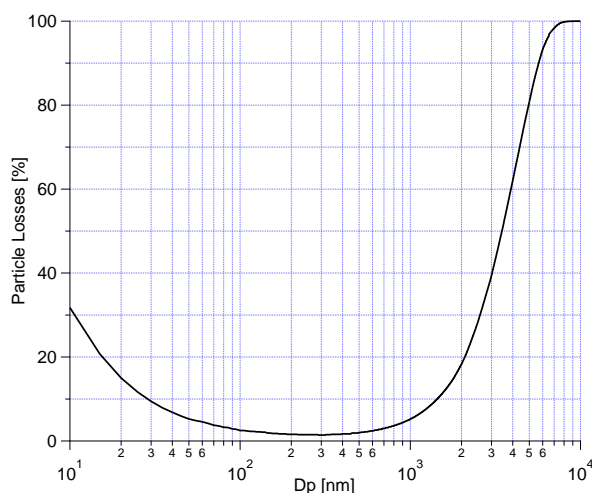
Figure A1. Size-dependent particle loss through the inlet at CVAO station.

Appendix B: MAAP correction

Figure B1 shows the PDF of σ_{abs} (red line) and $\sigma_{abs,corr}$ (black line). Although corrections for the scattering artifact for filter-based absorption photometers are highly uncertain, the results clearly underline the importance of the correction.
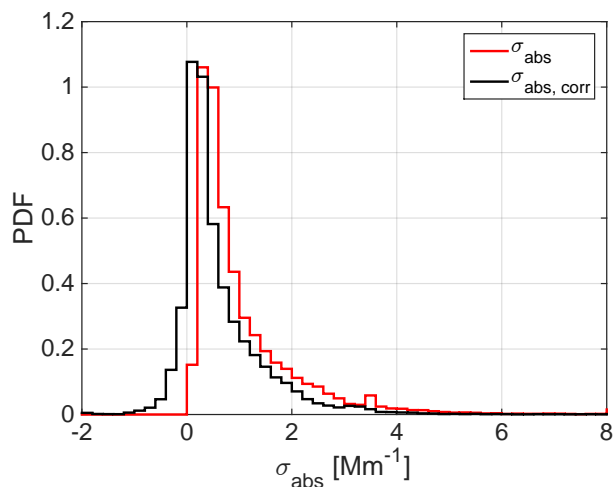
Figure B1. Probability density functions (PDFs) of σ_{abs} as red line and $\sigma_{abs,corr}$ as black line.

Appendix C: Input parameters normalization and determine classification clusters

5 To make sure PNC in all size ranges are assessed equally when applying the K-means algorithm, we normalized PNC in different size ranges separately based on a min-max normalization:

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{C1}$$

where x_i is i_{th} PNC value in a specific size range. For PNC in a specific size range, the minimum PNC value of that size range gets transformed into a 0, the maximum PNC value gets transformed into a 1, and every other value gets transformed into a

10 decimal between 0 and 1.

The elbow method is a useful graphical tool to estimate the optimal number of clusters k for a given task. Theoretically, if k increases, the within-cluster sum of squared errors (a.k.a. distortion) will decrease. This is because the samples will be closer to the centroids they are assigned to. In the elbow method, distortion begins to decrease rapidly and then switches to a more slowly decrease at a certain k value. Figure C1 shows the elbow plot as a blue line. The distortion started decreasing in a linear

15 fashion after k=5. Therefore, k=5 is the optimal number of clusters. Also, we used the Calinski-Harabasz score (Caliński and Harabasz, 1974), which computes the ratio of dispersion between and within clusters. The vertical black dashed line indicates the highest Calinski-Harabasz value which occurred, which also is at a number of five clusters. Therefore, overall, the optimal number of clusters was derived to be five, here, which is the value we then used within this study.
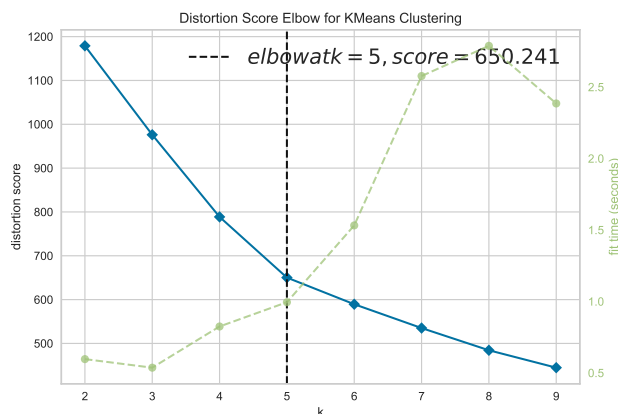
Figure C1. The distortion as a function of k (a.k.a. elbow plot) is shown as blue line. The highest Calinski-Harabasz score is shown as dashed black line. The consumed time as a function of k is shown as dashed green line.

Author contributions. XG conceived the study and wrote the manuscript with contributions from HW, TM and FS. TM performed particle number size distribution and light absorbing carbon measurements. SH performed cloud condensation nuclei measurement. XG evaluated and analyzed the data and developed analysis methods. XG, HW, TM, SH, JV, AW and FS discussed the results. All of the co-authors proofread and commented the manuscript.

5 *Competing interests.* The authors declare that they have no conflict of interests.

Acknowledgements. The authors acknowledge the Leibniz Association SAW funding for the project "Marine biological production, organic aerosol particles and marine clouds: a Process Chain (MarParCloud)", SAW-2016-TROPOS-2. We acknowledge the SOPRAN (Surface Ocean Processes in the Anthropocene) project, supported by the Bundesministerium für Bildung und Forschung (BMBF) under grant number 03F0611J. We acknowledge the SAMUM (Saharan Mineral Dust Experiment) project, founded by the German Research Foundation under
10 grant MU 2669/1–1 and grant FOR 539. The authors acknowledge the CVAO site manager, Luis Neves, to maintain SMPS, APS and CCN measurements at CVAO. We acknowledge Thomas Conrath and Kay Weinhold for their technique support.

References

Atwood, S. A., Kreidenweis, S. M., DeMott, P. J., Petters, M. D., Cornwell, G. C., Martin, A. C., and Moore, K. A.: Classification of aerosol population type and cloud condensation nuclei properties in a coastal California littoral environment using an unsupervised cluster model, Atmospheric Chemistry and Physics, 19, 6931–6947, https://doi.org/10.5194/acp-19-6931-2019, https://www.atmos-chem-phys.net/19/6931/2019/, 2019.

Bohren, C. F. and Huffman, D. R.: Absorption and scattering of light by small particles, John Wiley & Sons, 2008.

Brooks, S. D. and Thornton, D. C.: Marine Aerosols and Clouds, Annual Review of Marine Science, 10, 289–313, https://doi.org/10.1146/annurev-marine-121916-063148, https://www.annualreviews.org/doi/abs/10.1146/annurev-marine-121916-063148, 2018.

Caliński, T. and Harabasz, J.: A dendrite method for cluster analysis, Communications in Statistics, 3, 1–27, https://doi.org/10.1080/03610927408827101, https://www.tandfonline.com/doi/abs/10.1080/03610927408827101, 1974.

Carslaw, K. S., Boucher, O., Spracklen, D. V., Mann, G. W., Rae, J. G. L., Woodward, S., and Kulmala, M.: A review of natural aerosol interactions and feedbacks within the Earth system, Atmospheric Chemistry and Physics, 10, 1701–1737, https://doi.org/10.5194/acp-10-1701-2010, https://www.atmos-chem-phys.net/10/1701/2010/, 2010.

Christopoulos, C. D., Garimella, S., Zawadowicz, M. A., Möhler, O., and Cziczo, D. J.: A machine learning approach to aerosol classification for single-particle mass spectrometry, Atmos. Meas. Tech., 11, 5687–5699, https://doi.org/10.5194/amt-11-5687-2018, https://www.atmos-meas-tech.net/11/5687/2018/, 2018.

DeMott, P. J., Sassen, K., Poellot, M. R., Baumgardner, D., Rogers, D. C., Brooks, S. D., Prenni, A. J., and Kreidenweis, S. M.: African dust aerosols as atmospheric ice nuclei, Geophysical Research Letters, 30, n/a–n/a, https://doi.org/10.1029/2003GL017410, http://dx.doi.org/10.1029/2003GL017410, 2003.

DeMott, P. J., Hill, T. C. J., McCluskey, C. S., Prather, K. A., Collins, D. B., Sullivan, R. C., Ruppel, M. J., Mason, R. H., Irish, V. E., Lee, T., Hwang, C. Y., Rhee, T. S., Snider, J. R., McMeeking, G. R., Dhaniyala, S., Lewis, E. R., Wentzell, J. J. B., Abbatt, J., Lee, C., Sultana, C. M., Ault, A. P., Axson, J. L., Diaz Martinez, M., Venero, I., Santos-Figueroa, G., Stokes, M. D., Deane, G. B., Mayol-Bracero, O. L., Grassian, V. H., Bertram, T. H., Bertram, A. K., Moffett, B. F., and Franc, G. D.: Sea spray aerosol as a unique source of ice nucleating particles, Proceedings of the National Academy of Sciences, 113, 5797–5803, https://doi.org/10.1073/pnas.1514034112, http://www.pnas.org/content/113/21/5797.abstract, 2016.

Dupart, Y., King, S. M., Nekat, B., Nowak, A., Wiedensohler, A., Herrmann, H., David, G., Thomas, B., Miffre, A., Rairoux, P., D'Anna, B., and George, C.: Mineral dust photochemistry induces nucleation events in the presence of SO2, Proceedings of the National Academy of Sciences, 109, 20 842–20 847, https://doi.org/10.1073/pnas.1212297109, https://www.pnas.org/content/109/51/20842, 2012.

Fomba, K. W., Müller, K., van Pinxteren, D., and Herrmann, H.: Aerosol size-resolved trace metal composition in remote northern tropical Atlantic marine environment: case study Cape Verde islands, Atmos. Chem. Phys., 13, 4801–4814, https://doi.org/10.5194/acp-13-4801-2013, http://www.atmos-chem-phys.net/13/4801/2013/, 2013.

Fomba, K. W., Müller, K., van Pinxteren, D., Poulain, L., van Pinxteren, M., and Herrmann, H.: Long-term chemical characterization of tropical and marine aerosols at the Cape Verde Atmospheric Observatory (CVAO) from 2007 to 2011, Atmos. Chem. Phys., 14, 8883–8904, https://doi.org/10.5194/acp-14-8883-2014, http://www.atmos-chem-phys.net/14/8883/2014/, 2014.

Gong, X., Wex, H., Müller, T., Henning, S., Voigtländer, J., Wiedensohler, A., and Stratmann, F.: Ground-based measurements on aerosol particles microphysical properties at Cabo Verde from 2008 to 2017, https://doi.org/10.1594/PANGAEA.921321, https://doi.org/10.1594/PANGAEA.921321, 2020.

Gong, X., Wex, H., van Pinxteren, M., Triesch, N., Fomba, K. W., Lubitz, J., Stolle, C., Robinson, T.-B., Müller, T., Herrmann, H., and Stratmann, F.: Characterization of aerosol particles at Cabo Verde close to sea level and at the cloud level – Part 2: Ice-nucleating particles in air, cloud and seawater, Atmospheric Chemistry and Physics, 20, 1451–1468, https://doi.org/10.5194/acp-20-1451-2020, https://www.atmos-chem-phys.net/20/1451/2020/, 2020a.

Gong, X., Wex, H., Voigtländer, J., Fomba, K. W., Weinhold, K., van Pinxteren, M., Henning, S., Müller, T., Herrmann, H., and Stratmann, F.: Characterization of aerosol particles at Cabo Verde close to sea level and at the cloud level – Part 1: Particle number size distribution, cloud condensation nuclei and their origins, Atmospheric Chemistry and Physics, 20, 1431–1449, https://doi.org/10.5194/acp-20-1431-2020, https://www.atmos-chem-phys.net/20/1431/2020/, 2020b.

Gysel, M. and Stratmann, F.: WP3 - NA3: In-situ chemical, physical and optical properties of aerosols, Deliverable D3.11: Standardized protocol for CCN measurements, Tech. rep., http://www.actris.net/Publications/ACTRISQualityStandards/tabid/11271/language/en-GB/Default.aspx, 2013.

Gysel, M., Weingartner, E., and Baltensperger, U.: Hygroscopicity of Aerosol Particles at Low Temperatures. 2. Theoretical and Experimental Hygroscopic Properties of Laboratory Generated Aerosols, Environmental Science & Technology, 36, 63–68, https://doi.org/10.1021/es010055g, https://doi.org/10.1021/es010055g, 2002.

Haywood, J. M., Francis, P. N., Glew, M. D., and Taylor, J. P.: Optical properties and direct radiative effect of Saharan dust: A case study of two Saharan dust outbreaks using aircraft data, Journal of Geophysical Research: Atmospheres, 106, 18 417–18 430, https://doi.org/10.1029/2000jd900319, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900319, 2001.

Hess, M., Koepke, P., and Schult, I.: Optical Properties of Aerosols and Clouds: The Software Package OPAC, Bulletin of the American Meteorological Society, 79, 831–844, https://doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2, https://doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2, 1998.

Hoppel, W., Frick, G., and Larson, R.: Effect of nonprecipitating clouds on the aerosol size distribution in the marine boundary layer, Geophysical Research Letters, 13, 125–128, https://doi.org/doi:10.1029/GL013i002p00125, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/GL013i002p00125, 1986.

Kaaden, N., Massling, A., Schladitz, A., Müller, T., Kandler, K., Schütz, L., Weinzierl, B., Petzold, A., Tesche, M., Leinert, S., Deutscher, C., Ebert, M., Weinbruch, S., and Wiedensohler, A.: State of mixing, shape factor, number size distribution, and hygroscopic growth of the Saharan anthropogenic and mineral dust aerosol at Tinfou, Morocco, Tellus B, 61, 51–63, https://doi.org/doi:10.1111/j.1600-0889.2008.00388.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0889.2008.00388.x, 2009.

Kandler, K., SchüTz, L., Deutscher, C., Ebert, M., Hofmann, H., JäCkel, S., Jaenicke, R., Knippertz, P., Lieke, K., Massling, A., Petzold, A., Schladitz, A., Weinzierl, B., Wiedensohler, A., Zorn, S., and Weinbruch, S.: Size distribution, mass concentration, chemical and mineralogical composition and derived optical parameters of the boundary layer aerosol at Tinfou, Morocco, during SAMUM 2006, Tellus B: Chemical and Physical Meteorology, 61, 32–50, https://doi.org/10.1111/j.1600-0889.2008.00385.x, https://doi.org/10.1111/j.1600-0889.2008.00385.x, 2009.

Karydis, V. A., Kumar, P., Barahona, D., Sokolik, I. N., and Nenes, A.: On the effect of dust particles on global cloud condensation nuclei and cloud droplet number, Journal of Geophysical Research: Atmospheres, 116, n/a–n/a, https://doi.org/10.1029/2011JD016283, http://dx.doi.org/10.1029/2011JD016283, 2011.

Kelly, W. P. and McMurry, P. H.: Measurement of Particle Density by Inertial Classification of Differential Mobility Analyzer–Generated Monodisperse Aerosols, Aerosol Science and Technology, 17, 199–212, https://doi.org/10.1080/02786829208959571, https://doi.org/10.1080/02786829208959571, 1992.

Köhler, H.: The nucleus in and the growth of hygroscopic droplets, Transactions of the Faraday Society, 32, 1152–1161, 1936.

5 Kok, J. F., Ridley, D. A., Zhou, Q., Miller, R. L., Zhao, C., Heald, C. L., Ward, D. S., Albani, S., and Haustein, K.: Smaller desert dust cooling effect estimated from analysis of dust size and abundance, Nature Geoscience, 10, 274–278, 2017.

Kok, J. F., Ward, D. S., Mahowald, N. M., and Evan, A. T.: Global and regional importance of the direct dust-climate feedback, Nature communications, 9, 1–11, 2018.

Korhonen, H., Carslaw, K. S., Spracklen, D. V., Mann, G. W., and Woodhouse, M. T.: Influence of oceanic dimethyl sulfide emissions on

10 cloud condensation nuclei concentrations and seasonality over the remote Southern Hemisphere oceans: A global model study, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/https://doi.org/10.1029/2007JD009718, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009718, 2008.

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

15 Modini, R. L., Frossard, A. A., Ahlm, L., Russell, L. M., Corrigan, C. E., Roberts, G. C., Hawkins, L. N., Schroder, J. C., Bertram, A. K., Zhao, R., Lee, A. K. Y., Abbatt, J. P. D., Lin, J., Nenes, A., Wang, Z., Wonaschütz, A., Sorooshian, A., None, K. J., Jonsson, H., Seinfeld, J. H., Toom-Sauntry, D., Macdonald, A. M., and Leaitch, W. R.: Primary marine aerosol-cloud interactions off the coast of California, Journal of Geophysical Research: Atmospheres, 120, 4282–4303, https://doi.org/10.1002/2014JD022963, http://dx.doi.org/10.1002/2014JD022963, 2015.

20 Müller, T., Henzing, J. S., de Leeuw, G., Wiedensohler, A., Alastuey, A., Angelov, H., Bizjak, M., Collaud Coen, M., Engström, J. E., Gruening, C., Hillamo, R., Hoffer, A., Imre, K., Ivanow, P., Jennings, G., Sun, J. Y., Kalivitis, N., Karlsson, H., Komppula, M., Laj, P., Li, S.-M., Lunder, C., Marinoni, A., Martins dos Santos, S., Moerman, M., Nowak, A., Ogren, J. A., Petzold, A., Pichon, J. M., Rodriquez, S., Sharma, S., Sheridan, P. J., Teinilä, K., Tuch, T., Viana, M., Virkkula, A., Weingartner, E., Wilhelm, R., and Wang, Y. Q.: Characterization and intercomparison of aerosol absorption photometers: result of two intercomparison workshops, Atmospheric Measurement Techniques,

25 4, 245–268, https://doi.org/10.5194/amt-4-245-2011, https://www.atmos-meas-tech.net/4/245/2011/, 2011.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7, 1961–1971, https://doi.org/10.5194/acp-7-1961-2007, http://www.atmos-chem-phys.net/7/1961/2007/, 2007.

Petzold, A. and Schönlinner, M.: Multi-angle absorption photometry—a new method for the measurement of aerosol light absorption and atmospheric black carbon, Journal of Aerosol Science, 35, 421 – 441, https://doi.org/https://doi.org/10.1016/j.jaerosci.2003.09.005, 2004.

30 Petzold, A., Ogren, J. A., Fiebig, M., Laj, P., Li, S.-M., Baltensperger, U., Holzer-Popp, T., Kinne, S., Pappalardo, G., Sugimoto, N., Wehrli, C., Wiedensohler, A., and Zhang, X.-Y.: Recommendations for reporting "black carbon" measurements, Atmospheric Chemistry and Physics, 13, 8365–8379, https://doi.org/10.5194/acp-13-8365-2013, 2013.

Pfeifer, S., Birmili, W., Schladitz, A., Müller, T., Nowak, A., and Wiedensohler, A.: A fast and easy-to-implement inversion algorithm for mobility particle size spectrometers considering particle number size distribution information outside of the detection range, Atmospheric

35 Measurement Techniques, 7, 95–105, https://doi.org/10.5194/amt-7-95-2014, https://amt.copernicus.org/articles/7/95/2014/, 2014.

Pöschl, U.: Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects, Angewandte Chemie International Edition, 44, 7520–7540, https://doi.org/10.1002/anie.200501122, https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.200501122, 2005.

Prather, K. A., Bertram, T. H., Grassian, V. H., Deane, G. B., Stokes, M. D., DeMott, P. J., Aluwihare, L. I., Palenik, B. P., Azam, F., Seinfeld, J. H., Moffet, R. C., Molina, M. J., Cappa, C. D., Geiger, F. M., Roberts, G. C., Russell, L. M., Ault, A. P., Baltrusaitis, J., Collins, D. B., Corrigan, C. E., Cuadra-Rodriguez, L. A., Ebben, C. J., Forestieri, S. D., Guasco, T. L., Hersey, S. P., Kim, M. J., Lambert, W. F., Modini, R. L., Mui, W., Pedler, B. E., Ruppel, M. J., Ryder, O. S., Schoepp, N. G., Sullivan, R. C., and Zhao, D.: Bringing the ocean into the laboratory to probe the chemical complexity of sea spray aerosol, Proceedings of the National Academy of Sciences, 110, 7550–7555, https://doi.org/10.1073/pnas.1300262110, http://www.pnas.org/content/110/19/7550.abstract, 2013.

Pringle, K. J., Tost, H., Pozzer, A., Pöschl, U., and Lelieveld, J.: Global distribution of the effective aerosol hygroscopicity parameter for CCN activation, Atmos. Chem. Phys., 10, 5241–5255, https://doi.org/10.5194/acp-10-5241-2010, http://www.atmos-chem-phys.net/10/5241/2010/, 2010.

Quinn, P. K., Collins, D. B., Grassian, V. H., Prather, K. A., and Bates, T. S.: Chemistry and Related Properties of Freshly Emitted Sea Spray Aerosol, Chemical Reviews, 115, 4383–4399, https://doi.org/10.1021/cr500713g, http://dx.doi.org/10.1021/cr500713g, 2015.

Quinn, P. K., Coffman, D. J., Johnson, J. E., Upchurch, L. M., and Bates, T. S.: Small fraction of marine cloud condensation nuclei made up of sea spray aerosol, Nature Geoscience, 10, 674, https://doi.org/10.1038/ngeo3003 https://www.nature.com/articles/ngeo3003#supplementary-information, https://doi.org/10.1038/ngeo3003, 2017.

Reddington, C. L., Carslaw, K. S., Stier, P., Schutgens, N., Coe, H., Liu, D., Allan, J., Browse, J., Pringle, K. J., Lee, L. A., Yoshioka, M., Johnson, J. S., Regayre, L. A., Spracklen, D. V., Mann, G. W., Clarke, A., Hermann, M., Henning, S., Wex, H., Kristensen, T. B., Leaitch, W. R., Pöschl, U., Rose, D., Andreae, M. O., Schmale, J., Kondo, Y., Oshima, N., Schwarz, J. P., Nenes, A., Anderson, B., Roberts, G. C., Snider, J. R., Leck, C., Quinn, P. K., Chi, X., Ding, A., Jimenez, J. L., and Zhang, Q.: The Global Aerosol Synthesis and Science Project (GASSP): Measurements and Modeling to Reduce Uncertainty, Bulletin of the American Meteorological Society, 98, 1857–1877, https://doi.org/10.1175/BAMS-D-15-00317.1, https://doi.org/10.1175/BAMS-D-15-00317.1, 2017.

Reutter, P., Su, H., Trentmann, J., Simmel, M., Rose, D., Gunthe, S. S., Wernli, H., Andreae, M. O., and Pöschl, U.: Aerosol- and updraft-limited regimes of cloud droplet formation: influence of particle number, size and hygroscopicity on the activation of cloud condensation nuclei (CCN), Atmospheric Chemistry and Physics, 9, 7067–7080, https://doi.org/10.5194/acp-9-7067-2009, https://www.atmos-chem-phys.net/9/7067/2009/, 2009.

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

Rolph, G.: Real-time environmental applications and display system (READY) website, http://www. arl. noaa. gov/ready. php, 2003.

Sassen, K., DeMott, P. J., Prospero, J. M., and Poellot, M. R.: Saharan dust storms and indirect aerosol effects on clouds: CRYSTAL-FACE results, Geophysical Research Letters, 30, https://doi.org/doi:10.1029/2003GL017371, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL017371, 2003.

Schladitz, A., Müller, T., Nowak, A., Kandler, K., Lieke, K., Massling, A., and Wiedensohler, A.: In situ aerosol characterization at Cape Verde, Part1: Particle number size distributions, hygroscopic growth and state of mixing of mrine and Saharan dust aerosol, Tellus B, 63, 531–548, https://doi.org/10.1111/j.1600-0889.2011.00569.x, http://dx.doi.org/10.1111/j.1600-0889.2011.00569.x, 2011.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, Bulletin of the American Meteorological Society, 96, 2059–2077, https://doi.org/10.1175/bams-d-14-00110.1, https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-14-00110.1, 2015.

van Pinxteren, M., Fomba, K. W., Triesch, N., Stolle, C., Wurl, O., Bahlmann, E., Gong, X., Voigtländer, J., Wex, H., Robinson, T.-B., Barthel, S., Zeppenfeld, S., Hoffmann, E. H., Roveretto, M., Li, C., Grosselin, B., Daële, V., Senf, F., van Pinxteren, D., Manzi, M., Zabalegui,

N., Frka, S., Gašparović, B., Pereira, R., Li, T., Wen, L., Li, J., Zhu, C., Chen, H., Chen, J., Fiedler, B., von Tümpling, W., Read, K. A., Punjabi, S., Lewis, A. C., Hopkins, J. R., Carpenter, L. J., Peeken, I., Rixen, T., Schulz-Bull, D., Monge, M. E., Mellouki, A., George, C., Stratmann, F., and Herrmann, H.: Marine organic matter in the remote environment of the Cape Verde islands – an introduction and overview to the MarParCloud campaign, Atmospheric Chemistry and Physics, 20, 6921–6951, https://doi.org/10.5194/acp-20-6921-2020, https://www.atmos-chem-phys.net/20/6921/2020/, 2020.

von der Weiden, S. L., Drewnick, F., and Borrmann, S.: Particle Loss Calculator - a new software tool for the assessment of the performance of aerosol inlet systems, Atmos. Meas. Tech., 2, 479–494, https://doi.org/10.5194/amt-2-479-2009, http://www.atmos-meas-tech.net/2/479/2009/, 2009.

Weinzierl, B., Petzold, A., Esselborn, M., Wirth, M., Rasp, K., Kandler, K., Schütz, L., Koepke, P., and Fiebig, M.: Airborne measurements of dust layer properties, particle size distribution and mixing state of Saharan dust during SAMUM 2006, Tellus B, 61, 96–117, https://doi.org/doi:10.1111/j.1600-0889.2008.00392.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0889.2008.00392.x, 2009.

Wex, H., Dieckmann, K., Roberts, G. C., Conrath, T., Izaguirre, M. A., Hartmann, S., Herenz, P., Schäfer, M., Ditas, F., Schmeissner, T., Henning, S., Wehner, B., Siebert, H., and Stratmann, F.: Aerosol arriving on the Caribbean island of Barbados: physical properties and origin, Atmos. Chem. Phys., 16, 14 107–14 130, https://doi.org/10.5194/acp-16-14107-2016, http://www.atmos-chem-phys.net/16/14107/2016/, 2016.

Wiedensohler, A.: An approximation of the bipolar charge distribution for particles in the submicron size range, Journal of Aerosol Science, 19, 387–389, https://doi.org/http://dx.doi.org/10.1016/0021-8502(88)90278-9, //www.sciencedirect.com/science/article/pii/0021850288902789, 1988.

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

Wilson, T. W., Ladino, L. A., Alpert, P. A., Breckels, M. N., Brooks, I. M., Browse, J., Burrows, S. M., Carslaw, K. S., Huffman, J. A., Judd, C., Kilthau, W. P., Mason, R. H., McFiggans, G., Miller, L. A., Najera, J. J., Polishchuk, E., Rae, S., Schiller, C. L., Si, M., Temprado, J. V., Whale, T. F., Wong, J. P. S., Wurl, O., Yakobi-Hancock, J. D., Abbatt, J. P. D., Aller, J. Y., Bertram, A. K., Knopf, D. A., and Murray, B. J.: A marine biogenic source of atmospheric ice-nucleating particles, Nature, 525, 234–238, https://doi.org/10.1038/nature14986, http://dx.doi.org/10.1038/nature14986, 2015.

Zender, C. S., Miller, R. L. L., and Tegen, I.: Quantifying mineral dust mass budgets:Terminology, constraints, and current estimates, Eos, Transactions American Geophysical Union, 85, 509–512, https://doi.org/10.1029/2004EO480002, 2004.